
UDK 159.932:534

534. 159

Pregledni rad

Prihvaćeno 10.06.2000.

Juraj Bakran, Vlasta Erdeljac i Nikolaj Lazić
Filozofski fakultet, Zagreb
Hrvatska

ANALIZA ZVUČNOG OKRUŽENJA

SAŽETAK

U članku se raspravlja o doprinosu pojedinih svojstava govornog zvuka procesu perceptivnog odvajanja mnoštva simultano prisutnih zvukova kojima smo okruženi. Uz pregled doprinosa pojedinih autora toj relativno malo istraživanoj temi (posebno kod nas) i komentar na temelju iskustava iz vlastitih istraživanja, predlaže se nekoliko (za naše govorno područje) novih termina nužnih za razumijevanje problema.

Ključne riječi: govorni zvuk, zvučni tok, zvučno okruženje, psihoakustika

UVOD

Tema o kojoj se ovdje raspravlja do sada kod nas nije sustavno obrađivana, unatoč tome što je poznata i stalno prisutna. Neka od kod nas objavljenih temeljnih istraživanja, pridonijela su ovoj raspravi, ali nisu interpretirana u tom svjetlu. Neprestano smo izloženi mnoštvu simultanih, pomiješanih zvukova. Da bismo se u toj mješavini mogli snalaziti, moramo biti u stanju prepoznavati pojedine smislene cjeline. U tom procesu važnu ulogu ima odluka o tome koji dijelovi osjetilnog stimulusa pripadaju istom objektu ili događaju iz okoline. O čemu se zapravo radi, o kojem pojedinom, konkretnom zvuku, može se prepoznati samo u slučaju prave kombinacije osjetilnih elemenata. *Gestalt* psiholozi početkom ovog stoljeća otvorili su pitanje organizacije percepcije, ali u svojim istraživanjima pretežno su se bavili vizualnim kanalom, dok je problem auditivne organizacije ostao zapostavljen. Jedan od razloga za to je i relativno zaostajanje tehnike manipuliranja zvukom. Tek zamahom digitalne tehnologije obrade zvuka posljednjih desetljeća omogućeno je sustavno laboratorijsko proučavanje auditivne percepcije. Problem o kojem se ovdje govori ne odnosi se samo na govor, već je to univerzalan auditivni problem. Ono što Cherry (1953) naziva *cocktail party* efektom, Bregman (1994) na globalnom auditivnom (akustičkom) planu naziva problemom analize zvučnog okruženja (*auditory scene analysis*).

Da bismo pobliže definirali temu koja nas ovdje zanima, potrebno je uvesti jedan novi pojam (termin). Što mi kao slušatelji zapravo radimo? Mi neprestano iz mješavine zvukova koji nas okružuju izdvajamo i u cjeline povezujemo akustičke elemente koji pripadaju istom izvoru. To je posebna vrsta filtriranja koju klasični (elektro) akustički filtri ne mogu obaviti, osim u sasvim specifičnom slučaju kad zvukovi koje treba odvojiti iz mješavine zapremaju različite, ne-preklapajuće dijelove zvučnog spektra. Opisujući kompleksne zvukove kojima smo okruženi, pa i govorni zvuk, koristimo se nizom termina kao što su: tonovi, šumovi, harmonici, tranzijenti, formanti itd. Ono što nedostaje za opis zvučnog okruženja ili *cocktail party* problema jest izraz koji će označiti skup ili niz zvučnih elemenata koji predstavljaju cjelinu, odnosno, najčešće zvučne elemente koji pripadaju istom izvoru zvuka. Da bi označio pripadnost zvučnih elemenata jednoj cjelini, Bregman (1994) predlaže izraz *stream*, što se najneposrednije može prevesti kao "zvučni tok". Kriterij po kojem nešto predstavlja jedan (isti) zvučni tok ili cjelinu u smislu ove rasprave, isključivo je perceptivan, dakle ne radi se o objektivnom akustičkom nego o psihoakustičkom određenju pojma. Uvedeni pojam, "zvučni tok" ili "zvučna cjelina", ne može se zamijeniti nekim od postojećih. U glazbi jedan zvučni tok može biti jedna melodija (tema), jedna ritmička figura ili zvuk istog instrumenta. U opisu govornog zvuka to može biti jedan vokal, riječ ili rečenica. Načelo se jednako odnosi na cjelokupno auditivno područje, na cjelokupno zvučno okruženje. Zajedničko svim ovim primjerima jest to da mora postojati unutarnja kohezija koja ujedinjuje akustičke elemente.

Proces percepcije, da bi uspješno obavio snalaženje u kompleksnom zvučnom prostoru, u mješavini zvukova, neprestano mora koristiti dva međusobno komplementarna mehanizma: razdvajanje (*segregation*) i spajanje (*integration*) zvučnih tokova. Oba ova mehanizma djeluju i na sukcesivnom i na simultanom planu. Odvajanje i spajanje nazvali smo komplementarnim mehanizmima, a ne međusobno isključivim, zato što djeluju istovremeno i rezultat nije kategorično opredjeljenje nego pretpostavka ponderirana vjerojatnošću. Ovu ideju možemo ilustrirati ishod bilo kojeg slušnog testa: čak i kada su signali nedvosmisleni, u velikoj skupini slušatelja nikad odgovori nisu potpuno podudarni. To potvrđuje navedenu pretpostavku o komplementarnom djelovanju razdvajanja i spajanja u procesu percepcije jer pokazuje da kod nekih ispitanika prevagne integrirajući mehanizam, a kod nekih separirajući.

Načelo objedinjavanja zvučnih elemenata u cjeline često se uspoređuje s osjetom vida. S obzirom na to da je spektrogram zapravo slika zvuka, moglo bi se očekivati da je upravo to medij u kojem se principi ponašanja osjeta na vizualnom planu neposredno poklapaju s analizom osjeta na auditivnom planu. Vizualno područje (kad je o spektrogramu riječ) pokazuje se značajno inferiornim auditivnoj domeni. Zbog toga pokusi u projektu *visible speech* opisani u istoimenoj knjizi (Potter i sur. 1947) koja niz godina predstavlja početnicu za očitavanje spektrograma, nisu doveli do željenih rezultata. Da podsjetimo, izum spektrografa omogućio je da se predloži ideja po kojoj bi se slušno hendikepiranim osobama zvuk (govora) približio pomoću vizualnoga medija. Projekt nije uspio unatoč tome što su ispitanicima predočeni samo, u akustičkom smislu, relativno jasni govorni uzorci. Pravi se problem pojavljuje tek kad govorni uzorci više nisu dostupni u svojem "čistom" obliku nego su, na jedan od bezbroj načina, izobličeni ili maskirani drugim zvukom. Takva je, zapravo, normalna svakodnevna govorna situacija. Govornici ne komuniciraju zatvoreni u laboratorijskim uvjetima bez buke, prisustva drugih zvukova i ne izgovaraju riječi u takozvanom čistom obliku. Unatoč svim tim negativnim učecajima na prijenos govornog signala, govorna je komunikacija začuđujuće uspješna.

GOVOR KAO ORGANIZIRANI SLIJED AKUSTIČKIH ELEMENATA

Govor je organiziran u vremenu "slijeva nadesno", odnosno u govoru je važan redosljed akustičkih elemenata. Brzi slijed različitih vrsta zvukova koji dolaze od pojedinoga govornika moraju se okupljati u jedan tok i ne smiju se istovremeno sekvencijalno povezivati sa zvukovima koji dolaze od drugoga govornika. Redosljed akustičkih elemenata, sekvencijalno povezivanje važno je i na razini pojedinih glasnika. Identifikacija onih glasnika govora kod kojih je vremenski slijed bitan element njihove organizacije (npr. afrikate) ovisi o redosljedu relevantnih informacija koje se pripisuju jednom zvučnom toku. Tako se tišina pred šumom afrikate ne smije interpretirati kao prekid jedinog zvučnog toka jer u tom slučaju ne bi bilo percepcije afrikate. Slušatelj mora tu tišinu interpretirati kao događaj unutar istog zvučnog toka (slijed zvučnih elemenata istog

izvora, govornika). Isto tako, ako jedan govornik prestane a drugi započne govoriti, slučajna tišina koja tako nastane ne smije signalizirati okluziju.

