

Beszédpercepció és pszicholingvisztika

Mády Katalin

A pszicholingvisztikai modellek többsége a nyelvi feldolgozás legkisebb egységként a beszédhangok szintjét teszi fel. A modellek – a hagyományos nyelvészeti szinteknek megfelelően – többnyire különbséget tesznek az absztrakt, azaz fonológiai, és a konkrét, azaz fonetikai kategória között. Utóbbi közvetlen kapcsolatban áll a beszéd fizikai megvalósulásával, tehát a percepció láncban a bemenet utáni első, a produkció láncban a kimenet előtti utolsó mentális láncszemet alkotja. Az absztrakt fonológiai egységeket egyrészt az jellemzi, hogy szoros kapcsolatban állnak a nyelv felsőbb szerkezeti egységeivel (morfológia, szintaxis, lexikon), tehát ekként a nyelvi rendszer részei, másrészt hogy a legtöbb hipotézis szerint állandó tulajdonságokkal rendelkeznek, azaz invariáns nyelvi kategóriákat alkotnak. A beszédhangok, azaz a fonetikai szint elemei viszont igen sokfélék lehetnek, és sokszor korántsem könnyű eldönteni, milyen közös jegyek alapján ítélt egy anyanyelvi beszélő két hangot azonos osztályhoz tartozónak. Például az alveoláris, azaz a felső fogmeder érintésével képzett /n/ hang a környezet hatására jelentős változásokon megy keresztül: a *len* szóban alveoláris, a *kengyel* szóban palatális, a *ménkü* szóban veláris, a *kénbűz* szóban labiális. Ez a hely szerinti hasonulás könnyen megmagyarázható a követő mássalhangzó képzési helyével, az azonban már kevésbé, hogy miért hajlamosabb a hasonulásra az /n/, mint más nazális mássalhangzók, vagy hogy bizonytalan képzés esetén miért hajlunk inkább arra, hogy egy hangot alveolárisként halljunk, mint hogy valamely más képzési helyet tulajdonítsunk neki. (Erre a későbbiekben még visszatérünk).

A fonetikai szint beszédhangjait tehát a variabilitás, vagy más szóval az állandóság hiánya jellemzi, míg az absztrakt fonémák észlelése nem képzelhető el invariabilitás nélkül - másképp hogyan lennének képesek a beszédben előforduló sok száz hangot ötvennél kevesebb fonémához hozzárendelni? Általánosságban azt mondhatjuk, hogy a beszédpercepcióval foglalkozó kutatások fő célja a beszédhangok képviselte sokféleségben azokat az ún. **felismerési kulcsokat** megtalálni, amelyek az emberi agy számára lehetővé teszik a releváns jegyek felismerését és a redundáns jegyek figyelmen kívül hagyását. Más szóval a beszédpercepció a pszicholingvisztikai modellek fonológiai és fonetikai szintje között található részfolyamatokat igyekszik leírni és modellezni.

A beszédészlelés a kommunikációs lánc harmadik nagy egységét képi a beszédprodukció és a nyelvi jel után, következésképpen függvénye mind a beszélőszervekkel előállítható akusztikumnak, mind magának az akusztikai jelnek. Más szóval: a percepció értelemszerűen arra a frekvenciasávra és azokra az időbeli felbontásokra van utalva, amit egy emberi beszélő képes produkálni, és amit a közvetítő közeg – a levegő – továbbítani tud. Mint látni fogjuk, az emberi fül különösen jól alkalmas az ember által létrehozható rezgések észlelésére, és a levegőben való hallásra rendezkedett be. Számos fonetikai elmélet abból indul ki, hogy nemcsak a hallás alkalmazkodik a beszédprodukcióhoz, hanem a beszélő is a percepcióhoz: öntudatlanul olyan jeleket produkál, amelyekre az emberi fül különösen fogékony. A produkció és a percepció tehát nemcsak az előállított, ill. észlelendő akusztikai jelen keresztül kapcsolódik egymáshoz, hanem azáltal is, hogy a beszélő ismeri a percepció működését, a hallgató pedig a produkciós folyamatokat. Ebben az értelemben a kommunikációs lánc metaforája félrevezető, hiszen nem egymás után csatolt, hanem egyszerre ható folyamatokról van szó. A beszédpercepció leírásakor tehát szem előtt kell tartanunk a beszédprodukció folyamatát is.

A fejezet tagolása a következő: elsőként a fül anatómiájáról és a hallás fiziológiai alapjairól lesz szó. A leírás a percepció folyamatok megértéséhez szükséges alapvető ismeretekre szorítkozik. A második fejezetben az alapvető szegmentális és szupraszegmentális egységek felismerésére térünk ki, amelyeket a harmadik fejezetben különféle percepció elméletek fényében vizsgálunk meg újra. Végül a negyedik fejezetben egyes percepció elméletek által felvetett kérdésekre keresünk válaszokat.

¹ A fonetikai leírásokban nem egységes a szögletes és ferde zárójel használata: egyes rendszerekben a /n/ jelölés a fonémára, az [n] pedig a kiejtett beszédhangra vonatkozik, mások viszont a ferde zárójelt használják a durvább, a szögleteset pedig a finomabb fonetikai átírásra. E fejezet az utóbbi hagyományt követi.

