

# Neka istraživanja auditivnog pamćenja i percepcije

---

**Fulgosi, Ante**

**Other document types / Ostale vrste dokumenata**

*Publication year / Godina izdavanja:* **2000**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:257:862492>

*Rights / Prava:* [Attribution-NonCommercial-NoDerivatives 4.0 International/Imenovanje-Nekomercijalno-Bez prerada 4.0 međunarodna](#)

*Download date / Datum preuzimanja:* **2025-03-13**



*Repository / Repozitorij:*

[SUVAG Polyclinic Repository](#)

**ANTE FULGOSI**

**NEKA ISTRAŽIVANJA AUDITIVNOG PAMĆENJA I PERCEPCIJE**



Poliklinika SUVAG, Zagreb

Predavanje održano u Centru za rehabilitaciju sluha i govora

Centar SUVAG – za rehabilitaciju slušanja i govora, Zagreb 1973.

*Pretisak*

*Poliklinika SUVAG, Zagreb,*

*2000.*

## **Auditivno i vizuelno pamćenje**

Neka novija istraživanja pokazuju da, iako postoje razlike između vizuelnog i auditivnog kratkoročnog pamćenja, između njih postoji i jedna vrlo važna interakcija. Ikoničko /časovito/ vizuelno pamćenje kako je nazvao Neisser /1967./ može perzistirati samo jedno vrlo kratko vrijeme /oko jedne sekunde/. Ako se želi zadržati ono što postoji u takvom ikoničkom pamćenju kroz neki duži period, taj sadržaj mora biti rekodiran i zadržan u jednom drugom mediju koji omogućuje duže pamćenje. Ukoliko do takvog rekodiranja ne dođe, sadržaj ikoničkog vizuelnog pamćenja nestaje. Ovdje postoji jedan interesantan odnos između retencije nekih vizuelnih obrazaca. Smatra se da oni vizuelni obrasci, koji su dobro asociirani s nekim poznatim slušnim oznakama /imenima/ mogu biti zadržani u pamćenju na taj način što bivaju rekodirani ili zamijenjeni tim auditivnim oznakama. Na taj način medijum u kojem dolazi do trajne retencije je auditivni medijum, a ne vizuelni. Mnogi su eksperimenti ukazali na takvu mogućnost. Tu u prvom redu mislimo na eksperimente u kojima je ispitivan fenomen akustičke zamjene ili akustičkog brkanja /acoustic confusability/. U tim pokusima utvrđeno je da kod pamćenja slova ili brojeva koji su prezentirani vizuelno dolazi do zamjenjivanja i brkanja zavisno od toga koliko su ta slova ili brojevi akustički, a ne vizuelno, slični jedan drugome. Takve su eksperimente izveli Conrad /1964/, Sperling i Spealman /1970/, Wickelgren /1966/ i drugi.

Smatra se da je trajnije pamćenje auditivno rekodiranog vizuelnog sadržaja moguće zbog toga što je kod kratkoročnog auditivnog pamćenja lakše ponavljanje, nego li kod vizuelnog kratkoročnog pamćenja.

Ostaje nejasno kako je moguće pamtiti vizuelne sadržaje duže od ikoničkog intervala onda kada ti sadržaji nemaju odgovarajuće verbalne oznake. Noton i Stark /1971/ smatraju da takvi sadržaji bivaju zadržani u obliku verbalnih diskripcija. Međutim, ne izgleda vjerojatno da bi svaki vizuelni sadržaj koji se pamti duže od ikoničkog intervala morao biti preveden u takav oblik. Ispitivanja pokazuju da se vrlo kompleksni vizuelni oblici, koji se nedaju svesti na izdeferencirane verbalne deskripcije, mogu vrlo dobro pamtiti. /Nickerson, 1965, Shepart 1967/.

Mnogi istraživači slažu se da je pamćenje vizuelnih podataka za koje postoje verbalne oznake /imena, nazivi i sl./ drugačije od pamćenja vizuelnih sadržaja za koje takvih oznaka nema. Ako je to točno, onda bi prvi vizuelni sadržaji trebali u toku pamćenja pokazivati sličan tok i karakteristike pamćenja koje imaju auditivni sadržaji u toku kratkoročnog zadržavanja,

dok bi oni drugi trebali pokazivati u toku pamćenja promjene koje su karakteristične za vizuelnu retenciju. Rezultati nekih ispitivanja, koja su obavljena u vezi s tom hipotezom, nisu dala potvrdne rezultate /Nickerson i Pew, 1972, Snodgrass, 1971/.

### **Vremenska dimenzija auditivne percepcije**

Perceptivni proces općenito karakteriziraju operacije prepoznavanja ili identifikacije. Te operacije imaju svoju vremensku dimenziju. Prepoznavanje nekog podražaja zahtijeva dva procesa: analizu i sintezu informacije koju podražaj sadrži u nekom senzornom obliku. Pretpostavlja se da na području vizuelnog pamćenja vidne predodžbe koje su stvorene fiksiranjem, drže informaciju dotle dok traje proces prepoznavanja. Nasuprot tome, kod slušanja podražaj se kontinuirano mijenja, a informacija koju on nosi mora ipak na neki način biti dostupna analizi i sintezi, ukoliko želimo da bude identificirana. Zbog toga Massaro /1972/ pretpostavlja da mora postojati jedna “auditivna predperceptivna predodžba” koja traje izvjesno vrijeme i osigurava takvu analizu i sintezu. Karakteristike takvih slušnih predodžaba bile bi drugačije od karakteristika vidnih predodžaba. Kod vidnih predodžbi najvažnija osobina bio bi prostorni raspored, dok bi kod slušnih predodžbi najvažnija osobina senkvencijalni raspored.

Usprkos takvih razlika postoje neke sličnosti u vremenskim tokovima slušne i vidne percepcije.

Katkada su psihofizičke karakteristike jedne i druge percepcije sasvim različite. Tako se smatra da na području sluha nema ničeg što bi bilo slično miješanju boja kod vida. Međutim, kao i kod vidne percepcije, tako i kod slušne percepcije, odnosno podražaja, moraju postojati distinktivne karakteristike koje omogućuju prepoznavanje rasporeda. Kod slušne percepcije, odnosno slušnog podražaja, te bi distinktivne karakteristike trebale biti određene mikrostrukturom slušnog podražaja, tj. modulacijama pritiska u vremenu i to u vrlo kratkim vremenskim intervalima.

Međutim, te distinktivne karakteristike ne mogu biti prepoznane istog časa kad su prisutne, jer bi percepcija morala biti istovremena podražaju. Nasuprot tome, pretpostavlja se, da su vremenske jedinice slušnog podražaja uskladištene kroz kratko vrijeme u predperceptivnom skladištu koje, na taj način, omogućuje njihovu obradu. Postojanje predperceptivne faze /skladišta/ dobro utvrđeno na području vida /Averbach i Coriell, 1961, Haber, 1969, Kahneman 1968, Neisser, 1967, Sperling 1960, 1967/.