Percepcija redosljeda akustičkih elemenata za razumijevanje govora ima ključno značenje. Uz prosječan tempo artikulacije oko 5 slogova u sekundi (Bakran 1984) fonemi se izmjenjuju tempom bržim od 10 u sekundi, odnosno prosječno im je trajanje manje od 100 ms. Međutim, ako se govor umjetno ubrza do 30 fonema u sekundi (uz zadržavanje iste tonske visine) slušatelji i dalje mogu razumjeti govor bez vremenske konfuzije, odnosno, percipiraju ispravan slijed glasnika. Nasuprot tome, ako se niz u kojem se izmjenjuju niski i visoki tonovi ubrza na sličan način, ono što se u sporom tempu percipira kao jedan zvučni tok, počinje se razdvajati u pod-tokove (*substreams*), tonovi se povezuju u tokove koji imaju sličnu tonsku visinu.

Zanimljivo je da ljudi mogu razumjeti brzu izmjenu glasnika govora unatoč tome što je govor slijed kvalitativno različitih zvukova (frikativ je različiti tip zvuka od vokala), a nisu u stanju percipirati redosljed nepovezanih zvukova (fíćuk, šum, zujanje, vokal). Slijed takvih zvukova mora biti mnogo sporiji od slijeda glasnika govora da bi se ispravno identificirao. Za razumijevanje govora slušatelj mora ispravno odrediti redosljed glasnika jer izmijenjeni redosljed može predstavljati drugu riječ ili ne-riječ. Warren (1982) upozorava da se kod slušanja govora ne mora nužno prepoznati svaki konstitutivni element nego da slušatelji provode vrstu globalne analize, a rezultat uspoređuju s pohranjenim podacima u memoriji. Djeca prepoznaju riječi a da ne moraju biti svjesna od kojih su elemenata one sastavljene. Međutim, ovakvo objašnjenje ne rješava problem odjeljivanja zvučnih tokova u govoru.

Zanimljiv pokus kojim se pokazuje specifična struktura govornog zvuka proveo je Bregman (1994). On je pretpostavio da djeca počinju riječi prepoznavati kao cjeline zato što uočavaju da se isti zvučni tokovi (riječi) pojavljuju u različitim okolinama. Niz riječi bez stanke izgovarao je monotonim glasom tako da su se one ponavljale u različitom redosljedu. Zatim je snimku reproducirao naopako. Na taj je način dobiven potpuno neprepoznatljiv niz zvukova. Nakon jednog sata, uspio je izdvojiti sve cjeline koje su se ponavljale. Potom je istim tempom kojim su se izmjenjivali glasnici govora snimio niz nepovezanih zvukova i isto ih tako grupirao u "riječi". Takvi se nizovi ni nakon dugog slušanja nisu mogli izdvojiti u prepoznatljive cjeline. Prema tome, u govornom zvuku mora postojati unutarnja kohezija koja slušatelju olakšava povezivanje u perceptivne cjeline.

Warren (1982) proučava kako slušatelji određuju redosljed nizova nepovezanih glasnika. Vrlo je teško odrediti redosljed vokalnih segmenata u trajanju od 200 ms izrezanih iz prirodno izgovorenih vokala. Zadatak je lakši ako vokalni elementi traju 150 ms, s 50 ms tišine među segmentima. Još je lakše odrediti redosljed ako to nisu izrezani elementi nego stvarno izgovoreni vokali s vlastitim prirodnim početkom i završetkom. Određivanje redosljeda takvih zvučnih elemenata mnogo je bolje nego određivanje redosljeda raznovrsnih zvukova (zvižduk, šum, zujanje, ton...) u istom tempu, ali je mnogo lošije nego određivanje elemenata brzog povezanoga govora. Superiornost određivanja redosljeda vokala u odnosu na određivanje redosljeda raznovrsnih zvukova može

se djelomično protumačiti korištenjem jezičnih sposobnosti. Moguće je da se niz doživljava (čuje) kao višesložni verbalni izričaj. Uspješnost je prema tome bolja što su elementi niza sličniji prirodnom izgovoru.

S obzirom na to da se u prirodnom govoru izmjenjuju raznovrsni slogovi sastavljeni od kombinacija konsonanata i vokala, može se pretpostaviti da će uvođenje konsonantskih elemenata u pokuse identifikacije redosljeda govornih elemenata olakšati percepciju. Lackner i Goldstein (1974) pokusom su provjerili da li postojanje okluziva /b/ ili /p/ koji unose kratku pauzu i brzi tranzijent među vokalima može olakšati detekciju redosljeda. Iznenađuje rezultat da je uspjeh u detekciji najlošiji kada se izmjenjuju slogovi s konsonantom i slogovi bez konsonanta u nizu CV-V-CV-V. Uspješnost je u takvom slučaju bila ispod razine slučajnosti. Zbog toga su se slogovi s konsonantima odvojili u poseban zvučni tok pa su slušatelji umjesto "de-o-be-a-de-o..." registrirali "de--be--de--..." i "--o--a--o..." kao odvojene, simultano reproducirane tokove.

Noteboom i sur. (1976) sistematski su istraživali djelovanje fundamentalne frekvencije sintetskih vokala na odvajanje tokova. Varirali su F0 od 0 do 20 polutonova, a trajanje segmenata od 100 ms razmaknuto je 0 do 400 ms. Rezultati su pokazali jasan "trade-off" efekt (mijenjanje veličine jedne varijable utječe na efikasnost druge) između frekvencijske različitosti i trajanja tišine između sukcesivnih stimulusa. Ako je pauza između sukcesivnih zvučnih elemenata bila oko 100 ms, razlika njihovih fundamentalnih frekvencija ne smije biti veća od 2-3 polutona kako bi se niz mogao percipirati kao jedan zvučni tok. Uz veći razmak među susjednim zvučnim elementima, frekvencijska razlika može biti veća. Uz razmak od 200 ms kao isti zvučni tok mogli su se percipirati vokalni elementi kojima se frekvencija osnovnog tona razlikovala do 10 polutonova.

ULOGA HARMONIČKIH ODNOSA I F0

Djelovanje tonske visine primijećeno je u istraživanjima selektivne pažnje. Ako se od ispitanika traži da "pokrivaju" jedan od dva simultana govora ("shadowing" - eksperimentalni postupak u kojem ispitanici sa što manjim vremenskim pomakom ponavljaju zadani tekst), oni će tu zadaću jednostavnije obaviti kad su ta dva glasa različite tonske visine (Norman 1976). Uspješnosti zasjenjivanja također pomaže ako su spektri dvaju signala ograničeni na različita frekvencijska područja. U ovakvim i sličnim istraživanjima selektivne pažnje korišten je prirodni povezani govor tako da su akustičke okolnosti bile vrlo kompleksne. Istraživanje koje se opisuje u nastavku mnogo je analitičnije u akustičkom smislu.

Najprije, potrebno je pokazati da je slušni mehanizam u stanju istovremeno postojanje dviju različitih fundamentalnih frekvencija iskoristiti za odvajanje zvučnih tokova. U jednom pokusu Brox i Noteboom (1982) promatrali su kako će ispitanici ponavljati besmislene rečenice koje su reproducirane simultano s čitanjem kontinuiranog teksta jedne priče. U jednom dijelu test-materijala muški je govornik povišenim registrom i normalnim intonacijskim varijacijama izgovarao besmislene rečenice, a u drugom ih je izgovarao svojim

normalnim glasom. Kako se i očekivalo, mnogo je uspješnije bilo ponavljanje besmislenih rečenica kad je interferirajući signal bio bitno različite fundamentalne frekvencije.