1 Hallás és pszichoakusztika

Az emberi észleléssel általános értelemben a pszichofizika foglalkozik. E tudományág azt vizsgálja, hogy fizikai konkrét nagyságok és az ember által észlelt mértékük között milyen összefüggések vannak: pl. súly és szubjektív nehézségérzet, hangerő és hangosság. Ezen belül a pszichoakusztika a hanghullámok érzékelésének részterülete. A percepció fonetika vagy pszichofonetika e tudományterülettel részben határos, hiszen szintén az akusztikai jelek észlelésére koncentrál, ám csupán azokra, amelyek a beszéd szempontjából relevánsak, tehát az emberi beszédképző szervekkel létrehozhatóak (kb. 50 Hz és 10 kHz között). Mégsem mondhatjuk, hogy a percepció fonetika a pszichoakusztika része lenne, hiszen nemcsak a jelek észlelésével, hanem azok nyelvi értelmezésével is foglalkozik, ami nem tartozik a pszichofizika vagy -akusztika kutatási területéhez.

Mielőtt rátérnénk az emberi nyelv szempontjából releváns akusztikai információk feldolgozására, tegyünk egy kis kitérőt, és nézzük meg röviden, milyen biológiai alapokra épül a beszédpercepció, azaz hogyan zajlik az akusztikus jel feldolgozása a fülben és az agyban.

1.1 Az emberi hallás anatómiája és fiziológiája

Az emberi hallószerv két részre tagolható: a perifériás hallószervként működő fülre és a központi idegrendszer hangérzékelésben érintett részére, amely a központi hallószervet alkotja, és az észlelt ingereket a nyelvpercepció nagyobb egységei felé továbbítja.

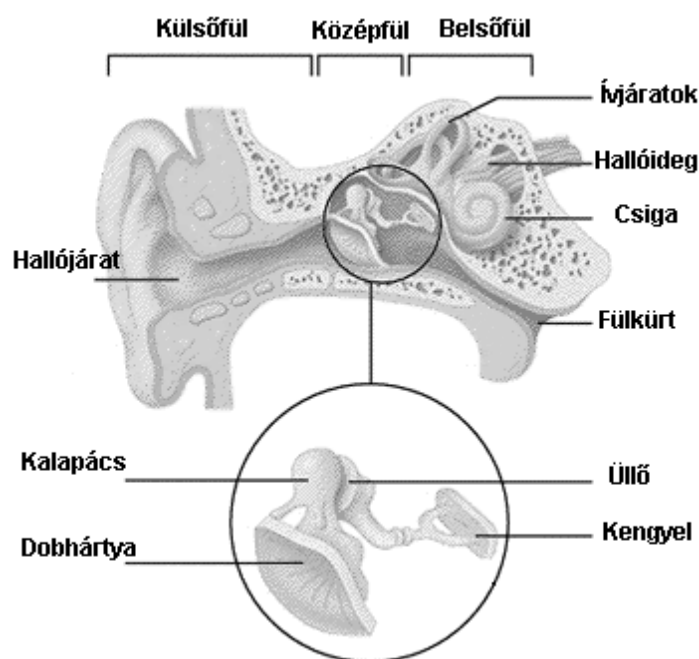
1.1.1 A perifériás hallószerv

A perifériás hallószerv három részre tagolható: a külső, közép- és belfülre. Ezek szerepe a hallásban a következő (ld. Pompino-Marschall 2003, Pétursson & Neppert 1991, Pfitzinger 2006, Reichel 2007, magyarul Szentágothai 1971, Bolla 1995):

- **külsőfül:** hanghullámok vétele és felerősítése,
- **középfül:** hanghullámok mechanikus továbbítása és felerősítése,
- **belsőfül:** hidraulikus rendszer, benne a Corti-szervvel, ahol az ingerek transzformációja végbemegy és továbbítódik az idegrendszerbe.

A **külsőfül** a fej két oldalán, a koponyacsontokhoz kb. 15-os dőlésszögben csatlakozó fülkagylóból és a külső hallójáratból áll. A fülkagyló helyzete és asszimmetrikus alakja elősegíti a hanghullámok irányának betájolását, egyrészt azért, hogy a hátulról jövő hanghatásokat a kagyló leárnyékolja, másrészt mert a hátulról érkező hanghullámokat a hallójárat elülső fülkagylóporcai lényegesen gyengébben verik vissza. Az irányfüggő hallás egyben frekvenciafüggő is: a beszéd szempontjából lényeges frekvenciatartományt (50–10 000 Hz) akkor észleljük legpontosabban, ha a hangforrás a test szagittális ² keresztmetszetétől 25–50 fokra található. Triviális, de lényeges körülmény a hangforrás irányának meghatározása szempontjából, hogy két fülünk van, kb. 20 cm-re egymástól, ami a pontos hely észlelését az eltérő intenzitásoknak és az észlelés eltérő időpontjának köszönhetően nagyban megkönnyíti (ld. 1.1.3 rész). Ennek azért van nagy jelentősége, mert szemben más emlősökkel, az ember nem képes arra, hogy fülét a hangforrás irányába fordítsa.

2 A testen előlről hátra húzott függőleges képzeletbeli választóvonal.



1. ábra: A fül anatómiai felépítése. (Forrás: www.fulspecialista.hu.)

A hanghullám, pontosabban a hangnyomás a fülkagylóból a külső hallójáratba kerül, azaz egy kb. 2,5 cm hosszú csőbe, amely méreténél fogva a 3 400 Hz és 10 000 Hz körüli frekvenciatartományokat viszonylag széles sávban (2–5 kHz és 8–12 kHz között) 2–4-szeresükre erősíti; ezáltal a beszéd szempontjából fontos tartományok intenzitása 12 dB-lel nő. A hallójáratot a rovarokra, gombákra és baktériumokra ártalmatlan fülzsír védi a fertőzéstől és a kiszáradástól, és a dobhártya választja el a mögötte található középfülettől. A hallójárat tekervényes alakja egyszersmind a dobhártya mechanikus védelmére is szolgál.