## **Neke opće karakteristike predperceptivne predodžbe**

Iz Geštaltističke psihologije je poznato da u percepciji vizuelnih informacija važnu ulogu igraju forma i raspored. Slično bi trebalo biti i kod predperceptivnog uskladištenja na području sluha. Fluktuacije zvučnog pritiska koje su vremenski organizirane određivale bi karakteristične i distinktivne aspekte podražaja odnosno auditivne percepcije.

U vezi s tim postoji jedan veći broj pitanja koja su od važnosti za slušnu percepciju. Navest ćemo neka:

1/ Koje varijable podražaja sačinjavaju ili utječu na formiranje perceptivnih jedinica slušne informacije?

2/ Ako postoji takva perceptivna jedinica na području sluha, kolika je njena veličina?

3/ Kakva je perceptivna jedinica kod govora? Da li je to fonem, slog ili riječ?

4/ Ako jedinice slušne informacije zaista perzistiraju u sistemu predperceptivnog uskladištenja, koliko je njihovo trajanje tj. trajanje auditivne predodžbe?

5/ Koliko su stabilne i otporne te predperceptivne predodžbe na podražaje koji su im prethodili ili koji ih slijede?

6/ Da li je obrada slušnih informacija paralelan ili senkvencijalan proces?

7/ Da li su perceptivne jedinice određene i nekim unutarnjim determinantama kako nam to govori teorija psihološkog časa /Stroud, 1955/ ili su determinirane samo karakteristikama podražaja?

Ova i slična pitanja i problemi bili su predmetom brojnih istraživanja.

## **Maskiranje auditivnih podražaja i detekcija**

Postoji veliki broj istraživanja koja su se bavila maskiranjem i detekcijom slušnih podražaja /Deathpage i Evans, 1969, Elliot, 1962, 1967, Homick, Elfner i Boothe, 1969, Raab, 1963, Samojlova, 1959/.

Elliot /1967/ je koristio eksperiment maskiranja unatrag da bi proučavao razvoj konture slušne frekvencije u funkciji intervala maskiranja i dužine trajanja podražaja. Test-tonovi trajali su između 10 i 100 milisekundi. Ti su tonovi bili maskirani šumom sličnog trajanja i intenziteta od 75 dB SPL. Izračunavao je pragove detektiranja za intervale maskiranja od 10 do 100 ms.

Što je interval bio duži, to je maskirajući efekat bio kraći. Takav je efekat bio sasvim neznan onda kad je šum dolazio tek 100 ms iza tona. Ovi su rezultati potvrđeni i u novijim studijama.

Rezultati ovih istraživanja pokazuju da ako se ton ili šum maskiranja pojavljuje manje od 100 ms iza test-ona, prag detektiranja test-ona je povećan. Nadalje, prag detektiranja je inverzno povezan sa intervalom trajanja tišine između test-ona i maskirajućeg tona ili šuma. Ovi podaci govore bi, dakle, u prilog postojanju predperceptivnog slušnog uskladištenja koje je neophodno za detekciju slušnog signala. Međutim, ima i drugih, jednostavnijih, interpretacija ovakvih rezultata.

Jedna takva interpretacija jeste da bi podražaj maskiranja koji je zadan iza test-ona mogao dostizati test – podražaj u toku njegovog putovanja slušnim živcem i stizati istovremeno u slušni centar. Na taj način snizio bi se omjer signali: šum, pa bi slabija ili smanjena detekcija mogla biti tako objašnjena.

Zbog toga je interesantno pitanje kako glasnoća podražaja utječe na vrijeme potrebno za detektiranje test-ona. McGill /1961, 1963/ je pokazao da postoji inverzni odnos između glasnoće i jednostavnog vremena reakcije. Ako, poput Sternberga /1969/ i Dondersa pretpostavimo da vrijeme jednostavne reakcije na ton uključuje tri faze: fazu detekcije, fazu odluke i fazu odgovaranja, onda bi intenzitet podražaja trebao utjecati samo na komponentu detekcije. Zbog toga vrijeme reakcije predstavlja smisleni način istraživanja odnosa efikasnosti detekcije i intenziteta. McGill /1961/ je ispitanicima prezentirao tonove od 1000 Hz koji su slučajno varirali s obzirom na intenzitet. Pronašao je da, kad je amplituda porasla od 30 na 100 dB, da se medijalno vrijeme reagiranja spustilo od 216 na 120 ms. To znači da je vrijeme detekcije tona smanjeno za 100 ms, a to je upravo vrijednost koja je utvrđena kao efektivno vrijeme za maskiranje unatrag.

Ako maskirajući podražaj dostiže i prešire test-ton u slušnim putevima, onda razlika u vremenu reagiranja na test-ton i maskirajući šum mora biti jednaka vremenskom toku maskiranja unatrag.

Homick i dr. /1969/ je ispitivao maskiranje tona u funkciji intenziteta maskirajućeg šuma. Rezultati pokazuju da povećanje intenziteta šuma sa 70 na 90 dB utječe na vremenski tok maskiranja. Maskirni šum od 90 dB imao je istu efikasnost kao i maskirni šum od 70 dB, ako je bio odgođen za 20 ms. Npr., ispitanik je mogao detektirati jedan test – signal ako je prije

šuma od 70 dB postojao period od 30 ms. Ako je šum iznosio 90 dB, ispitaniku je bio potreban interval od 50 ms da bi detektirao test – signal.

Moglo bi se dakle reći, da je za šum od 70 dB potrebno 20 ms više nego li za šum od 90 dB da bi bio detektiran. Vremena reakcije koje je utvrdio McGill na tonove upravo takvog intenziteta to potvrđuje. Njegovi su ispitanici trebali upravo 20 ms više vremena da bi reagirali na 70 dB nego li na 90 dB. Očigledno, rezultati pokazuju da se maskiranje događa zbog toga što drugi podražaj /maskirajući/ dostiže prvi /test – signal/ i na taj način smanjuje omjer signal:šum. Prema tome, pokusi s maskiranjem unatrag mogu se objasniti bez pretpostavljanja postojanja predperceptivnog uskladištenja.

Što je s maskiranjem unaprijed? Fenomen maskiranja unaprijed također je vrlo dobro dokazan /Deatherage i Evans, 1969, Elliot, 1962, Homick i dr 1969/. Neki smatraju da je taj fenomen maskiranja unaprijed slabiji ili manji od fenomena maskiranja unatrag. Homick i dr /1969/ nalazi istu količinu maskiranja kod jednog i drugog fenomena.

Kod maskiranja unaprijed otpada naravno tumačenje fenomena na osnovu dostizanja, jer je glasniji podražaj zadan prvi. Taj je podražaj već završen prije zadavanja test – signala. Znači, maskiranje unaprijed ne bi smjelo biti ukoliko nema auditivno predodžbe maskirajućeg signala koja traje duže od trajanja samog signala i koja omogućuje detektiranje test signala. Prema tome ako postoji maskiranje unaprijed onda se ono može objasniti **samo** na osnovu postojanja auditivne predperceptivne predodžbe koja nadživljuje sam auditivni podražaj. Rezultati brojnih eksperimenata ukazuju zaista na postojanje takve predodžbe i pokazuju da je takva predodžba, koja nadživljuje sam podražaj, po svojoj prirodi jednaka samom podražaju.