Djelovanje tonske visine provjereno je i preciznim kontroliranjem tonske visine digitalnom obradom. Na taj se način isključuje djelovanje intonacije prirodnoga govora, njezinoga kontinuiteta, kao objedinjavajućeg faktora. Ustanovljeno je da se broj pogrešaka smanjuje s povećavanjem razlike u fundamentalnim frekvencijama. Uz razliku frekvencija od tri polutona broj pogrešaka smanji se za 20 %. Međutim, uz razliku od oktave, ponovno se povećava broj pogrešaka zbog toga što se frekvencije harmonika jednog i drugoga glasa poklapaju. Disonantan frekvencijski odnos pogoduje odvajanju tokova.

Korištenje dviju različitih fundamentalnih frekvencija za odvajanje dvaju glasova za slušni mehanizam ozbiljan je problem. Slušni mehanizam u svakom trenutku mora detektirati dvije različite fundamentalne frekvencije da bi registrirao dvije različite tonske visine, usto mora nekako oformiti dva odvojena spektra. Svaki od njih mora sadržavati ne samo harmonike, nego i njihove relativne intenzitete. To je potrebno zato što identitet zvučnih glasnika ovisi o relativnim intenzitetima pojedinih harmonika i o njihovim promjenama u vremenu. Da bi se moglo odlučiti koja su dva vokala prisutna, kompleksan oblik intenzitetskih odnosa pojedinih harmonika mora se razložiti u dva odvojena spektralna oblika.

KONTINUIRANA PRIRODA FREKVENCIJE OSNOVNOG TONA

Već se iz opisanih primjera i pokusa u kojima se manipuliralo frekvencijom tonova, može razabrati da tonska visina ima vrlo važnu ulogu za percepciju zvučnog kontinuiteta. Percepciju tonske visine u govoru omogućuje (kvazi) periodičan karakter zvučnoga govora (detaljnije o teorijama percepcije tonske visine vidjeti Moore 1990). Dijapazon i varijacije tonske visine u govoru nisu proizvoljne veličine. One djelomice ovise o fiziologiji proizvodnje harmoničnoga zvuka, a djelomice o paradigmama jezičnog funkcioniranja. Unatoč tome što osjet tonske visine pobuđuje periodičan, harmoničan zvuk, koji se u govoru vrlo brzo izmjenjuje s neperiodičnim (zvučni i bezzvučni glasnici), osjet tonske visine ne mijenja se naglo, a pogotovo ne u ritmu izmjene zvučnih i bezzvučnih glasnika. Prema tome postoje tri tipa ograničenja u kretanju tonske visine: jedno je relativna sporost promjene, drugo je jezična zadanost oblika promjene, i treće je ograničenje koje predstavljaju individualne karakteristike govornika (dijapazoni varijacija tonske visine kod različitih se govornika međusobno razlikuju).

Važnost intonativnog kontinuiteta pokazuje jednostavan pokus. Ako se iz povezanog govora izdvoje pojedine riječi te se one poredaju u jedan drugi, također smisleni niz, rezultat je teško razumljiv. Slušateljima se čini da zvuk dolazi iz različitih izvora i imaju poteškoća integrirati ga u smislenu cjelinu. Treisman (1960) provela je pokus u kojem je ispitanicima u jedno uho (lijeva slušalica) emitiran jedan tekst, a u drugo uho (desna slušalica) drugi tekst. Imali su zadatak ponavljati samo tekst koji se čuje u jednom (lijevom) uhu. U jednom trenutku

preklopnikom bi se zamijenile strane, tako da tekst koji je stizao u lijevo uho dode u desno, i obratno. Ispitanici bi nakon prebacivanja teksta koji su ponavljali iz lijevog uha u desno, ponovili još nekoliko riječi koje su nastavak istog teksta, bez obzira na to što se promijenilo uho, i tek su se onda ponovno vratili na zadatak, ponavljanju onoga što čuju u lijevom uhu.

Ovom se pokusu može prigovoriti to da se u njemu ne može odvojiti djelovanje niza čimbenika koji održavaju kontinuitet, npr. prostorna i semantička povezanost. U ponovljenom pokusu odvaja se djelovanje semantičke povezanosti i inotacijskog kontinuiteta. U trenutku prebacivanja iz jednoga u drugo uho, jednom je zadržan intonacijski oblik a promijenjen smisao, drugi puta obrnuto. Obje vrste "preklopnika" uzrokovale su pogreške u obavljanju zadatka, no greške nisu bile istoga tipa. Kada je prekinut semantički kontinuum, ispitanici bi zapeli u ponavljanju jer se nastavak ne očekuje na temelju konteksta, ali ne bi ponovili riječi u suprotnom uhu. Međutim, kad je prekinut intonativni kontinuum, ispitanici bi ponovili još koju riječ koja je stigla u suprotno uho jer one riječi koje su stizale u "zadano" uho nisu pripadale istom intonativnom obliku (kontinuitetu).

U pokusima s prirodnim govorom ne mogu se dobro kontrolirati sve dimenzije koje simultano djeluju na objedinjavanje i odvajanje zvučnih tokova. Zato su Darwin i Bethel-Fox (1977) pokusima sa sintetičkim govorom pokazali upravo nevjerovatnu važnost kontinuiteta fundamentalne frekvencije za percepciju slijeda govornih elemenata. Oni su sintetizirali slijed formantata za /uau/ s pripadajućim tranzijentima i nepromijenjenom fundamentalnom frekvencijom (130 Hz). Takav se slijed čuje kao jedan slog. Međutim, ako se formantski oblik zadrži, a fundamentalna se frekvencija promijeni na polovici svakog tranzijenta tako da se dobije slijed: nisko, visoko, nisko, s frekvencijama 110, 170, 110 Hz, potpuno se gubi prijašnji perceptivni dojam te se čuju dva sloga niskog tona i jedan slog visokog tona. Na pitanje koji su to slogovi, odnosno koji glasnik čuju prije /a/, ispitanici su odgovarali /b/, zato jer je promjenom fundamentalne frekvencije na polovici tranzijenta od /u/ do /a/ prekinut zvučni tok i percipira se skraćeni tranzijent koji je nalik na tranzijent od /b/. Ovim se pokusom pokazalo da je diskontinuitet fundamentalne frekvencije odvojio zvučne elemente u različite tokove unatoč kontinuitetu spektralnog oblika.

Kontinuitet frekvencije osnovnog tona može u jedan zvučni tok objединiti dva vokala između kojih je interpoliran bezvučni konsonant. Pitanje je po čemu je taj interpolirani konsonant dio istoga zvučnoga toka, odnosno kako to da spektralni diskontinuitet koji predstavlja interpolirani konsonant ne signalizira drugi zvučni tok. Vjerojatan odgovor na to pitanje jest da se radi o kontinuiranosti spektralnih promjena (vidjeti iduće poglavlje). Prekid zvuka za okluziju, odnosno rubovi toga prekida sadrže spektralni kontinuum. Međutim, u ovom slučaju simultano djeluje spektralni kontinuum koji se manifestira adekvatnim tranzijentima i kontinuum fundamentalne frekvencije. Ekstrapoliramo li rezultate Darwina i Bethell-Foxa, možemo zamisliti što će se u sintetiziranoj riječi čuti ako se razbije kontinuitet toka fundamentalne frekvencije prije i poslije bezvučnog okluziva. moguće je da će se to čuti kao dva različita glasnika, kao što je moguće i da se uopće ne percipira okluziv interpoliran između dvaju vokala. To znači da je

upravo kontinuitet fundamentalne frekvencije zaslužan za percepciju bezvučnog okluziva u takvom okruženju.

FORMANTI KAO FAKTOR OBJEDINJAVANJA ZVUČNIH TOKOVA

Istraživanja u vezi s grupiranjem simultanih komponenata uglavnom se odnose na grupiranje formanata jer oni predstavljaju relevantne akustičke znakove za identitet velikog dijela glasnika. Ima i drugih razloga za odabir formanata za materijal istraživanja: teorija njihova nastajanja prilično je zaokružena, a osim toga, oni predstavljaju jednostavne akustičke elemente pomoću kojih se vokali i okluzivi mogu lako sintetizirati.