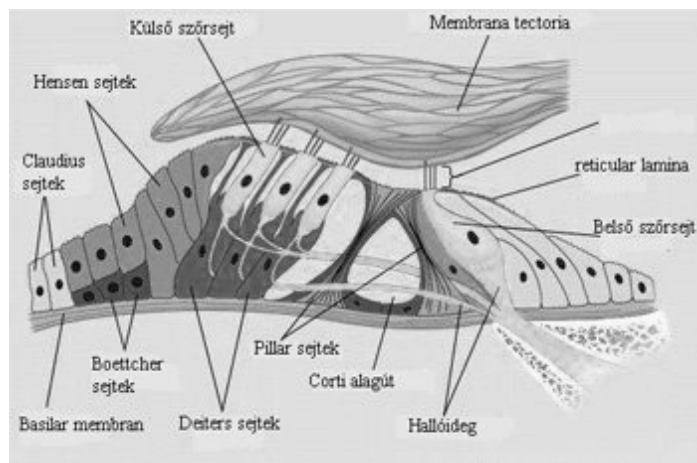
A **középfül** fő feladata, hogy a levegővel töltött külsőfül, valamint a folyadékkal töltött belsőfül közötti átmenetet biztosítsa. A középfület az ún. **fülkürt** vagy Eustach-kürt köti össze az orrgarattal, ez gondoskodik a külvilággal való nyomáskiegyenlítésről. A fülkürt többnyire zárt állapotban van, hogy a középfület védje a fertőzésektől, és csupán nyeléskor és ásításkor nyílik meg; ilyenkor jön létre a nyomáskiegyenlítés az atmoszférikus nyomással (ezért nyelünk automatikusan, ha a repülő változtatja a magasságát, vagy a vonat alagútba ér). A középfül bejáratát a dobhártya alkotja, amelynek vibráló felülete kb. 0,5 cm, vastagsága pedig 0,1 mm. A dobhártya sajátfrekvenciája 500 Hz és 3 500 Hz között van, ebben a tartományban tehát szinte veszteségmentesen továbbítja a rezgéseket. A dobhártyához csatlakozik a három parányi hallócsont: a kalapács, az üllő és a kengyel. E három csont az emberi szervezetben található csontok legkisebbike, amelyek az egyedfejlődés során nem nőnek, tehát már a csecsemőkorra elérik teljes nagyságukat.

A hallócsontocskák továbbítják a hanghullám okozta nyomást a belsőfül bemeneteként szolgáló ovális ablakhoz, amelynek felülete 0,03 cm, tehát a dobhártya felületénél egy nagyságrenddel kisebb. Ennek köszönhetően a hangnyomás 22-szeresére, azaz 27 dB-lel nő,³ amikor a középfület elhagyja. A megnövekedett intenzitás elősegíti a hanghullámok továbbterjedését a belsőfül kiegészítő folyadékban. A középfül hallócsontocskái mellett a belsőfül védelmére is szolgálnak: ha ugyanis túl erős (85 dB-nél nagyobb) hanginger éri a fület, a kengyelizom (*musculus stapedius*) reflexszerűen összehúzódik, ezáltal a csontocskák alkotta emelőrendszer megmerevedik, és a hangnyomást csökkentett intenzitással továbbítja.⁴ Nem

³ Ez az arány arra vezethető vissza, hogy a decibel nem lineáris, hanem logaritmikus mértékegység (ld. Vicsi: *A beszéd akusztikai fonetikai leírása* c. fejezet). A kétszeres intenzitás definíció szerint a kétszeres hangnyomás 6 dB-nyi szintnövekedésnek, azaz 3 dB-nyi intenzitásnövekedésnek felel meg.

⁴ Mivel a kengyelizom reakcióideje 100–200 ms között van, ez a reflex nem véd a hirtelen zajoktól.

egyértelmű, hogy a dobhártya feszítőizma (*musculus tensor tympani*) milyen szerepet játszik: egyes szerzők szerint a *musculus stapedius*hoz hasonlóan hangtompító hatást fejt ki, mások szerint viszont éppen ellentétes hatást fejt ki: a hallócsontocskák megmerevítésével elősegíti a hanghullámok jobb továbbítását.

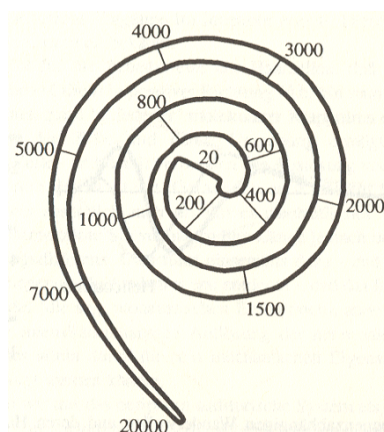


2. ábra: A fedőhártya, a hozzá kapcsolódó külső és belső szőrsejtek, valamint az alaphártya (itt: basilar membran). (Forrás: www.fulspecialista.hu.)

A **belsőfület** az előcsarnok vagy pitvar (*vestibulum*) és a csontos, valamint a hártyás labirintus alkotja. Benne található az egyensúlyi szerv három félkörös ívjáratával, valamint a voltaképpeni hallószerv, a csiga (*cochlea*). Az emberi fülben található csiga két és félszer tekeredik fel. Menetét két, membránok által határolt csatorna alkotja: a felső folyadékter vagy tornáclépcső (*scala vestibuli*) és az alsó folyadékter vagy dobúri lépcső (*scala tympani*), amelyek a csiga csúcsánál (*helicotrema*) találkoznak. Előbbi az ovális ablaktól a csiga csúcsáig tart, utóbbi onnan kifelé halad, és egy membrán, az ún. **kerek ablak** zárja le. Mindkét csatornát folyadék (perilympha) tölti ki. A felső folyadékteret alulról a Reissner-hártya, az alsó folyadékteret felülről az alaphártya (*membrana basilaris*) határolja, a kettő között helyezkedik el az endolymphával töltött középső járat, a *scala media*. Az alaphártyához kapcsolódik a voltaképpeni hallószerv, az ún. Corti-szerv.