Dakle, maskiranje unaprijed objašnjava se smanjenjem omjera signal: šum zbog postojanja auditivne predodžbe.

Međutim, Erikson i Johnson /1964/ su pokazali postojanje predperceptivnog ehoičnog pamćenja u trajanju do 10 sekundi /Bryden, 1971, Neisser, 1967/. Erikson i Johnson su prezentirali ispitanicima jedan jedva čujni ton dok su oni bili zabavljeni čitanjem štiva. Svjetlo je bilo ugašeno ili simultano s tonom ili u zaostatku od 1-10 sekundi nakon tona. Svjetlo je gašeno i onda kada nije prezentiran ton. Čim je svjetlo nestalo, ispitanik se trebao izjasniti da li je neki ton bio prezentiran unatrag u intervalu od 10 do 15 sekundi ili ne. Rezultati su pokazali da postotak točnih odgovora o postojanju ili nepostojanju tona opada sa dužinom intervala. Međutim, i kod 10 sek. zaostatka ispitanici su mogli dosta pouzdano



odrediti postojanje ili nepostojanje tona. Iz toga se zaključuje na postojanje ehoičkog pamćenja.

Međutim, eksperiment nije sasvim čist, a moguća je i interpretacija da su kod gašenja svjetla ispitanici tražili predodžbu. Zbog toga možemo zaključiti da što se tiče maskiranja detekcije unaprijed kao i unatrag ispitivanja ne potvrđuju postojanje predperceptivnog auditivnog uskladištenja ili pamćenja.

### **Maskiranje slušnog prepoznavanja**

Eksperimenti maskiranja prepoznavanja stvoreni su sa svrhom da se ispituju kvantitativne karakteristike sistema uskladištenja slušne formacije i vremenskog toka obrade takove slušne informacije /Massaro, 1970, 1971, 1972/.

Eksperiment prepoznavanja sastoji se u slijedećem. Ispitanik najprije nauči da identificira ili prepozna dva ili više zvučnih signala. Recimo dva kratka tona različitih visina moraju biti prepoznata po visini. Ako se radi s maskiranjem unatrag, prezentira se najprije prvi ton, zatim dolazi prazni interval, pa onda ton maskiranja.

Test-ton i ton maskiranja su jednakog intenziteta, tako da ovaj drugi ne može dostići prvog kao što je to slučaj s maskiranjem detekcije.

Massaro /1970/ je izveo slijedeći eksperiment. Ispitanik je trebao identificirati tonove od 770 i 870 Hz koji su trajali 20 ms. Maskirajući ton je imao 820 Hz. Svi tonovi su imali intenzitet od 81 dB. Tihi interval je trajao 0,20, 40,80,160,250,350 i 500 ms. Ton maskiranja je trajao 500 ms. Rezultati su prikazani na slici 1.

Iz krivulje se vidi da je poboljšanje identifikacije do 250 ms, a da dalje produžavanje tihog intervala više ne utječe na efikasnost prepoznavanja.

Interpretacija ovih rezultata je slijedeća. Nakon test-tona od 20 ms ostaje predperceptivna predodžba koja služi za obradu informacije i prepoznavanja. Ako je ta predperceptivna predodžba ometena prebrzim javljanjem maskirajućeg tona, rekognicija je nemoguća ili slabija. Rezultati dakle pokazuju da novi ton, maskirajući, završava obradu informacije sadržane u predodžbi i onemogućava identifikaciju prvog tona. Budući da utjecaj traje do 250 ms, to je opravdano pretpostaviti da je to vrijeme trajanja ili nestajanja takve predodžbe. To ujedno pokazuje da je ispitaniku potrebno najmanje 250 ms nakon 20 ms podražaja da bi taj podražaj prepoznao.

Usporedbe maskiranja detekcije unatrag i maskiranja prepoznavanja pokazuju da tu djeluju različiti procesi.

Elliot /1967/ i Homick i dr. /1969/ su pokazali da količina maskiranja detekcije unatrag ovisi o frekvencijskoj sličnosti tonova. Maskiranje je to veće što je sličnost tonova veća. To je u skladu sa spomenutom interpretacijom u smislu omjera signal: šum. Slični je slučaj i kod siltultanog maskiranja.

Kod maskiranja prepoznavanja takav efekt frekvencijske sličnosti nije utvrđen. Sličnost frekvencija tonova uopće ne utječe na efikasnost maskiranja prepoznavanja. Taj je rezultat u skladu s interpretacijom da kod prepoznavanja maskirajući ton prekida perceptivnu obradu slušne predodžbe test-tona. Nadalje, utvrđeno je da ako se maskirajući ton prezentira kontralateralno od test-tona, onda kod prepoznavanja ne dolazi do smanjenja maskirajućeg efekta. To pokazuje da je predperceptivna slušna predodžba lokalizirana na centralnom a ne na perifernom nivou.

Međutim, kod dihotičke stimulacije i maskiranja detekcije, detekcija je gotovo onemogućena.

Nadalje, iako maskiranje unaprijed smanjuje detekciju, Massaro /1970/ nije našao takvog efekta kod maskiranja prepoznavanja unaprijed. Toga nema, jer predodžba koju je stvorio test-ton, zadan nakon maskirajućeg tona, ukida predodžbu maskirajućeg tona. To je zbog njihovog jednakog intenziteta kod maskiranja, dočim kod detektiranja postoji uvijek razlika u intenzitetu.

Zbog svega toga zaključuje se da je maskiranje detekcije usko povezano uz psiho-fizičke procese, kao što je to uostalom slučaj i kod simultanog maskiranja, dok je auditivno prepoznavanje povezano s vremenskim tokom perceptivne obrade.

Ako želimo sumirati rezultate ovih eksperimenata, onda bismo mogli reći slijedeće. Eksperiment s maskiranjem prepoznavanja pokazuje da kratki ton ostavlja slušnu predodžbu koja brzo slabi i nestaje. Ta se predodžba koristi za identifikaciju samog tona.

Ostaju dva pitanja:

1/ Kakva je priroda predodžbe koja ostaje nakon prezentacije kratkog slušnog podražaja?

2/ Kako maskirajući podražaj interferira s perceptivnom obradom te slušne predodžbe?