Većina znanstvenika koji se bave govorom vjeruju da su formanti smislene perceptivne činjenice i da auditivni sustav slijedi te spektralne vrhove i njihove promjene u vremenu tijekom procesa prepoznavanja govora. Na spektrogramu je relativno lako vizualno razdvojiti formante. S druge strane, nije jednostavno "začuti" pojedini formant kao zasebnu boju. Ima istraživača koji sumnjaju u to da formanti imaju središnju ulogu u percepciji govora. Oni misle da su formanti samo akustička baza za perceptivnu analizu globalnih kvaliteta kao što su kompaktnost, oblik spektra ili brzina spektralne promjene (Stevens i Blumstein 1981).

Broadbent i Ladefoged (1957) prvi su postavili pitanje grupiranja formanata u situaciji simultanog pojavljivanja formanata koji pripadaju različitim glasicima. Pitanje je kako slušni sustav zna koju kombinaciju formanata treba odabrati da bi se oblikovao vokal. Njihovo je objašnjenje u tome da je presudno ritmično pulsiranje u funkciji frekvencije osnovnog tona. Uho grupira one formante koji imaju identično pulsiranje. To su autori zaključili na temelju pokusa u kojem su sintetizirali jednostavnu rečenicu, tako da su prvom i drugom formantom pridružili jednom identičnu F0, a drugi put različite F0. Rezultat je emitiran ispitanicima tako da su prvo oba formanta emitirana u isto uho, a potom u različita uha. Ispitanici su trebali procijeniti čuju li jednog ili dva govornika, i jesu li oni u na istom mjestu u prostoru. Kad je za sintetiziranje obaju formanata upotrebljena identična F0, bez obzira na to jesu li oni emitirani u isto ili u različita uha, ispitanici su čuli da se radi o jednom glasilniku. Oni su, isto tako, čuli da je signal bio u onom uhu u koje je emitiran prvi formant. Nasuprot tome, kad su formanti sintetizirani različitim F0, bez obzira na to je li intonacija bila prirodno modulirana ili monotona, čuli su dva govornika. Posebno je važan dio pokusa u kojem su korištena dva posebna generatora za osnovni ton, svaki za svoj formant, i kad su slijedili isti intonativni oblik. Zbog neznatnih neslaganja sustava, oni su se povremeno razilazili u fazama. Unatoč istoj frekvenciji osnovnog tona, povezivanje formanata u tom slučaju bilo je slabije, a kad bi se oni emitirali u različita uha, čula bi se dva govornika. To pokazuje da je za povezivanje različitih spektralnih dijelova potrebna velika vremenska podudarnost (preciznost), i da nije dovoljno da su harmonici povezani samo s istom F0, frekvencijski, nego F0 mora biti podudarna i u fazi. Ovi su pokusi posebno važni jer iste rezultate pokazuju s ljudskim glasom u rečenicama kao i s negovornim zvukom.

Cutting (1976) je pokusima provjeravao djelovanje frekvencije osnovnog tona u povezivanju formanta emitiranih u različita uha u funkciji fonetske identifikacije. Radilo se o slogovima okluziv-vokal, koji se uvjerljivo mogu sintetizirati sa samo dva formanta. U slučaju kad se sintetizirao slog "da" i kad su oba formanta prezentirana u oba uha, samo je o frekvenciji osnovnog tona pojedinog formanta ovisilo čuje li se jedan ili dva govornika. Razlika u frekvenciji osnovnog tona od samo 2 Hz bila je dovoljna da se čuju dva različita glasa (govornika). Pritom je ispravno identificiran slog "da", a ne nešto drugo. To znači da slušatelji mogu kombinirati informacije iz lijevog i desnog uha da bi se uobličio identitet govornog elementa (sloga).

Postoji iznimka u tom općem ponašanju. Naime, kada postoje različite mogućnosti percepcije ovisno o tome koji se formanti grupiraju zajedno, podudarnost u frekvenciji osnovnog tona pomaže, priklanja se jednoj od mogućih percepcija. Ima slučajeva kad F1, F2 i F3 zajedno daju jedan perceptivni rezultat, a F1, F3 i F4 drugačiji. Ishod percepcije (identifikacije) ovisi o tome koja kombinacija formanta ima zajedničku frekvenciju osnovnog tona. Prema tome, frekvencija osnovnog tona, koja predstavlja važan čimbenik za mehanizam "primitivnoga" grupiranja, u problematičnim slučajevima može pomoći prepoznavanju fonema, premda je ono prvenstveno proces koji se temelji na obrascima (*schema-based*), dakle na naučenom.

KORELIRANE FREKVENCIJSKE PROMJENE

Fundamentalni ton ljudskoga glasa varira u vremenu. Te varijacije, dakle intonativni oblik, istodobno mijenjaju frekvencije harmonika. To znači, ako se osnovnom tonu frekvencija promijeni za 25%, onda se i frekvencije svih harmonika paralelno promijene za 25%. S druge strane, ako dvije osobe govore istovremeno, nije vjerojatno da se njihove intonacije mijenjaju paralelno. Paralelno pomicanje djeluje kao objedinjavajući faktor zvučnog toka, to jest oni harmonici koji se paralelno pomiču po frekvencijskoj skali, pripadaju istom zvučnom toku (glasu, govorniku). Pritom nije jasno da li grupiranje prema tom načelu samo odvaja glasove u tom smislu da su slušatelji svjesni da se radi o različitim glasovima ili takvo grupiranje akustičkih elemenata olakšava ujedno i fonetsku identifikaciju. Ovdje se mogu miješati dva efekta. S jedne strane, paralelno pomicanje harmonika može djelovati tako da ih se grupira u jednu cjelinu. S druge strane, sama činjenica frekvencijskog pomicanja harmonika pridonosi boljem ocrtavanju spektralnog oblika. To postaje osobito važno kad je fundamentalna frekvencija relativno visoka, s razmaknutim harmonicima. Paralelno pomicanje harmonika pojašnjava oblik koji ne mora biti posve definiran frekvencijski stacionarnim harmonicima. Prema tome, ne može se razlučiti koliko na odvajanje zvučnih tokova djeluje sama činjenica paralelnog pomicanja frekvencija harmonika, a koliko to što se zbog pomicanja frekvencija harmonika jasnije ocrta spektralni oblik, što omogućava bolju percepciju vokalne boje.

I kada govornici nastoje tonsku visinu održati stabilnom, frekvencija osnovnog tona neprestano se mijenja. Ta se pojava naziva mikromodulacijom ili

jitter. Pokazalo se da dodavanje malih frekventijskih nepravilnosti sintetiziranom glasu djeluje, ne samo tako da glas zvuči prirodnije, nego i tako da se harmonici bolje integriraju u jedan zvučni tok. Sintetizirani su različiti vokali, svaki svojim osnovnim tonom. Jedan skup stimulusa nije sadržavao nikakve frekventijske nepravilnosti, a drugi skupovi stimulusa varirali su frekventijski, pravilno (*vibrato*) ili nepravilno (*jitter*). Rezultati su pokazali da slušatelji jednostavnije prepoznaju, to jest odvajaju, vokale ako su frekventijski modulirani, bez obzira na to kakve su vrste bile modulacije, pravilne ili nepravilne. Osim toga, u slučaju kad F0 nije bila modulirana, slušateljima je bilo teško procijeniti o kojim se tonskim visinama radi, a kad su stimulusi bili modulirani, čuli su točno koje su to različite tonske visine. U ponovljenim pokusima, u jednom slučaju sintetizirani vokal modulira se tako da se mijenjanjem frekvencije F0 bolje ocrta spektralni oblik (mijenjaju se amplitude harmonika tako da uz promjenu frekvencije slijede spektralni oblik), a u drugom slučaju tako da amplitude harmonika ostaju nepromijenjene s promjenom F0 (pa se prema tomu mijenja spektralni oblik).

U oba slučaja podjednako je olakšana identifikacija vokala. To znači da modulirani vokali nisu jasniji samo zbog detaljnije ocrtanog spektralnog oblika. Vjerojatno je da modulacija snažnije pobuđuje živčani sustav.