A Corti-szervben található a hallás receptorsejtjei: a kb. 3 600 belső és 20 000 külső szőrsejt, valamint a rajtuk lévő csillószőrök (sztereociliumok). Ha az alaphártya egy hanghullámra reagálva kileng, akkor egyidejűleg elmozdítja a fölötte található **fedőhártyát** vagy fedőlemezt (*membrana tectoria*). Ezáltal a szőrsejtek elhajlanak vagy elmozdulnak, és a folyadék ionkoncentrációja megváltozik. Az inger továbbítódik a ganglion spirale, onnan pedig a hallóideg sejtjeihez (ld. 2. ábra). A szőrsejtek tehát másodlagos receptorsejtként működnek, hiszen belőlük közvetlenül nem váltható ki akciós potenciál, de a hallóideg dúca, a ganglion spirale szinaptikus úton hozzájuk kapcsolódó sejtjeiből igen.

A feltekeredett, teljes hosszában 32 mm hosszú alaphártya a csiga csúcsa felé egyre szélesedik: alapjánál az átmérője 0,04 mm, csúcsánál 0,5 mm, egyszersmind egyre rugalmasabbá is válik. A kengyel mozgása a belsőfület kitöltő folyadékban hullámot hoz létre, amely a csiga mentén vándorol, még hozzá frekvenciájától függően eltérő távolságra: a magas frekvenciák az alaphártyát már a csigába való belépéskor rezgésbe hozzák, és itt el is veszítik energiájukat, az alacsony frekvenciájú hullámok viszont továbbvándorolnak, az alaphártyát csak a csiga belsejét elérve mozdítják ki helyzetéből, ahol az ellenállása kisebb. Más szóval a csiga a rezgéseket – a Fourier-transzformációhoz hasonlóan – frekvenciaösszetevőire bontja (ld. a következő részben).



3. ábra: A vándorhullámok végpontjának helye a csigaszerűen feltekeredett alaphártyán.
(Forrás: Pompino-Marschall 2003:147.)

1.1.2 Halláselméletek

A csiga **tonotópiás** elrendezésű, tehát frekvenciaérzékelése helyfüggő (szomszédos frekvenciákra szomszédos területeken reagál, ld. 3. ábra). E felismerés Helmholtz nevéhez, közelebbről 1862-es rezonanciaelméletéhez kötődik. Ebben az alaphártyát egy hárfához hasonlóan írta le, ugyanis azt feltételezte, hogy az alaphártya párhuzamos, a hárfa húrjainak megfelelő száalából áll.

Később azonban kiderült, hogy az alaphártya elasztikus szerv, felépítése folyamatos, nem pedig húrszerű. Ennek megfelelően Békésy György, Nobel-díjas fizikus 1960-ban megalkotta a **vándorhullám-elméletet**, amelyben feltételezi, hogy a kengyel bizonyos frekvenciájú mozgása az alaphártyán végigvándorló hullámmozgást vált ki, amely – mint fentebb láttuk – az alaphártya anatómiai adottságaiból adódóan frekvenciától függően a membrán kezdeti vagy későbbi szakaszán okoz maximális elmozdulást.

1.1.3 A belsőfülből származó ingerek feldolgozása az agyban

Mint fentebb láttuk, a szőrsejtek szinapszissokkal kapcsolódnak a ganglion spirale idegszáaljaihoz. A ganglion spirale bipoláris magokból tevődik össze, és a hallóideg afferens, azaz a központi idegrendszer felé vezető pályáját alkotja. Az afferens idegpálya 95%-ban a belső, 5%-ban a külső szőrsejtekhez kapcsolódik.

A hallóideg kb. 30 000 idegszáalból áll, amelyeknek kb. 6%-a efferens, azaz a központi idegrendszerből a perifériás idegrendszerbe továbbít információt. Ezek túlnyomó része (kb. 90%-a) a külső szőrsejtekhez vezet, kis része a belső szőrsejtekhez, és az izomműködés szabályozásáért felelős.

Általánosságban tehát elmondhatjuk, hogy a külső szőrsejtek alapvetően az efferens, a belső szőrsejtek alapvetően az afferens idegpálya részei. Míg az efferens sejtek által kiváltott izomösszehúzódás aktívan befolyásolja az alaphártya viselkedését, az afferens belső szőrsejtek receptorsejteként működnek. Kísérleti eredmények arra utalnak, hogy a külső szőrsejtek növelni tudják az alaphártya egyes frekvenciákra való fogékonyságát. A visszacsatolásban keletkező zavar a fülzúgás (*tinnitus*) egyik feltételezett oka.

A csigától a halántéklebenyig húzódó idegpályát működése alapján két nagy egységre oszthatjuk: a **ventrális** (= has felőli) és a **dorzális** (= hát felőli) auditív pályára.

A ventrális pálya elsősorban a hanghatás irányának meghatározására szolgál, és elsősorban ipsilaterális száalakat tartalmaz. A pálya a nucleus cochlearis ventralison keresztül az azonos oldali oliva superiorhoz, onnan pedig az ipsilaterális (= azonos oldali) agykéreghez vezet. A felső olivában találkozik először a két fülből érkező információ. A ventrális pálya idegsejtjei nemcsak bizonyos frekvenciákra, hanem a rezgés bizonyos fázisaira is reagálnak. A felső oliva így értesül a két fül által eltérő futamidővel felvett ingerekről, ezáltal képes a hangforrás helyzetének meghatározására. A felső oliva emellett olyan sejteket is tartalmaz, amelyek a

beérkező jelek intenzitáskülönbségeit érzékelik, ezzel szintén támogatva a térbeli hallást. A ventrális rendszer kapcsolódik a vizuális rendszerhez is, és egyes összekötő sejtek innervációja csak akkor jön létre, ha bennük az auditív és a vizuális pályán azonos irányból érkező ingerek találkoznak (erre a 4.2 részben még visszatérünk).