I ovdje postoje brojni pokusi. Efron /1970, A,B,C,/ upotrebljavajući eksperimentalnu paradigmu Sperlinga /1967/, Habera i Standinga /1969/, analizirao je perceptivno trajanje tonova i šumova. Zadatak je ispitanika bio da prosudi da li se je drugi podražaj /zadan na nekom drugom modalitetu – svjetlo/ dogodio prije slušnog podražaja. Interval između jednog i drugog podražaja mijenja se sve dotle dok ispitanik ne počinje doživljavati pojavljivanje i nestajanje jednog i drugog podražaja istovremeno. Nezavisna varijabla je trajanje prvog, slušnog podražaja. Rezultati pokazuju da minimalna percepcija slušnog podražaja iznosi 130 ms.

Ako je trajanje prvog slušnog podražaja manje, njegova percipirana dužina raste. Recimo, ako je prvi šum trajao 30 ms, ispitanik neće percipirati vremenski interval između kraja tog podražaja i početka indeksnog podražaja ukoliko taj interval ne iznosi 100 ms.

Prema tome i ovi rezultati govore u prilog postojanja predperceptivne slušne predodžbe.

Ta predodžba slična je predodžbi koja je nastala za vrijeme slušnog podražaja, kao što je to slučaj na vidnom području s paslikom. Plomp /1964/ je proučavao brzinu nestajanja slušnog osjeta. Prezentiran je šum od 200 ms i odmah nakon njega drugi podražaj. Ukoliko nakon prvog podražaja ostaje paslika koja nestaje s vremenom, ispitanik ne bi smio primijetiti prazni interval. Dužina tog praznog intervala, koji ostaje nezamijećen, morala bi biti veća što je veća intenzitetna razlika između njih. Plomp je pokazao da se interval od 2,6 ms zapaža u 75% slučajeva između dva šuma od 65 dB. Nasuprot, ako je drugi šum 15 dB slabiji, interval mora biti povećan na 78 ms, da bi u 75% slučajeva bio zapažen.

Neke sovjetske studije također pokazuju postojanje slušne paslike.

Goldburt /1961/ je pokazao da drugi ton skraćuje percipirano trajanje prvog tona. Naime, što je tihi interval između dva tona duži, duža je i percepcija prvog tona. Drugi ton, dakle, može interferirati s percipiranim trajanjem prvog. On je pokazao da je efekat drugog tona na percipiranu dužinu prvog manji što je trajanje prvog tona veće. Slično imamo i na vidnom području.

I ovi su rezultati u skladu s koncepcijom o postojanju predperceptivne predodžbe. Ako perceptivna obrada predodžbe traje 250 ms, onda nema potrebe da ta predodžba perzistira ukoliko je samo trajanje tona bilo dovoljno dugačko. Obrada informacije sa samog tona, ukoliko je omogućena, dovoljna je da eliminira svaku pasliku.

Von Bekesy /1971/ simulirao je inhibiciju unatrag u koncertnim dvoranama. Prezentirao je ton od 1000 Hz u trajanju od 35 ms iz jednog izvora i ton od 1500 Hz iz nekoliko izvora postavljenih oko prvog. Ako je drugi ton bio prezentiran 60 ms iza prvog, on je reducirao percipirano trajanje i glasnoću prvog tona. Bekesy to objašnjava pretpostavkom da proces u mozgu, započet prvim tonom, traje oko 200 ms i da se inhibicija unatrag može dogoditi u tom vremenskom intervalu.

Svi ovi pokusi pokazuju da se slušna predodžba koja ostaje nakon kratkotrajnog tona ne razlikuje značajno od predodžbe koja ostaje ili koja nastaje za samog trajanja tona. Ispitanik stoga papredodžbu ili pasliku ne razlikuje od predodžbe koja nastaje za vrijeme trajanja samog podražaja i precjenjuje trajanje samog podražaja.

Massaro /1972/ je dobio slične rezultate na području prepoznavanja tonova. On je ispitivao što utječe više na prepoznavanje da li trajanje test – podražaja ili vrijeme obrade. Vrijeme trajanja podražaja je olakšavalo identifikaciju, ali je najkritičnije bilo vrijeme trajanja obrade. Pokazalo se da ako se vrijeme obrade drži konstantnim, povećanje trajanja samog tona ima samo neznatan efekat na prepoznavanje. Međutim, povećanje vremena za obradu značajno je poboljšavalo identifikaciju nezavisno od trajanja test-tona.

Kahn i Massaro /rukopis/ su pokazali da paslika mora biti slična test-tonu. Naime, može se logički pretpostaviti da maskirajući ton djeluje na prepoznavanje test-tona negativno, jednostavno zbog toga što odvlači pažnju ispitanika, pa je zbog toga prepoznavanje slabije. Ako je to tako, onda bi moralo biti svejedno što je ispitaniku prezentirano nakon test-tona – neki drugi ton ili neki sasvim drugačiji podražaj. Oni su upotrijebili i tonove i svjetlo. Kad je za maskiranje upotrebljavan ton, onda je prepoznavanje test-tona bilo to bolje, što je tihi interval između dva tona bio duži. S druge strane, kad je za maskiranje bilo upotrebljeno

svjetlo, onda takve zavisnosti nije bilo. Ovo pokazuje da slušna predodžba ima iste karakteristike kao i sam slušni podražaj i da druga slušna informacija interferira zbog toga s njom.

### **Vrijeme trajanja perceptivne obrade**

Vrijeme perceptivne obrade utvrđuje se minimalnim trajanjem podražaja. Warenn, Obusek, Farmer, Warenn /1969/ su pokazali da je ispitanicima bilo potrebno najmanje 300 ms za svaki od 4 neverbalna slušna podražaja pa da bi mogli utvrditi vremenski redosljed u kojem su se ti podražaji pojavljivali jedan za drugim. Utvrdili su, također da neiskusni ispitanici trebaju više vremena, aliiskusni ne manje od 300 ms.

Warren i Warren /1970/ su utvrdili da se četiri repetirana broja ili četiri repetirana samoglasnika /Thomas, Hill, Carrol i Bienvenido, 1970/ mogu prepoznati za 125 do 200 ms. To sugerira da postoji kvalitativna razlika u percipiranju govornog i negovornog sadržaja. Massaro /1972/ smatra da, kada bi se familijarnost sadržaja kojeg treba percipirati držala konstantnom, da bi te razlike nestale.

Vremena perceptivne obrade govornih elemenata mogu se utvrditi trajanjem samoglasnika u normalnom govoru. To je trajanje između 150 ms i 350 ms /Fletcher, 1953, House, 1961/. Budući da na fundamentalnoj frekvenciji, to je zbog toga, za takve govorne elemente, perceptivna obrada produžena /osobito ako nema tihog intervala ispred njih/. Gray /1942/ je utvrdio da je potrebno vrijeme za identifikaciju govornih elemenata znatno kraće ako je retroaktivni interval tih.