Chalikia i Bregman (1989) promatrali su identifikaciju simultano emitiranih parova sintetičkih vokala kojima je F0 sporo modulirana, slično kao govorna intonacija. Pritom je promatran utjecaj razlike (razmaknutosti) fundamentalnih frekvencija i načina njihova mijenjanja. Fundamentalni ton pojedinih vokala za jedan skup stimulusa bio je nepromijenjen, za drugi se mijenjao paralelno, a za treći u suprotnom smjeru. Obje vrste pomaka, i paralelni i suprotni, značajno su pomogli identifikaciji vokala u usporedbi s nepromijenjenom F0. U slučaju kad razmak F0 iznosi upravo jednu oktavu, a to se odnosi na maksimalni razmak koji se kod suprotnih pomaka ostvaruje samo na početku i na kraju, stimulusi s paralelnim pomakom značajno se lošije identificiraju. U načelu, veći frekventijski razmak između dviju fundamentalnih frekvencija pridonosi boljem odvajanju stimulusa. U slučaju kad se F0 mijenja u suprotnim smjerovima, zapravo je u većem dijelu trajanja stimulusa manji frekventijski razmak nego kod parova kod kojih se F0 nije mijenjala. To nije umanjilo uspješnost identifikacije.

KONTINUITET SPEKTRALNIH PROMJENA

S obzirom na to da je ustanovljena neobična važnost kontinuiranosti fundamentalne frekvencije za odjeljivanje i integriranje zvučnih tokova, može se postaviti pitanje ima li kontinuitet spektralnog oblika (frekvencije formanata) slično djelovanje. Kontinuiranost promjena spektralnog oblika na spektrogramima najbolje se može uočiti kad je govorna cjelina sastavljena samo od zvučnih elemenata. U tom je slučaju izvor zvuka neprestano ispod rezonancijskih šupljina i one sve djeluju na modifikaciju spektralnog oblika u skladu s artikulacijskim pokretima. Onda kada se bezvučni glasnici izmjenjuju sa zvučnima, također se može uočiti kontinuiranost spektralnih promjena (formanata). Međutim, kontinuitet toka formanata između zvučnih i bezvučnih glasnika, ne vidi se uvijek.

Kontinuitet formanata može se opaziti, samo ako zvučna struja prolazi istim rezonantnim prostorima. S obzirom na to da je izvor zvuka za zvučne glasnike na razini larinksa, zvuk mora proći cijeli sustav rezonantnih prostora. Šumni se zvuk, međutim, stvara na različitim mjestima govornog trakta. Na primjer, šum za /h/ stvara se prije negoli zračna struja uđe u oralnu šupljinu pa se zbog toga filtrira na sličan način kao vokali tako da su mu i formanti u kontinuumu sa susjednim vokalima. S druge strane, šum za /f/ stvara se između zuba i usana, i ne filtrira se više kroz sve oralne prostore pa zato njegova formantska struktura nije slična (kontinuirana) susjednim glasnica s kojima je u kontaktu. Prema tome, spektralni kontinuitet (formantski kontinuitet) "vidljiv" je i može se pratiti u kontaktu mnogih, ali ipak ne svih glasnika.

Postavlja se pitanje kako mogu govorni segmenti potpuno različitog spektralnog sastava i oblika (šumovi na primjer) biti objedinjeni u isti zvučni tok (s vokalima)? Budući da šum (frikativa) zajedno sa svojom okolinom stvara smislenu cjelinu, jedno od mogućih objašnjenja poziva se na znanje jezika: objedinjavanje u isti zvučni tok postiže se vježbom tijekom učenja jezika. Ipak, tranzijenti predstavljaju suptilne osobine akustičkog slijeda koje omogućuju "primitivnu" analizu i objedinjuju slijed u jedan zvučni tok. Tranzijenti sugeriraju artikulacijski kontinuum. Da se ne radi samo o naučenim oblicima, potvrđuje Bregmanovo iskustvo (Bregman, 1994). On je slušao govornika jezika s klikovima i unatoč tome što nije poznao jezik, točno je identificirao položaj klika unutar zvučnog slijeda. To je bilo moguće jer je klik proizveden artikulacijom, a nije mehanički superponiran govoru na slučajnom mjestu. Izgovoreni klik rezultat je konsonantskog pokreta koji istovremeno modificira spektralne osobine zvučne okoline i ta sinkronost može signalizirati slušnom mehanizmu da su te dvije pojave povezane. Nasuprot tome, Ladefoged i Broadbent (1960) i poslije drugi objavili su da slušatelji nisu u stanju odrediti mjesto slučajno interpoliranog klika u govornom toku. Takvo se ponašanje ispitanika primjećuje i u pokusu s restauracijom fonema. Ako se u govornom toku neki glasnik zamijeni komadićem šuma, slušatelji nadomještaju (u percepciji) zamijenjeni fonem kao da se ništa nije dogodilo, a najčešće nisu svjesni da je fonem bio zamijenjen. Osim toga, ispitanici često nisu u stanju točno odrediti na kojem su mjestu u govornom toku čuli da se pojavio šumni djelić (Erdeljac, 1997).

Bregman (1994) izvještava o pokusu u kojem je promatrano djelovanje spektralnog kontinuiteta u sintetiziranom govoru. Željelo se vidjeti kako formantski tranzijenti pridonose sekvencijalnoj integraciji slogova. Sintetizirani su različiti tipovi niza vokala /o, a, i, u/, svaki s istom fundamentalnom frekvencijom i stalnim frekvencijama prvih triju formanata. Jedan tip sastojao se od dugih vokala trajanja 120 ms bez tranzijenata. Drugi tip uključio je sintetizirani /b/ u sekvenciji CVC (/b/ zbog toga što se u engleskom može jednostavno sintetizirati samo s vokalnim tranzijentima, jer zvučna okluzija nije obavezan akustički znak zvučnih okluziva u engleskom). Kod toga stabilni dio formanata trajao je 30 ms, a tranzijenti po 45 ms. U trećem tipu nizova kratki vokali sintetizirani su tako da je uzet od prethodnog tipa samo stabilni vokalni dio od 30 ms, a tranzijenti su zamijenjeni tišinom. U četvrtom tipu, formanti povezanih vokala neposredno se

nadoveziju. Peti tip nizova može se nazvati pseudoslogovima zbog toga što je stabilnom djelu formanata dodan pomak tranzijentata na više, što ne može rezultirati prepoznatljivim slogovima.

Zadatak ispitanika bio je da ispišu redosljed vokala. Od ponuđenih tipova dva su niza bila u izrazitoj prednosti: onaj s vokalima neposredno spojenim tranzijentima i onaj s CVC slogovima. Lošije se identificirao redosljed vokala u nizovima u kojima oni nisu bili spojeni tranzijentima, a najlošije je prošao tip stimulusa s pseudoslogovima. Kod nepovezanih vokala slušatelji su često čuli dva vokala unutar jednog, a dva unutar drugog zvučnog toka. Pokazalo se također da je najefikasniji u objedinjavanju niza tranzijent najnižeg (prvog) formanta, a da najmanje utjecaja ima treći.

Ovakvo načelo odvajanja različitih zvučnih tokova na temelju suprotnog pomaka frekvencije osnovnog tona ne može se jednostavno primijeniti na frekvencije formanata. Formanti, u ovisnosti o promjenama oblika vokalnog trakta, mijenjaju frekvencije i paralelno i u suprotnim smjerovima, na prvi pogled neovisno, tako da korištenje takva suprotnoga pomaka frekvencija formanata za odvajanje glasova (zvučnih tokova) ne bi funkcioniralo. To da F1 i F2 unutar jedne riječi imaju suprotno kretanje redovita je pojava. Postavlja se pitanje: kako to da ih ne odvojimo i ne čujemo kao posebne zvukove? Jedan od mogućih odgovora na to pitanje mogao bi biti da ih objedinjuje zajednička harmonička struktura. Treba se sjetiti da su formanti zapravo skupine harmonika pojačane rezonancijom. Kad kažemo da se frekvencija formanta kreće prema dolje ili prema gore, mijenja se samo frekvencijsko područje koje će biti pojačano, a ne frekvencije harmonika. Zato, ako F0 raste, frekvencije svih harmonika rastu, a istovremeno, moguće je da frekvencija prvoga formanta (F1) raste, a frekvencija drugoga (F2) da pada. Prema tome, ako se dogodi da dva formanta imaju različitu harmoničku strukturu, oni će se odvojiti u dva zvuka, ali se neće odvojiti u različite zvučne tokove na temelju suprotnog frekvencijskog pomaka.