A kontralaterális felépítésű dorzális pálya bonyolult inger- és ingerületmintákat tartalmaz. Ez a pálya nemcsak egyes frekvenciákra, hanem ezek kombinációjára, zajra, frekvencia- és amplitúdóváltozásokra is érzékeny. Az agykéregben ezért az alaphártya tonotópiás szerveződése számos komplex percepciós folyamattal egészül ki.

A fentiekkel párhuzamos efferens idegpályának köszönhető az a képességünk, hogy bizonyos irányból érkező hangokra fokozottan tudunk koncentrálni (szelektív térbeli hallás). Egy nagyobb társaságban a magas alapzaj ellenére képesek vagyunk meghallani, mit mond beszélgetőpartnerünk, mert a környezetből érkező hangingereket képesek vagyunk kikapcsolni. Ezt a jelenséget koktélparti-effektusnak is szokás nevezni. Jól bizonyítja a szelektív térbeli hallás szerepét az a tény, hogy egy a partiról készült monó hangfelvételen nem tudjuk ugyanazokat a beszédfoszlányokat kiszűrni, azaz a környezeti zajt elnyomni.

1.2 Akusztikai jelek percepciója

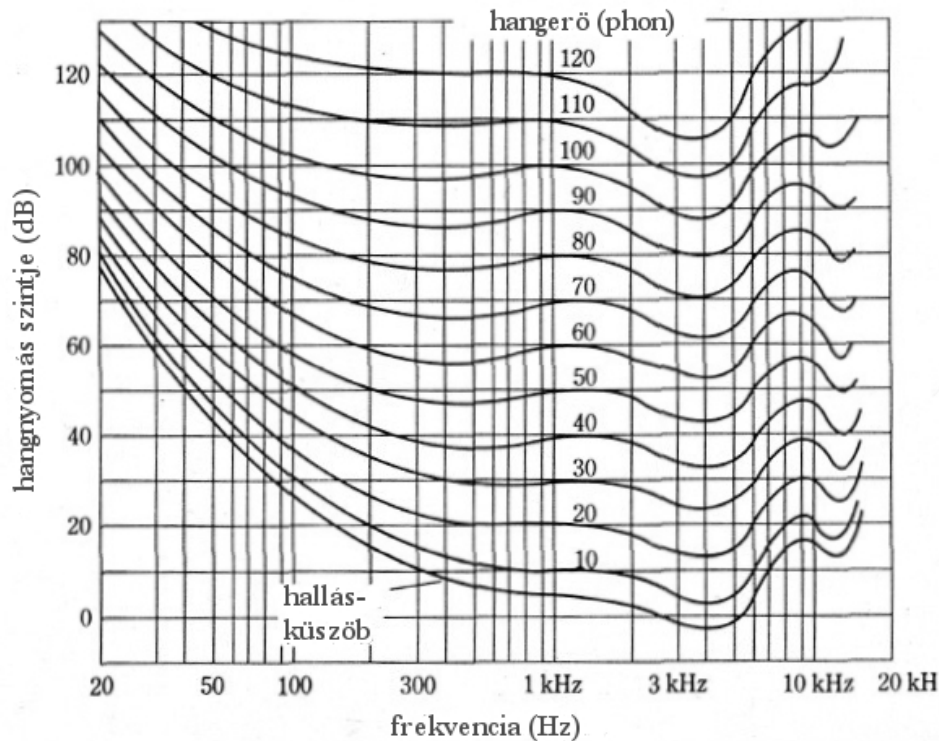
A beszédpercepció bemeneteként szolgáló akusztikai jel fizikai jellemzői objektíven mérhetőek, azonban az emberi észlelés ezeket a tulajdonságokat sajátos szűrőkön keresztül érzékeli. Más szóval, a nyelvi feldolgozás bemenete, tehát a perifériás és központi jelfeldolgozás terméke nem azonos a külsőfül bemenetével.

Mint a bevezetőben elmondtuk, az auditív észlelés egységeivel a pszichoakusztika foglalkozik. Mivel az észlelt jelhez közvetlenül nem férünk hozzá, e tudományág felismerései kísérletek eredményeire épülnek. A pszichoakusztika fő célkitűzése, hogy felfedje az akusztikai jel fizikai mérőszámai és az észlelés hozzájuk rendelt pszichikai mérőszámai közötti matematikai összefüggést.

1.2.1 Intenzitás, hangerő és hangosság

Az emberi hangészlelés alsó határa 20 Hz, felső határa 20 000 Hz, azaz 20 kHz. Ez az adat fiatal felnőttek hallására vonatkozik. Idősebb korban a magas frekvenciák észlelése fokozatosan romlik, ezért fordul elő, hogy egy idős ember nem hallja meg a csengőt a lakásában, de dörömbölésre ajtót nyit.