Massaro /1972/ je upotrebljavao kratke samoglasnike u eksperimentu prepoznavanja. Izgovoreni samoglasnici najprije su registrirani na istoj fundamentalnoj frekvenciji i s jednakom amplitudom. Jedan segment svakog samoglasnika bio je uz pomoć analogno – digitalnog kovertera, kojim je upravljao kompjuter, pretvoren u digitalnu formu. Za vrijeme pokusa taj je segment prezentiran uz pomoć digitalnog-analognog konvertera u trajanju od 20 ms. Upotrebljen je samoglasnik “i” u engleskoj riječi “heat” i samoglasnik “i” u engleskoj riječi “hit”. Maskirajući podražaj je bio jedan samoglasnik u trajanju od 270 ms. To je bio besmisleni samoglasnik, učinjen kombinacijom dvaju alternativnih segmenata uzetih od samoglasnika “e” u engleskoj riječi “hat” i samoglasnika “u” u engleskoj riječi “put” i trajao je 90 ms. Ispitanici su tri dana vježbali identifikaciju test – samoglasnika prije početka samog eksperimenta.

Rezultati eksperimenta prikazani su na slici 2.

Kako se iz slike 2. vidi, što je veći međusamoglasnički interval, to je veća vjerojatnost identifikacije i prepoznavanja. To pokazuje da je vrijeme obrade i ovdje kritično za identifikaciju samoglasnika. Međutim, ovdje je porast identifikacije u funkciji međusamoglasničkog intervala brži nego li kod identifikacije čistih tonova /slika 1/. **To znači da su dva samoglasnika više međusobno distinktivna nego li tonovi od 880 i 770 Hz, a to pokazuje da samoglasnici kao elementi govora mogu biti identificirani mnogo brže nego li tonovi i da su manje osjetljivi na maskiranje.**

Ovi eksperimenti pokazuju vremena perceptivne obrade za različite sadržaje, ali se ta vremena dosta slažu. Eksperimenti maskiranja prepoznavanja pokazuju da vrijeme perceptivne obrade traje između 120 i 250 ms. Identifikacija četiri slušna podražaja pokazuje da za svakog od njih, zavisno o njihovom karakteru vrijeme perceptivne obrade iznosi između 120 i 300 ms.

Postoji jedan prigovor da su se u ovim eksperimentima koristili prejednostavni materijali, ali taj prigovor otpada kao što ćemo vidjeti kasnije, u vezi s perceptivnom obradom govora.

Procjene vremena perceptivne obrade auditivnih informacija pokazuju se na taj način jednake procjenama trajanja predperceptivne slušne predodžbe. Sve to pokazuje da su prvih 100 do 250 ms trajanja podražaja kritički za proces prepoznavanja. Ukoliko je podražaj kraći, on će proizvesti slušnu predodžbu. To ujedno i pokazuje da obrada vrlo kratkih podražaja traje i za vrijeme tihog retroaktivnog intervala.

Eksperimenti maskiranja prepoznavanja pokazuju da kratki ton ili samoglasnik funkcionira kao perceptivna jedinica, a da samo prepoznavanje predstavlja odčitavanje informacije u toj jedinici. Ako se slušni podražaj brzo mijenja, kao što je to slučaj kod govora, onda perceptivna obrada mora biti nastavljena i nakon prezentacije perceptivne jedinice. Zbog toga je vrlo važno specificirati koje varijable samog podražaja određuju formiranje perceptivnih jedinica informacije.

## **Perceptivne jedinice**

Prema Massarovoj hipotezi jedna perceptivna jedinica ne može biti nikako duža od predperceptivne slušne predodžbe i nakon nje mora postojati tihi period za vrijeme kojega se odvija obrada.

Vremenska jedinica mora, nadalje, biti određena vremenskim karakteristikama pritiska, jer bi se inače spajali u perceptivnu jedinicu test-ton i maskirajući ton.

Hirah /1959/ je tražio od ispitanika da odredi koji je između dva tona prvi izbio. Početak svakog tona bio je varijabilan, ali je svršetak bio istovremen. Upotrebljavana su dva tona različite frekvencije. Kad je prvi ton bio ispred drugog 17 ms, onda je bilo 75% točnih identifikacija. Ovo je vrijeme mnogo kraće nego li vrijeme koje je utvrdio Massaro /250 ms/. Izgleda, dakle da ispitanik može dva tona integrirati u vremenu i smanjiti efekat maskiranja. To bi značilo da perceptivnu jedinicu u tom pokusu Hirsra čini ili određuje kvalitet para tonova. Npr. visok – nizak ton ili nizak – visok ton. Ako, kao kod Massara, imamo period tišine između dva tona, onda takve integracije nema i svaki je ton zasebna perceptivna jedinica.

Cjelokupno trajanje tonova kod Hirshea bilo je između 440 i 560 ms. Patterson i Green /1970/ su utvrdili da se vremenski raspored može percipirati i onda kad su kraći cjelokupni intervali trajanja. Oni su utvrdili da ispitanici mogu čak identificirati vremensku razliku u rasporedu ili početku od svega 1.5 ms. Taj je interval komparabilan intervalu koji je neophodan da bi se dva slušna signala čula kao dva, a ne kao jedan /Miller i Taylor, 1948/. Hirsh smatra, međutim, da ta dva zadatka nisu istog reda. On smatra da je zadatak određivanja da li se radi o jednom ili dva tona nižeg a zadatak određivanja vremenskog rasporeda višeg reda. On smatra da je za ovaj drugi zadatak potrebno 20 ms više nego li za prvi.

Perceptivna obrada predstavlja identifikaciju slušne informacije koja je, već zbog svoje prirode, uskladištena sekvencijalno. Slušna predodžba mora stoga ispravno reprezentirati sam podražaj i njegov sekvencijalni raspored. Zbog toga vrijeme obrade, prema mišljenju Massara, mora biti duže. Da bi se ono izmjerilo, potrebno je imati takve eksperimentalne mogućnosti da se obrada prekida u različitim vremenima od časa prestanka podražaja. To se postiže tako, da se prezentiraju dva tona koji se razlikuju po svom početku, a nakon toga se prezentira treći ton i to poslije kraćeg tihog perioda. Taj treći ton bi trebao maskirati vremenski raspored, prva dva, ali samo do 250 ms.

Studdert – Kennedy, Skandweiler i Schulman (1970) studirali su identifikaciju dva sloga tipa konsonant – vokal kao funkciju razlike u vremenu njihovog započinjanja. Dva su se sloga razlikovala u prvom konsonantu. Slogovi su uvijek bili izabrani između slijedećih: Da, Ga, Pa, Ta, Ka. Slogovi su prezentirani u trajanju od 250 ms i to na isto ili na različita uha. Utvrđeno je, ako su bili prezentirani istovremeno, da je jedan od njih bio identificiran u 65%

slučajeva, bez obzira na raspored. Nadalje utvrđeno je da kod monauralne prezentacije, povećanje razlike u početku izbijanja jednog i drugog sloga je poboljšavalo identifikaciju prvog sloga a slabilo identifikaciju drugog. Kod dihotičke prezentacije povećanje vremenske razlike u izbijanju jednog i drugog sloga dovodilo je do poboljšanja prepoznavanja jednog i drugog sloga. Ovi rezultati sugeriraju različite procese koji se događaju kod monauralne i dihotičke stimulacije. Monauralni su podaci u skladu s Hirshom. To se vidi po tome što, kad je prvi slog bio ispred drugog za 17 ms onda je bilo 75% točnih identifikacija. Perfektna identifikacija postignuta je kod vremenske razlike u izbijanju od 50 ms. Postojao je i maskirajući efekat prvog na drugi slog. Drugi je slog bio to lošije identificiran što je bila veća razlika u započinjanju. Kad je razlika iznosila 50 ms, onda je identifikacija drugog sloga bila na nivou slučajnog pogađanja. **Kod 50 ms razlika oba se sloga čuju kao jedan slog s određenim klikom na kraju.**

Kod dihotičke prezentacije, kad je drugi slog kasnio 50 ili više ms, onda je taj drugi slog bio identificiran u 85% slučajeva. Iz toga se vidi da dihotička prezentacija ne utječe na integritet sloga.