Slušatelji ne slušaju pojedine formante i nisu ih svjesni kao zasebne boje zvuka, nego njihova kombinacija predstavlja jednu boju. Takvo je slušanje u skladu s načinom nastajanja formanata. Formanti ne nastaju neovisno - mijenja se zapravo oblik rezonancijskog sustava. Pri tome, pojedini dijelovi artikulacijskog trakta, rezonantne šupljine, nisu neovisno zaslužne za frekvencije pojedinih formanata. Zbog toga govornici ne mogu svojom voljom mijenjati frekvenciju pojedinog formanta.

Pokazalo se da poteškoće u integriranju zvučnog toka nastaju i kada se nadovezuju različite širine formanata. Dakle, bez obzira na podudarnost središnjih frekvencija formanata, ako se naglo poveća širina formanta, to se čuje kao povećanje glasnoće u jednoj frekvencijskoj zoni. Isto tako glasnoća umetnutog šuma u govorni niz utječe na to hoće li (šum) biti integriran u zvučni tok ili će biti interpretiran kao neovisan zvučni događaj. To se primijetilo pri pokušaju sinteze konkatenacijom prethodno pripremljenih, spremljenih zvučnih elemenata oblika (Bakran i Lazić, 1998).

VAŽNOST LOKALIZACIJE IZVORA ZVUKA

Vidjeli smo kako je kontinuiranost fundamentalne frekvencije i formanata važna za integriranje govornih nizova. Oba ova kontinuiteta proizlaze iz postupnog mijenjanja položaja (pokretanja) artikulatora i oblika govornog trakta. Vokalni trakt ne može se trenutačno prebaciti iz jednoga položaja, primjerenog proizvodnji nekoga glasnika, u drugi položaj. Jedan drugi oblik kontinuiranosti proizlazi iz činjenice da su govornici u načelu na jednom određenom mjestu u prostoru ili se relativno polagano premještaju kroz prostor.

Položaj izvora zvukova u prostoru i binauralno slušanje nepobitno pripadaju najsnažnijim mehanizmima objedinjavanja i odvajanja zvučnih tokova. Načela lokalizacije odavno su poznata (vidjeti Stevens i Davis, 1960) i o njima ovdje nećemo raspravljati. Želimo samo istaknuti da govor, kao vrsta zvuka, maksimalno koristi sve poznate mehanizme lokalizacije, zato što je širokog spektra, dinamički i spektralno moduliran, pa se nikada ne događa da svi mehanizmi lokalizacije simultano zakažu. Onda kad jedan od mehanizama lokalizacije postaje nedovoljno uspješan zbog vrste zvuka, funkciju prostornog odvajanja preuzima drugi.

Važnost lokaliziranja izvora zvuka za razumljivost govora odavno je uočena u pokusima s maskiranjem. Maskirajući efekt manji je ako se signal (govor) i buka mogu prostorno odijeliti. Pokusom smo ustvrdili da razumljivost govora može porasti do 25% ako se izvori signala i buke razdvoje u prostoru (Bakran, neobjavljen rad). Dakle, unatoč tome što razine intenziteta signala i buke ostaju iste, razumljivost je olakšana. Radi se o tome da kod prostornog razdvajanja u oba uha ne dopire identičan signal. U navedenom pokusu u jednom će uhu signal biti jači, a u drugom će biti jača buka. S obzirom na to da se zvuk u istom uhu lakše integrira u jedan zvučni tok, prostorno odvajanje olakšalo je odvajanje zvučnih tokova.

Kad se sukcesivni elementi kontinuiranoga govora emitiraju pomoću slušalica, tako da se izmjenjuju lijevo i desno uho, sposobnost prepoznavanja pada. Prebacivanje signala u drugo uho djeluje kao tišina u prethodnom uhu, kao manjak signala, i to dovodi do pogrešne segmentacije. Prema tome, ako diskontinuiranost u prostornom smještaju zvuka predstavlja problem integraciji, zaključujemo da podudarnost mjesta olakšava integraciju (Huggins, 1964). Otežano prepoznavanje zbog premještanja (izmjenjivanja) signala od jednog do drugog uha može se primijetiti i na razini fonema. Ako se CV slog /mi/ prezentira tako da se u jednom uhu čuje samo /m/ a u drugom /i/, ispitanici će teško identificirati /m/ zbog toga što je smanjena mogućnost komparacije sukcesivnih spektralnih oblika, pa se ne registrira akustički znak za identifikaciju /m/ koji se sastoji u naglom proširenju spektra na prijelazu od /m/ u /i/.

SIMULTANA ORGANIZACIJA GOVORNOG ZVUKA

Dosad smo promatrali grupiranje i odvajanje u govornom zvuku u slučaju kad zvuk dolazi u različito vrijeme (sukcesivno). Postoje mehanizmi za grupiranje

i odvajanje akustičkih komponenata koje do uha stižu istovremeno. Ti mehanizmi moraju odvajati akustičke elemente različitih glasova (govornika) kako ne bismo čuh zvuk koji je slučajna smjesa različitih glasova. Oni također tako moraju integrirati komponente zvuka koji pripadaju istom glasu (govorniku) tako da se ne bi dogodilo da se kao odvojeni zvukovi (zvučni tokovi) čuju formanti ili harmonici istoga glasa. Pretpostavimo li da auditivni mehanizam najprije oformi nešto nalik neuralnom spektrogramu, kako bi detektirao i prepoznao neki glasnik, on mora objediniti spektralne komponente koje mu pripadaju. Simultane komponente kod negovornog zvuka mogu se odvajati na temelju različitih faktora: tonskom visinom, prostornim smještajem, spektralnim regijama, neovisnošću promjena.

ODNOS "PRIMITIVNIH" MEHANIZAMA PROCESIRANJA I MEHANIZAMA KOJI SU ZASNOVANI NA OBRASCIMA (SCHEMA-BASED)

U cijelom ovom radu pažnja je usredotočena na takozvane primitivne procese grupiranja i odvajanja zvučnih elemenata koji se temelje na svojstvima većine prirodnih zvukova: imaju neko trajanje, sporo se premještaju u prostoru i sastoje se od komponenata koje počinju i završavaju istovremeno. Međutim, ne smije se zanemariti da slušatelji imaju iskustvo i znanje o različitim vrstama zvukova kao što je govor, glazba i raznovrsni svakodnevni poznati zvukovi. Poznavanje zvukova pohranjeno je u obliku obrazaca, *schema* (Bregman, 1994). Svaki obrazac sadrži informacije o nekoj posebnoj pravilnosti koje se manifestiraju na različitim razinama. Govor se u tom kontekstu najčešće navodi kao primjer. Ljudi koji znaju jezik imaju posebne obrasce za pojedine glasnike govora, za pojedine riječi i načine njihova korištenja. Obrasci se aktiviraju kad se u zvuku prepoznaju neki elementi. Aktiviran obrazac potom djeluje na daljnji tijek percepcije. Ako se čuje "whisky", aktiviran je obrazac koji omogućuje slušatelju da u nastavku bude spreman čuti npr. "s ledom" (Erdeljac, 1997).