A hanghullámok terjedése a levegőben a hangnyomás változásával jön létre. A hangnyomás és az ebből származtatott hangintenzitás átlagos (effektív) értékét mérjük, és értékeit a több nagyságrendnyi átfogás miatt [dB] szintértékben adjuk meg (ld. Vicsi: *A beszéd akusztikai fonetikai leírása* c. fejezet). Az előző részekben kitértünk arra, hogy a fül anatómiai szerkezetének köszönhetően a fül egyes frekvenciatartományokat jobban felerősít, ebből következően a hallás érzékenysége frekvenciafüggő: az észlelés a 3 400 Hz körüli tartományban a legjobb, az 500 Hz-nél mélyebb, ill. 10 kHz-nél magasabb tartományokban kevésbé pontos. Míg a még észlelhető frekvenciatartomány két szélső értékén lévő hangokat csak 70 dB fölötti intenzitással halljuk meg, egy 3 400 Hz körüli hangot süketszobában már -10 dB-es intenzitással is. Ezért a frekvenciafüggő hangerőt saját érzeti mértékegységgel, a **phonnal** szokás megadni. Az azonos phon-értékkel jelölt frekvenciatartományt a kísérleti alanyok azonos hangosságúként ítélték meg. A phon-skála alapjául az 1 kHz frekvencia szolgál, amely hangereje phonban kifejezve számban megegyezik a hozzá tartozó decibelértékkel. Tehát ha egy 1 kHz-es hang intenzitása 20 dB, akkor ehhez 20 phon-os érték járul, 100 Hz-en viszont azonos hangosságélményhez 30 és 40 dB közötti intenzitásra van szükség. A még éppen észlelhető hangerő alsó határa a hallásküszöb, felső határa pedig a fájdalomküszöb.



4. ábra: A hangnyomás (dB) és hangosság (phon) összefüggése süketszobában mérve.

Mivel a phon-skála nem arányskála, nem árul el semmit a hanghullámok egymáshoz viszonyított hangosságáról. A szubjektív hangosságárányérzet kifejezésére a **son** alapú skálát szokás használni. A son-skála sarokpontját 1 kHz-es hangmagasság, 40 dB-es hangerő és 1 másodperces tartam képezi, amelynek hangossága 1 sonnak felel meg. A son-görbe tehát hangosságárány-érzeti skála, azaz kétszeres son-érték kétszeres hangosságélménnyel párosul. A phon-görbék ezzel szemben az egyenlő hangosságérzetet fejezik ki, a phon tehát a hangosságszintérzet mértékegysége (ld. ISO Szabvány R 226-1961).

1.2.2 A hangmagasság észlelése

A relatív hangmagasságot a zenében hagyományosan hangközökkel, azaz félhangokra épülő viszonyszámokkal szokás kifejezni: 12 félhang egy oktáv, amely 1:2 -es frekvenciaarányának felel meg, tehát a 440 Hz-es zenei, azaz az egyvonalas a -nál egy oktávval mélyebb kis a frekvenciája 220 Hz. Egymástól egy kvint távolságra lévő hangok frekvenciájának aránya 3:2, a kvartnyi távolság viszonya 4:3, a nagy szekundé 9:8 (részletesebben ld. Tarnóczy 1982).

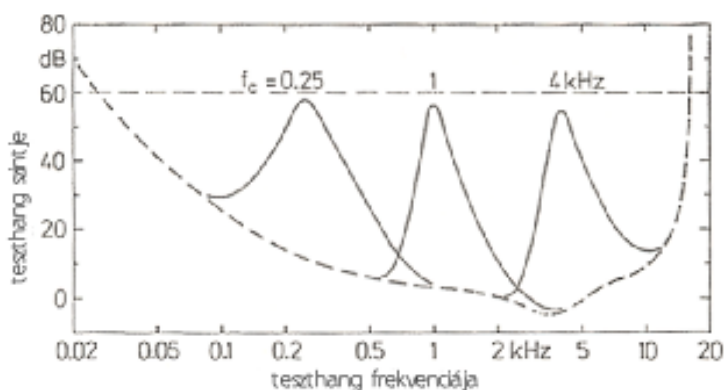
Az alaphártya által érzékelt frekvenciák, valamint a szőrsejtek eloszlásának köszönhetően az emberi fül a hangmagasságot különböző frekvenciákon eltérő pontossággal érzékeli. Míg 20 Hz és 500 Hz között az észlelés nagyjából lineáris, előlött a frekvencia és az észlelés helye közötti összefüggés logaritmikus. Ennek megfelelően az alacsonyabb frekvenciák észlelése jobb felbontású, mint a magasabbaké. Kísérletek során tesztelték, milyen frekvenciájú tiszta szinuszos hangokat érzékelünk egymástól azonos távolságra levőnek. A kapott értékeken alapszik a **mel**-skála (nevében a *melody*, azaz dallam szóra utal), kiindulásaként pedig egy 1 kHz-es, 40 dB-es hang szolgál, amelynek érzékelt magasságát 1 000 mel-ben állapították meg.⁵ Míg az észlelés 500 Hz-ig nagyjából lineáris, előlött a Hertz- és mel-értékek aránya logaritmikus. Így például 100 és 200 Hz távolságának 150 és 283 mel magasságú észlelet felel meg, az 1 000 és 2 000 Hz-es hangok között már kisebb a távolság: 1 000 és 1 521 mel, a

⁵ Fontos tudni, hogy a mel-t egyes források a fentiekől eltérően a Bark-skála alapján definiálják: 1 Bark = 100 mel (ld. Pompino-Marschall 2003).

10 000 és 20 000 Hz-nek megfelelő értékek aránya pedig még kisebb: 3 079 és 3 817 mel⁶ (ld. Tarnóczy 1982).

A hangmagasság észlelésében fontos szerepet játszanak a Fletcher (1940) által kimutatott ún. **kritikus sávok**. Ennek lényege, hogy egy adott frekvenciájú szinuszos hang észlelését adott körülmények között befolyásolhatja egy vele egyidőben lejátszott szélessávú fehér zaj. Ha a szinuszos hang és a fehér zaj frekvenciatartománya nem fedi egymást, a két hangot egymástól függetlenül észleljük. Ha azonban a szinuszos hang a fehér zaj középső frekvenciájával azonos, akkor a hangot – a két hangforrás intenzitásától függően – halkabban, vagy egyáltalán nem észleljük, azaz a fehér zaj **elfedi** azt. A jelenség a fehér zaj frekvenciatartományának közepétől távolodva is megfigyelhető, de egyre csökkenő mértékben: a zaj intenzitásának nagyobbának, ill. a szinuszos hang intenzitásának kisebbnek kell lennie a maszkolási jelenség kiváltására.