Upravo zbog toga Massaro smatra da se kod dihotičke stimulacije radi o dvije perceptivne jedinice i da druga perceptivna jedinica mora u paradigmi prepoznavanja utjecati na prvu, jer interferira s procesom perceptivne obrade prve jedinice. Rezultati eksperimenata potvrđuju ovakva očekivanja. Identifikacija prvog sloga je najniža ako je vremenski razmak između prvog i drugog sloga 20 ms.

Što je vremenski interval duži, to je identifikacija sloga bolja. Međutim, i kod 120 ms razlike prvi se slog identificira samo u 85% slučajeva.

Massaro zaključuje da je optimalno vrijeme za obradu prvog sloga oko 250 ms, a to znači isto kao i za čiste tonove. Prema tome, rezultati Studdert – Kennedyja i drugih /1970/ potvrđuju važnost perceptivne jedinice i vremenske perceptivne obrade za slušnu percepciju.

Greel, Boomsliker i Powers /1970/ pružaju i neke kliničke podatke o tome. Pacijenti kod kojih je utvrđen defekt u opskrbi krvlju mozgovnog debla trebaju 200 do 400 ms čistog tona od 1000 Hz da bi ga percipirali kao ton, a ne kao šum. Normalni slušalac treba za takvu percepciju samo 10 ms. Nadalje, utvrdili su da sinusoidalni val od 300 ms pacijent čuje kao šum, a tek od 400 ms kao ton. To bi, prema Meisseru /1967/ i koncepciji analize putem sinteze, značilo da informacija nije dovoljno dugo perzistirala da bi došlo do sinteze tonalnog osjeta koja se, sinteza, događa nakon analize podražaja.



## **Obrada kontinuiranog govora**

Vrlo važni i interesantni podaci o perceptivnim jedinicama dobiveni su i u studijama isprekidanog i alteriranog govora. Kod isprekidanog govora ispitaniku se prezentira samo polovina govora dok je druga polovina izbrisana i nadomještena tišinom. Kod toga se govor i tišina alternativno izmijenjuju. Miller i Lickeider /1950/, te Direks i Power /1970/ su pokazali da se u takvim slučajevima inteligibilnost govora nije značajno smanjila pod uvjetom da se prazni intervali pojavljuju svakih 50 odnosno svakih 5 ms.

Međutim, ako su ti intervali dugački 500 ms onda se samo polovina riječi uspijeva prepoznati. To pokazuje da je trajanje perceptivnih jedinica koje su kritične za percepciju monosilabičnih riječi između 50 i 500 ms.

Pokazalo se, nadalje, da redundanca perceptivne jedinice osigurava u ovom slučaju prepoznavanje tih jedinica iako su alternativne dionice takvih jedinica – tišina.

Huggins /1964/ je proučavao percepciju dijelova znanstvenih tekstova od strane laika. Utvrdio je da je inteligibilnost teksta bila najniža kod 1.5 do 5 prekida u sekundi, tj. za periode tišine od 100 do 330 ms. Budući da je to vrijeme i vrijeme trajanja slogova, on je došao do važnog zaključka da je perceptivna jedinica govora slog. Odstranjenje slogova moralo bi dakle dovoditi do najniže inteligibilnosti i prepoznavanja govora.

Postoji još jedna tehnika. To je tehnika alteriranja govora od jednog na drugo uho. Ako se alternacija događa tri puta u sekundi, onda dolazi do gubitka percepcije govora /Cherry, 1953, Cherry i Taylor, 1954/. Ako su i alternacije brže, onda one ne dovode do smetnji u percipiranju govora. Huggins je brzinu govora kovarirao s brzinom alternacije i došao ponovo do zaključka da slog predstavlja perceptivnu jedinicu govora, jer se u slučaju takve kovarijacije /veće brzine, više alternacije/ u svakom uhu nađe samo po dio istoga sloga. Kad je alternacija sporija, onda više ili veći dio svakoga sloga ostaje u istom uhu i identifikacija je bolja.

Huggins /1964/ navodi još neke podatke mjerenja vremenskih karakteristika dijelova govora. 18% totalnog trajanja govora nije ništa drugo nego li tišina. To se vrijeme vjerojatno koristi za stvaranje perceptivnih jedinica i za njihovu obradu. Utvrdio je da je srednje trajanje jednog sloga 200 ms s međukvartalnim raspršenjem od 150 do 250 ms. Iz ovoga se vidi da se trajanje sloga poklapa s ranije spominjanim vremenom perceptivne obrade. To je ustanovljeno i na vidnom području kod čitanja teksta /Haber, 1970/.

Ovi rezultati ukazuju, dakle, na slog kao perceptivnu jedinicu. Warren /1970/ je pokazao da je moguće neki fonem zamijeniti s nekim posve drugim zvukom, a da slušalac ipak čuje taj fonem ukoliko je ta zamjena dovoljno kratka. Warren je u svom pokusu izbrisao 120 ms fonema “s” u jednoj riječi uklopljenoj u rečenicu i taj dio fonema nadomjestio je u jednom slučaju snimkom kašlja, a u drugom čistim tonom. Samo jedan od 40 ispitanika je izjavio da je primijetio da jedan ton manjka, a i taj ispitanik je izabrao pogrešan ton kad ga se pitalo koji ton manjka. Štoviše, ispitanici nisu mogli dijagnosticirati u rečenici prisustvo stranog zvuka i onda kad im je to bio zadatak.

**Ovi rezultati pokazuju da se perceptivna jedinica govora može sintetizirati i s ograničenom količinom govorne informacije u predperceptivnoj slušnoj predodžbi.** Ta se sinteza događa u ostalom govornom kontekstu. Relevantni podaci prije i poslije izbrisanog dijela perzistiraju u predperceptivnom pamćenju i na osnovu njih se sintetizira dio koji nedostaje. Ako se fonem nadomjesti tišinom, prekid se odmah zapaža. Međutim, usprkos toga, kao što je poznato, ipak možemo identificirati riječ.