Za razliku od obrazaca, primitivni procesi djeluju nezavisno od znanja pa čak i neovisno o volji slušatelja. Djelovanje primitivnih mehanizama i ovih "top-down" procesa ne može se uvijek jasno razlikovati. Ima laboratorijskih situacija kojima se može isključiti djelotvornost primitivnih procesa. Ako se sintetiziraju dva vokala iste frekvencije osnovnog tona, istog trajanja i emitiraju se s istog mjesta, nema "primitivnih" elemenata u zvuku koji bi ih razdvojili u odvojene cjeline. Ipak, slušatelji to mogu. Jedino je objašnjenje da su aktivirani obrasci koji su se u tom zvuku prepoznaju. Drugi očit primjer odabiranja, odnosno, perceptivnog ishoda koji se temelji isključivo na aktiviranju obrazaca, nalazimo u pokusima s restauracijom fonema.

Ima eksperimentalnih iskustava u kojima se primitivni procesi sukobljavaju s obrascima. Ako se sintetizira vokal s dva formanta, ali tako da ih sačinjavaju harmonici koji pripadaju različitim fundamentalnim frekvencijama, slušatelji imaju neobičan slušni dojam: čuju dva odvojena zvuka, ali su u stanju čuti i vokal za koji je potreban čitav spektar, oba formanta.

Primitivni mehanizmi sortiraju signal u pojedine elemente, a mehanizmi koji se temelje na obrascima te elemente selektiraju i integriraju. Jedni i drugi procesi nisu osjetljivi na iste varijable. Npr. fundamentalna frekvencija grupe harmonika za primitivne je mehanizme vrlo značajna, a manje je važna za procese koji se temelje na obrascima.

KORELIRANE AMPLITUDNE PROMJENE

U vezi s vremenskom organizacijom zvukova treba spomenuti neke činjenice koje djeluju na odvajanje zvukova različitih izvora. U načelu, neovisni zvukovi ne započinju i ne završavaju istovremeno. Osim toga, često jedan zvuk traje kad se drugi pridodaje mješavini. Prema tome, asinkronija početaka i završetaka relativno je pouzdan znak za odvajanje zvučnih tokova. U jednoj od psihoakustičkih demonstracija (Houtsma i sur. 1987) izmjenjuje se harmoničan ton sastavljen od 10 harmonika s tonom koji sadržava sve te harmonike, osim jednog. Slušatelji u toj izmjeni postanu svjesni zvuka pojedinog harmonika iako njegov zvuk nisu posebno čuli kao sastavni element harmoničnog tona.

U slučaju izmjene vokala i nazala događa se da viši formanti budu naglo utišani (na početku nazala), odnosno pojačani (na početku vokala). To bi trebalo potaknuti odvajanje na taj način da se jedan zvuk čini kontinuiranim, a drugi pridodanim. To zaista možemo introspekcijom osjetiti ponavljamo li sekvencu /mamamama.../. Međutim, taj je efekt razdvajanja znatno manji od efekta objedinjavanja koji uzrokuje harmonički kontinuitet.

Pokusima je ustanovljeno da u slučaju maskiranja jednog vokala drugim, onda kad onaj maskirani počne nekoliko desetinki sekunde nakon maske, vokal se identificira kao da maske nema. Dapače, u tom slučaju efikasnost identifikacije ne ovisi više o razlici fundamentalnih frekvencija dvaju vokala. U prirodnom govoru nagle promjene energije događaju se na granicama okluziva i to okluzive čini otpornima na maskiranje.

Darwin (1984) je proučavao djelovanje (čistog) tona ubačenog u sintetizirani vokalni spektar. Ustanovio je da čak i kad se pridodani ton posve uklapa u harmoničku strukturu, ako nije sinkron s početkom vokala, čuje se kao odvojen zvuk i što je kasnije uključen, manje djeluje na percepciju vokalske boje tako da nakon 250 ms uopće više nema na nju utjecaja.

Ako se sintetiziraju dva odvojena formanta za prijelaz okluziv - vokal, i ako se ti formanti emitiraju svaki u jedno uho, razumljivost je oko 90 %. Međutim, ako nisu sinkroni, razumljivost uz vremenski pomak od 160 ms pada na 50 %. Identifikacija od 50 % i nije tako loša ima li se na umu da je konsonant pretežno definiran vokalnim tranzijentom trajanja otprilike 50 ms. Ako perceptivni proces treba informaciju o oba formanta (F1 i F2), on ih ne može spojiti u jedan osjet, ali može "koordinirati" informaciju jednoga i drugoga. Pokazalo se da asinkronija više otežava koordiniranje informacija pojedinih formanta nego nesklad u frekvenciji osnovnog tona (Cutting, 1976). Na temelju svih ovih primjera možemo zaključiti da sinkronizirane amplitudne promjene različitih

dijelova spektra mogu pridonijeti pravilnom pridruživanju pojedinih spektralnih komponenata istom izvoru.

Harmoničnost se u načelu opisuje kao ekvidistantnost frekvencijskih komponenata, odnosno kao poseban, cjelobrojni odnos između frekvencija pojedinih harmonika i frekvencije osnovnog tona. Međutim, kad se pogleda tzv. "široki" spektrogram, harmonici se (u načelu) ne vide, a harmoničan zvuk prepoznaje se prema okomitim crtama razmaknutim upravo za period osnovnog tona koje sežu do visokog dijela spektra. Time se očrtava sinkronija sastavnih komponenata. Ta sinkronija posebno je važan izvor informacija u visokom dijelu spektra kad se zbog ograničenja auditivnog sustava (širina kritičnog pojasa) pojedini harmonici teško razlučuju. Kod glasnika koji sadrže i harmoničan ton i šum, kao kod zvučnih frikativa, amplitudna modulacija koju uzrokuje periodičnost laringalnih impulsa vidljiva je i u šumnom dijelu spektra i ta sinkronija pomaže objediniti jake niske harmonične sastavne komponente sa šumom.

ZAKLJUČAK

Čini se da slušni mehanizam rješava problem analize zvučnog okruženja na dva različita načina: upotrebom takozvanih primitivnih procesa auditivnoga grupiranja i upravljajući slušanje obrazaca koje se temelje na znanju (jezika), poznavanju zvukova. Ovi se mehanizmi međusobno nadopunjuju.

Integriranje i odvajanje zvukova temelji se na analizi koja smjesu zvukova dijeli na elemente. Nakon toga uključuje se strategija objedinjavanja elemenata koji pripadaju istoj cjelini, i to u vremenskom slijedu (sekvencijalno grupiranje), i u spektralnom smislu (simultano grupiranje). Najočitije je načelo grupiranja po sličnosti. To se odnosi na fundamentalnu frekvenciju, vremensku bliskost, spektralnu sličnost, prostornu bliskost i intenzitetsku sličnost. Jedan od važnih elemenata analize jesu frekvencijski odnosi među sastavnim elementima zvuka. Što su zvučni elementi frekvencijski udaljeniji, manja je vjerojatnost da pripadaju istom zvučnom toku. Drugo je od djelotvornih načela integriranja elemenata analize harmoničnost frekvencijskih komponenata. Auditivni sustav odvojeno grupira harmoničke elemente koji pripadaju istom osnovnom tonu, formira odvojene osjete tonske visine i odvojene spektralne boje. Istovremeno, pojedini harmonici ne čuju se kao zasebni elementi. Sljedeće značajno načelo objedinjavanja elemenata zvučne analize jest korelacija promjena pojedinih zvučnih elemenata. Simultana promjena frekvencijskih i intenzitetskih odnosa signal je za objedinjavanje u istu zvučnu cjelinu. Način na koji se ponaša harmoničan govorni zvuk, za to je idealan primjer. Tu treba imati na umu paralelne frekvencijske promjene harmonika u funkciji govorne intonacije kao i male nehotične, frekvencijske modulacije (*jitter*). Simultane amplitudne promjene u različitim dijelovima spektra pomažu njihovu integriranje. To se opet odnosi na dvije razine: velike spore promjene i one kojima je izvor u periodičnosti harmoničnog zvuka koja se proteže po čitavom spektru i uzrokuje sinkronu neuralnu pobudu. Uz sinkronost treba spomenuti i povezivanje prostorno podudarnih sinkronih elemenata i korelaciju s vizualnim kanalom. Prostorno

povezivanje osim činjenice o različitosti zvuka koji dopire u različita uha, uključuje svijest o relativno sporim promjenama položaja izvora zvuka. Kontinuirani zvukovi lakše se povezuju u isti zvučni tok od diskontinuiranih. Poseban je fenomen koji se povezuje s analizom zvučnog okruženja iluzija kontinuiteta. Ako jedan tihi trajni zvuk mjestimično prekinemo i nadomjestimo mnogo jačim zvukom posve drugih osobina, slušatelji imaju iluziju da tihi zvuk nije prekidan nego da traje "ispod" jačega, odnosno da je samo maskiran. Taj se princip kao istraživačka metoda obilno koristi u pokusima s takozvanom "restauracijom fonema".