A kritikus sávok módosító hatását szemlélteti az 5. ábra. A csendes környezetre vonatkozó hallásküszöb (szaggatott görbe) fölfelé módosul 60 dB-es fehér zaj bejátszásakor. Ha a zaj középső frekvenciája 250 Hz, 1, ill. 4 kHz, amint ezt az ábrán láthatjuk, akkor egy 50 dB-es hangerejű szinuszos hangot teljesen elfed. A kritikus sáv szélessége abból a frekvenciasávból adódik, amely egy adott frekvencia körül befolyásolja a sáv közepére eső hang észlelését (az ábrán a folyamatos vonallal jelölt sávok szélessége).



5. ábra: Folyamatos vonalak: 250 Hz, 1 kHz és 4 kHz körüli, 1 kritikus sáv szélességű, 60 dB intenzitású szélessávú zaj által kiváltott hallásküszöb-módosulás. Szaggatott vonal: hallásküszöb csendes környezetben. (Fastl & Zwicker 2006:64 nyomán.)

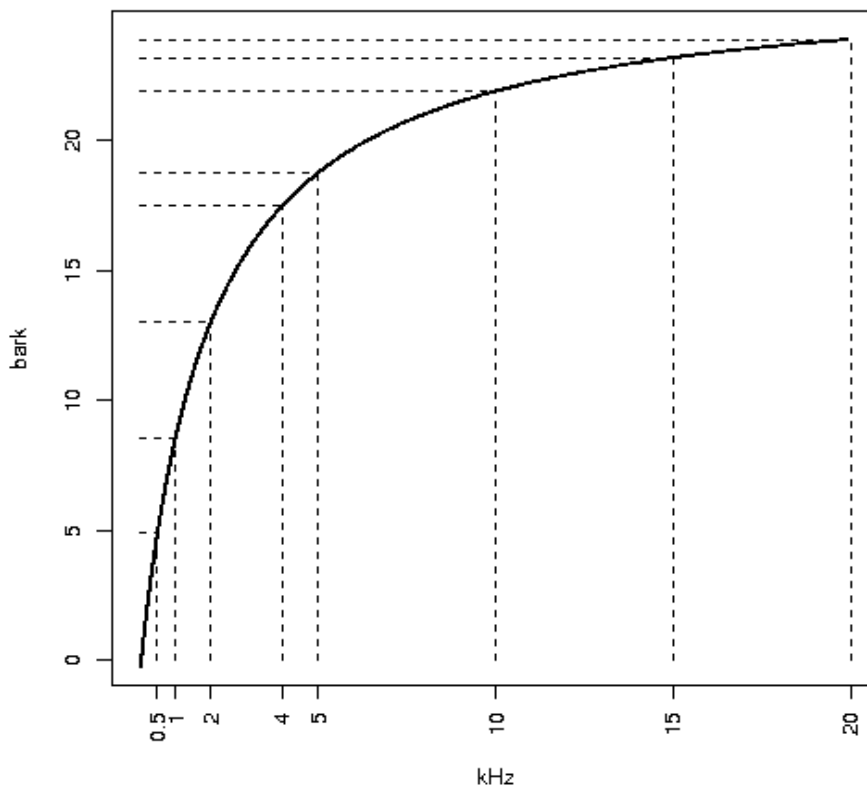
Az ábrán megfigyelhető, hogy az elfedés erőteljesebb a magasabb frekvenciák irányába, mint az alacsonyabbakéba. Ez azzal függ össze, hogy a csiga a magasabb frekvenciákra a bemenethez közelebb reagál, tehát ezek a frekvenciák nem, vagy csupán csekély mértékben befolyásolják a cochlea bemenetétől, azaz az ovális ablaktól távolabbra található alacsonyabb frekvenciák észlelését. Fordított esetben viszont a hanghullám keresztülhalad a membrán magasabb frekvenciák észleléséért felelős területein is, ezért valószínűbb, hogy ezek működését is befolyásolja.

A kritikus sávok megfeleltethetőek a csiga frekvenciafelbontó képességének, és fontos szerepet játszanak a percepcióban: ha ugyanis fülünket egyszerre több hang éri, és ezek egy kritikus sávon belül vannak, akkor intenzitásuk összegződik, de nem észleljük őket különálló hangokként. A kritikus sávok (értékeit ld. Vicsi: *A beszéd akusztikai fonetikai leírása* c. fejezet) szolgálnak a Zwicker által 1961-ben kialakított, Heinrich Barkhausen tiszteletére elnevezett **Bark-skála** alapjául. A Bark a szűrők sorszáma, és az emberi hallásra jellemző alsó 24 kritikus sávnak felel meg 20 és 15 500 Hz között (Fastl & Zwicker 2006, Tarnóczy 1982).

A kritikus sávoknak, valamint a csiga tonotópiás felépítését tükröző hangmagasság-skáláknak igen fontos szerepe van az emberi beszédpercepció folyamatok leírásában, hiszen – ellentétben a frekvencia- és félhangalapú mértékegységekkel – kifejezésre juttatják azt a tényt, hogy az emberi hallás érzékenysége frekvenciafüggő. Eppen ezért a kritikus sávok 500 Hz alatt

6 Kerekített értékek.

nagyjából 100 Hz szélesek, e fölött egyre nagyobb értéket vesznek fel. Ezt szemlélteti a 6. ábra: az x-tengelyen jelölt értékek között egyre nagyobb a távolság, míg a nekik megfelelő Bark-értékek között nagyjából azonos a különbség: 500, 1 000, 2 000, 4 000 és 10 000 Hz hangmagasság-észlelete 4,9; 8,5; 13,0; 17,5 és 21,9 Bark.