Sada se postavlja pitanje da li dolazi do restauracije fonema ukoliko je predperceptivna predodžba izbljedita. To pitanje nije razjašnjeno i tu nedostaju eksperimentalna istraživanja. Moglo bi se očekivati da će ispitanik biti u mogućnosti zapaziti fonem koji nedostaje.

Cherry i Willey /1967/ i Holloway /1970/ pokazali su da ako se filtriraju samo jako naglašeni govorni glasovi, to smanjuje percepciju govora. Takva sekvenca glasova i tišine ne sadrži ritmičku prirodu normalnog govora. Međutim, ako se periodi tišine koji su nastali brisanjem normalnog govora, ispune bijelim šumom niskog intenziteta ponovo se uspostavlja inteligibilnost govora kao da je govor potpun. Prema tome, šum koji je umetnut umjesto govora, predstavlja u stvari most za organizaciju perceptivnih govornih jedinica, a čim su one formirane, inteligibilnost je restaurirana.

Liberman, Cooper, Schenkweiler i Studdert – Kennedy /1967/ iznijeli su, međutim, drugačije shvaćanje.

Oni smatraju da je bazična govorna jedinica fonem, a ne slog. Dakle, za percepciju govora trebala bi biti najvažnija percepcija fonema. To bi značilo da, ukoliko nema percepcije fonema, nema ni percepcije govora. Akustičke signale koje nose informaciju potrebnu za takvu percepciju oni analiziraju s obzirom na formante. Na slici 3 prikazan je rezultat jedne takve formantne analize fonema “Di” i “Du”.

Na slici 3. prikazana su prva dva formanta za “Di” i “Du”. Tamne površine predstavljaju koncentraciju zvučne energije. Očigledno je da je druga formantna struktura neophodna, i različita, za percepciju “D” – segmenta u fonemima “Di” i “Du”. “D” je perceptivno jednak u oba slučaja, ali je njegova akustička osnova, kako se to vidi na slici 3, potpuno različita. Tako, u slučaju fonema “Di” imamo porast od 2200 Hz na 2600 Hz. Kod fonema “Du” imamo pad tog formanta od 1200 Hz na 700 Hz. Autori su otišli i dalje i pokazali da, ako se ovi oblici energije prezentiraju izolirano za slušanje, onda se fonem “Di” čuje kao glisando ili zvižduk u visokim frekvencijama a fonem “Du” kao brzo padajući zvižduk na niskom području. Prema tome rezultati ovog eksperimenta ne pružaju podršku njihovim autorima u tvrdnji da fonem može biti perceptivna jedinica govora, jer se ovi rezultati mogu vrlo jednostavno i lako razumjeti ako se prihvati da je osnovna perceptivna jedinica slog. Ti su rezultati potpuno u skladu s tom koncepcijom.

Drugi rezultati istih autora pokazuju, da ako se prezentiraju samo raniji dijelovi fonema “Di”, dobivaju se negovorni glasovi. Ako se tome dodaju i drugi dijelovi, nikada se ne čuje fonem “D”, već se percepcija odmah izmijeni u “Di”. Dakle, s porastom dužine podražaja, percepcija se mijenja od negovornog glasa u klaster sastavljen od konzonanta i vokala. Dakle, i ovi rezultati govore više u prilog sloga kao perceptivne jedinice.

Postoje i drugi, noviji podaci koji govore da percepcija fonema dolazi tek iza percepcije sloga. Savin i Bever /1970/ prezentirali su ispitanicima besmislene slogove. Zadatak je ispitanika bio da što brže reagiraju na jedan znak u konsekvenci glasova. Vrijeme reagiranja bilo je kraće ako je bio slog, nego li ako je bio fonem u tom slogu.

Warren /1971/ proučavao je vrijeme identifikacije potrebno za prepoznavanje određenog znaka u kontinuiranom govoru. Našao je isto. Vrijeme reakcije je kraće za skupine fonema, nego li za izolirane foneme.

To pokazuje da je identifikacija slogova brža nego li identifikacija fonema. Massaro smatra, zbog toga, da u predperceptivnom uskladištenju dolazi do identifikacije slogova da se fonemi identificiraju nakon toga jednostavno iz samog pregleda identificiranog sloga. Warrenovi rezultati /1970/ o nadomještavanju ispuštenog fonema stranim zvukom u skladu su s ovom interpretacijom.

Lieberman i drugi /1967/ pokušali su proizvesti govor rekonstrukcijom fonema, ali nisu uspjeli. To također pokazuje da fonemi ne mogu biti perceptivne jedinice govora.

Nasuprot tome Pettersen, Wang i Silversten /1958/ proizveli su govor jedinicama koje su bile dugačke samo polovinu sloga. Zbog toga neki su autori skloni da odbace i Libermanovo tumačenje o paralelnoj obradi auditivne informacije.

Analiza prethodne slike /slika 3/ pokazuje da su 100 prvih ms važni za percepciju, a da ostalih 200 ms ne daju nikakvu novu informaciju. Budući da u normalnom govoru samoglasnici traju od 150 do 250 ms, jasno je da postoji dovoljno vremena za obradu tih zvučnih signala.

### **Auditivno kratkoročno pamćenje**

Za vrijeme perceptivne obrade, informacije se ekstrahiraju ili iz samog podražaja ili iz predodžbe. Sintetizirana percepcija ulazi kao jedinica u kratkoročno pamćenje. Kratkoročno pamćenje predstavlja ovdje još jedan neophodni konstrukt, jer se karakteristike slušnog inputa gube s novim slušnim ulazom. Da bi se zadržale već sintetizirane informacije i omogućila informacija s drugim sintetiziranim informacijama neophodno je pretpostaviti postojanje kratkoročnog pamćenja. Ovdje su analizirani u brojnim eksperimentima fenomeni interferencije koji su slični ili analogni fenomenima retrogradnog maskiranja prepoznavanja. Međutim, to je jedan sasvim drugi nivo istraživanja.

### **Ostali utjecaji**

U ovom kratkom pregledu nemoguće je dati kompletnu sliku stanja i razumijevanja percepcije govora ili prepoznavanja kompleksni auditivnih obrazaca jer kod tih procesa su važni osim kratkoročnog još i dugoročno pamćenje a zatim i procesi odlučivanja.

### **Uređaji za prepoznavanje govora**

U ovom kontekstu interesantna su istraživanja mogućnosti da se govor prepoznaje od strane posebnih uređaja, odnosno da se stvore uređaji koji će moći prepoznavati govorne obrasce. Poznati su slični napori, i uspjesi, koji su postignuti na vizuelnom području /perceptroni/. Što se tiče govora, danas su stvoreni već prvi kompjuteri koji su osposobljeni za govornu produkciju. To pokazuje koliko se daleko došlo u proučavanju i razumijevanju govornih karakteristika i zakonitosti u govornoj produkciji. Međutim, danas još ne postoje strojevi za

prepoznavanje govora koji bi bili toliko efikasni koliko i kompjuteri koji govore. To pokazuje da je problem percepcije govora složeniji od problema produkcije govora.