U cijelom ovom radu pozornost je usredotočena na takozvane primitivne procese grupiranja i odvajanja zvučnih elemenata koji se temelje na svojstvima većine prirodnih zvukova: imaju neko trajanje, sporo se premještaju u prostoru i sastoje se od komponenata koje počinju i završavaju istovremeno. Međutim, ne smije se zanemariti da slušatelji imaju iskustvo i znanje o različitim vrstama zvukova kao što je govor, glazba i raznovrsni svakodnevni poznati zvukovi. Znanje sadrži informacije o nekoj posebnoj pravilnosti koje se manifestiraju na različitim razinama. Govor se u tom kontekstu najčešće navodi kao primjer. Ljudi koji znaju jezik imaju posebne obrasce za pojedine glasnike govora, za pojedine riječi i načine njihova korištenja. Obrasci se aktiviraju kad se u zvuku prepoznaju neki elementi. Aktivirani obrazac potom djeluje na daljnji tijek percepcije. Primitivni procesi djeluju neovisno o znanju, pa čak i neovisno o volji slušatelja. Djelovanje primitivnih mehanizama i "top-down" procesa ne može se uvijek jasno razdvojiti.

REFERENCIJE

- Bakran, J. (1984). *Vremenska organizacija hrvatskoga standardnog govora*. Disertacija. Filozofski fakultet u Zagrebu.
- Bakran, J. & Lazić, N. (1998). Fonetski problemi difonske sinteze hrvatskoga govora. *Govor XV*, br. 2, 103-116.
- Bregman, A. S. (1994). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.
- Broadbent, D. F. & Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America* 29, 708-710.
- Brox, J. K. L. & Noteboom, S. L. (1982). Intonation and perceptual separation of simultaneous voices. *Journal of Phonetics* 10, 23-36.
- Chalikia, M. H. & Bregman, A. S. (1989). The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation. *Perception & Psychophysics* 46, 487-497.
- Cherry, E. C. (1953). Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America* 25, 975-979.
- Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review* 83, 114-140.

- Darwin, C. J. (1984). Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America* 76, 1636-1647.
- Darwin, C. J. & Bethell-Fox, C. E. (1977). Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance* 3, 665-672.
- Erdeljac, V. (1997). *Prepoznavanje riječi*. Zagreb: SOL, Ibis.
- Huggins, A. W. F. (1964). Distortion of temporal patterns of speech: Interruptions and alternations. *Journal of the Acoustical Society of America* 36, 1055-1065.
- Houtsma, A. J. M., Rossing, T. D. & Wagenaars, W. M. (1987). *Auditory Demonstrations*. Institute for Perception Research, Eindhoven, The Netherlands. (CD s popratnom knjžicom)
- Ladefoged, P. & Broadbent, D. E. (1960). Perception of sequence in auditory events. *Quarterly Journal of Experimental Psychology* 12, 162-160.
- Lackner, J. R. & Goldstein, L. M. (1974). Primary auditory stream segregation of repeated consonant-vowel sequences. *Journal of the Acoustical Society of America* 56, 1651-1652.
- Moore, B. C. J. (1990). *An Introduction to the Psychology of Hearing*. London, San Diego. Academic Press.
- Norman, D. A. (1976). *Memory and Attention: An Introduction to Human Information Processing*. New York: Wiley (citirano prema Bregman 1994).
- Noteboom, S. G., Brokx, J. P. L. & De Rooij, J. J. (1976). Contributions of prosody to speech perception. In: W.J.M. Levelt and G.B. Flores d'Arcais (eds.) *Studies in the Perception of Language*, Chichester: Wiley. (citirano prema Bregman 1994).
- Potter, R. K., Kopp, G. A. & Green, H. C. (1947). *Visible Speech*. Van Nostrand.
- Stevens, K. N. & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In: P.D. Eimas & J. L. Miller (eds.) *Perspective in the Study of Speech*. Hillsdale, N.J.: Erlbaum.
- Stevens, S. S. & Davis, H. (1960). *Hearing, its Psychology and Physiology*. New York: Wiley.
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology* 12, 242-248.
- Warren, R. M. (1982). *Auditory Perception: A New Synthesis*. New York: Pergamon.

Juraj Bakran, Vlasta Erdeljac and Nikolaj Lazić
Faculty of Philosophy, Zagreb
Croatia

AUDITORY SCENE ANALYSIS

The paper discusses the contribution of individual features of the speech signal to the process of perceptual separation of the myriad of simultaneously occurring sounds we are surrounded with. Contributions of a number of authors to this relatively poorly investigated topic are reviewed and discussed related to the results of own research, and some new Croatian terminology is suggested that is essential for understanding the problem.

It seems that the auditory mechanism solves the problem of analysis of the auditory scene in two ways: by the so-called primitive processes of auditory grouping, and by controlling the listening by the schemas that are based on knowledge (of language) and recognition of sounds. These mechanisms are mutually complementary. Integration and segregation of sounds are based on the analysis that segregates the mixture of sounds into elements. Subsequently the strategy of grouping the elements that belong to the same unit is switched on, both temporally (sequential grouping) and spectrally (simultaneous grouping). The most obvious is the principle of grouping according to similarity. This relates to fundamental frequency, temporal proximity, spectral similarity, spatial proximity and intensity similarity. One of the important elements of analysis is the frequency relationship among the sound constituents. The further apart the sound elements with respect to their frequencies, the smaller the probability of their belonging to the same sound stream. The second efficient principle of integrating the elements of analysis is the harmonicity of frequency components. The auditory system groups separately harmonic elements that belong to the same fundamental frequency, forms separate sensations of pitch and separate spectral timbres. At the same time, individual harmonics are not heard as individual elements. The next important principle of grouping the elements of acoustic analysis is the correlation of the changes in the individual acoustic elements. Simultaneous change of the frequency and intensity relations is a signal to group into the same acoustic unit. An ideal example is the way in which a harmonious speech sound behaves. One should keep in mind parallel frequency changes of harmonics depending on speech intonation, as well as small unintentional frequency modulations (jitter). Simultaneous amplitude changes in various parts of the spectrum aid in their integration. This again relates to two levels: large slow changes and those that originate in the periodicity of the harmonic sound, that spreads throughout the entire spectrum and causes synchronous neural excitation. Besides synchronicity one should also mention the association of spatially analogous synchronous elements and the correlation with the visual channel. Spatial association, apart from the fact about the difference in sounds that reach the two ears, includes the awareness of the

relatively slow changes in the location of the sound source. Continuous sounds are more readily associated into the same sound stream than the discontinuous ones.

The entire article is focused on the so-called primitive processes of grouping and separating the sound elements that are based on the properties of most natural sounds: they are of certain duration, they travel slowly in space and they are made of components that begin and end simultaneously. However, one should keep in mind that the listeners have experiences and knowledge about different types of sound, such as speech, music and numerous familiar environmental sounds. Knowledge contains information about a special regularity manifested at various levels. In that context, speech is most often quoted as example. People who are competent in a language have special "schemas" for each speech sound, for words and for their applications. The schemas are activated when some elements are recognized in the sound. The activated schema in turn affects the subsequent process of perception. Primitive processes act independently of knowledge, even independently of the will of the listener. The activity of primitive mechanisms and the "top-down" processes cannot always be clearly distinguished.

Key words: speech sound, sound stream, auditory scene, psychoacoustics