Među najvažnijim pitanjima koje treba riješiti da bi se stvorili strojevi za prepoznavanje govorne produkcije, spomenut ćemo slijedeće.

Prvo je pitanje varijabiliteta govornika. Postoje vrlo velike razlike u govornoj produkciji različitih ljudi. Uređaj koji bi trebao prepoznavati govor morao bi biti tako konstruiran da reagira na značajne a ne reagira na beznačajne razlike u produkciji govora različitih ljudi. Zatim postoje i veoma velike razlike u govornoj produkciji istog čovjeka u različitim vremenima. Osim toga, to su i velike razlike među ljudima iz različitih dijelova istog govornog područja. Dalji problem je problem govorne jedinice tj. gdje započinje a gdje završava perceptivna jedinica govora. Zatim tu je i problem namjernih govornih varijacija u intonaciji brzini i slično koje varijacije nose dodatnu informaciju osim semantičke informacije.

U naporima da se stvore uređaji za percepciju govora kao govorne jedinice najčešće su do sada uzimani fonemi. Ti fonemi proučavani su detaljno s obzirom na dva aspekta. Jedan aspekt je aspekt produkcije fonema i tu su proučavani procesi i pokreti artikulacije koji se javljaju u govornom aparatu čovjeka prilikom izgovaranja tih fonema a koji se mogu objektivno utvrditi. Drugi aspekt je akustički. Tu se opet objektivno mogu proučavati varijacije u amplitudi, frekvenciji i druge. Proučavanja jednog i drugog aspekta dala su ogroman broj objektivnih pouzdanih i valjanih podataka koji se koriste za kompjuterski govor /Fulgosi, rukopis/.

Na osnovu rezultata ovih istraživanja danas je moguće specificirati foneme ili na osnovu diskripcije ponašanja govornog aparata, /pokreta/ ili na osnovu fizikalnih promjena izvora vibracije /Jakobson, Fant i Halle, 1963/.

Postoje, ipak, relativno efikasni programi za prepoznavanje govorne produkcije. Ti su programi stvoreni na osnovu poznavanja govorne akustike i osnovu poznavanja lingvistike. Takvi programi mogu razlikovati čitav niz aspekata govornog materijala. Na primjer, s pomoću takvih programa mogu se razlikovati otvoreni od zatvorenih glasova. Uz pomoć njih moguće je distingvirati mjesto konstrikcije u govornom aparatu prilikom produkcije glasa i na osnovu toga razlikovati glasove kod kojih nema drugih razlika nego li razlike u mjestu konstrikcije govornog aparata. Takvi su glasovi recimo P i T. Programi su, nadalje, osposobljeni da razlikuju trajanje i afrikaciju. Sve su to dimenzije koje su važne za analizu

procesa produkcije govora. Tome se, naravno pridružuju fizikalne dimenzije samog glasa. Jedna od njih je npr. periodičnost ili neperiodičnost glasa. Periodični su, kao što je poznato, vokali i još neki drugi glasovi, a neperiodični su nenaglašeni suglasnici.

Spektrografske analize otkrivaju i druge mogućnosti distinkcije govornog materijala i govornih elemenata koji se mogu upotrijebiti za konstrukciju uređaja za prepoznavanje govora. Te frekvencijske komponente čine sklop pojedinog glasa /fonema/. To su u prvom redu formanti, naročito prvi i drugi formant svakog glasa. Ako se u koordinantnom sustavu na apscisu nanese frekvencija prvog formanta nekog fonema a na koordinatu frekvencija drugog formanta, onda je pokazano da svaki samoglasnik engleskog jezika zauzima drugo područje u tom koordinantnom sustavu. Ovakve analize ukazuju na važnost multidimenzionalnih /analiza svih mogućih razlika/ pristupa u proučavanju govornih elemenata. Nadalje, te analize pokazuju na koji se način, i izmjenom kojih formanta, mogu pojedini glasovi pretvarati jedni u druge, pa, prema tome, koji su formanti važni za distinkciju a koji ne.

Osim ovih karakteristika, postoje i druge koje su također važne za percepciju govora. Tu se obično spominje emocionalni sadržaj, semantički sadržaj, gramatička pravila, dob, spol, položaj, naglasak i druge karakteristike govornika.

Postoji već nekoliko operabilnih programa za prepoznavanje govora. Ovdje ćemo ukratko opisati program koji su stvorili Hughes i Hemdal /1965/.

Program Hughesa i Hemdala predstavlja, u stvari, jedan niz binarnih operacija odlučivanja o karakteristikama govornog uzorka koji se nalazi u analizi. Obično se taj program ilustrira na besmislenim riječima i to tako da se jedan veći broj besmislenih riječi koje je govornik izgovorio najprije propusti kroz 35 filtera. Hughes i Hemdal su na izlazu iz filtera uzimali filtrirane uzorke svakih 16 ms. Ovakvo trajanje je dovoljno da se za svaki fonem dobije po nekoliko uzoraka. Početak i kraj svake riječi obilježeni su tišinom.

Svaki uzorak filtriranog govornog materijala od 16 ms najprije se testira s obzirom na to da li je to period tišine ili u sebi sadrži zvukove. Ukoliko takav uzorak predstavlja uzorak glasova, onda se prelazi na rješavanje drugog binarnog problema, tj. da li on sadrži "zvučne" glasove ili "bezvučne" glasove.

To se otkriva analizom frekvencije koje sadrži takav uzorak, tj. da li uzorak ima frekvencije ispod 350 Hz ili ne. Ukoliko uzorak ne sadrži "zvučne" glasove, istražuje se treće pitanje, tj. da li se ispred uzorka nalazio period tišine ili ne. Ukoliko je prije tog uzorka bio period tišine, onda uzorak mora sadržati neki od slijedećih fonema /engleskih/: T, P, K, ili Ch. Ukoliko

prije tog uzorka nije bio period tišine, onda je to uzorak koji mora sadržati neki od fonema: F, S, ili Th.

Ukoliko je uzorak sadržavao “zvučne” glasove, onda se pristupa analizi frekvencije. Ukoliko dominiraju visoke frekvencije, onda taj uzorak mora sadržavati neki od ovih fonema: Ž, Z, V, /delta/. Ako nema visokih frekvencija onda se ispituje da li postoji tišina ispred uzorka. Ako postoji onda se radi o glasovima B, D, G ili Dž. Ukoliko nije bilo tišine ispred uzorka, onda uzorak sadrži ili konzonante, sonante ili vokale. Ukoliko su glasovi jaki, tj. s mnogo energije, onda se radi o vokalima, a ako je energija mala onda su to konzonanti M, N, L, W, R. Nakon toga analiza se produžava.

Ispitivanja Hughes i Hemdaelovog programa za percipiranje govornih sadržaja su pokazala da taj program može prepoznati glasove u besmislenim riječima s točnošću od 90%, tj. s istom tolikom točnošću s kolikom ih točnošću može prepoznati i čovjek.