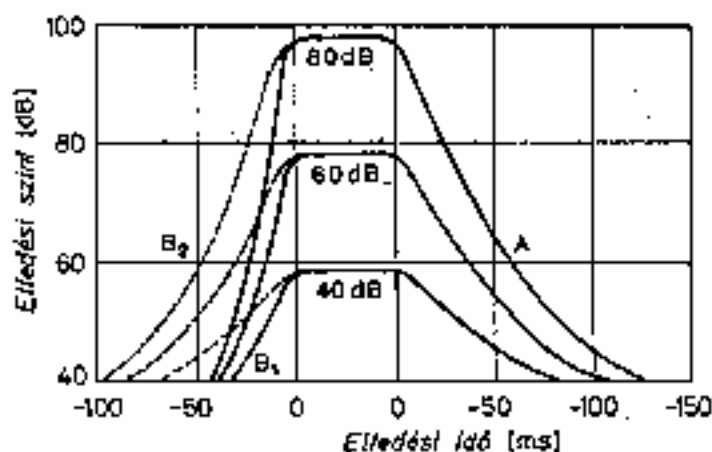


6. ábra: A frekvenciaérték (kHz) és a fül által észlelt hangmagasság (Bark) függvénye.

1.2.3 Az idő szerepe az észlelésben

Az időbeli észlelés szintén frekvenciafüggő. Ha az akusztikai jelbe szünetet iktatnak be, akkor a 2 és 6 kHz közötti tartományban a fül már 2–4 ms közötti szünetet is képes felismerni, míg a minimálisan felismerhető szünet tartama 22 ms körül van.

Az elfedési jelenségnek szintén van időbeli vonatkozása. Ha egy hangingerre 200 ms-on belül újabb, alacsonyabb intenzitású hanginger következik, a második ingert nem észleljük bizonyos időtartam- és intenzitáskülönbség mellett. Ez az ún. **utóelfedés** (*forward masking*). Érdekes módon a jelenség ellenkező irányban is megfigyelhető: ha egy gyengébb ingerre erősebb inger következik, bizonyos körülmények között csak a második ingert észleljük. Az **előelfedés** (*backward masking*) lényegesen kisebb időbeli eltérés esetén lép csak fel, és azzal magyarázható, hogy a nagyobb intenzitású második inger felismerési ideje rövidebb, ezért nyomja el a lassabban észlelt gyengébb ingert. Ha nagyobb a hanginger intenzitása, az elfedés nagyobb időbeli eltérés mellett is kimutatható. Az előelfedés esetén szerepet játszik az is, hogy a hangingert azonos vagy különböző fülön keresztül hallják-e a kísérleti alanyok: az azonos fülön történő bejátszás esetében nagyobb időkülönbség is kiváltja az elfedést (ld. 7. ábra).



7. ábra: Időbeli elfedés 40, 60, és 80 dB-es fehér zaj esetén. A: utóelfedés hatása, B₁: különböző fület érő hangok, B₂: azonos fület érő hangok (ld. bővebben Tarnóczy 1984.)

2 Szegmentális és szuprasegmentális elemek percepciója

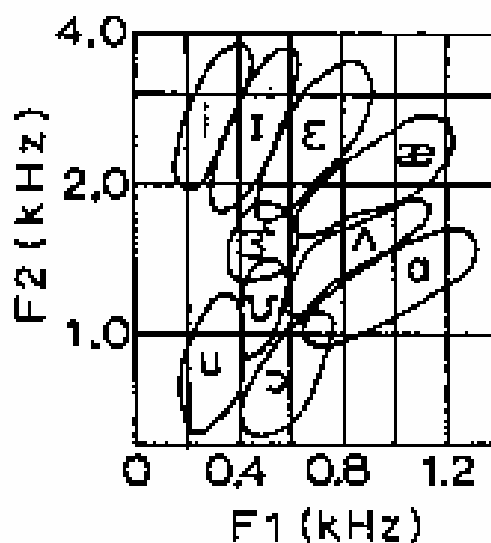
A percepciókísérletek alapjául **természetes** vagy **szintetizált** beszédet használhatunk fel. A természetes beszéd hátránya, hogy akusztikai jellemzőit nehéz kontrollálni, a szintetizált beszéd viszont nem feleltethető meg egyértelműen az emberi beszédnek. Sok kísérletben a kétféle módszer egyesítésével próbálkoznak, azaz természetes beszédhangokat módosítanak pl. idő-, frekvenciaszerkezetüket vagy intenzitásukat tekintve, így vonva részleges ellenőrzés alá akusztikai sajátosságaikat. Ez a módszer mindkét típus előnyeit és hátrányait egyesíti. A percepciókísérletek örök dilemmája éppen ezért az, hogy a kísérletek eredményei mennyire érvényesek a természetes körülmények között zajló beszédészlelésre. Könnyen elképzelhető ugyanis, hogy ha egy kísérlet arra „tanít”, hogy egyes akusztikai jellemzőknek különleges figyelmet tulajdonítsunk, akkor ezek jelentősége megnő, miközben a mindennapi beszédértés során messze nem akkora ezen jellemzők szerepe. A laboratóriumi körülmények között végzett kísérletek eredményeit tehát mindig óvatosan kell értelmeznünk.

2.1 Magánhangzók

A hangokat fonológiai szinten a generatív fonológiához kapcsolódva gyakran bináris szempontok alapján szokás jellemezni: hosszúság, ajakkerekítés, feszesség, zöngesség, zöreinesség stb. Ezeket a megkülönböztető jegyeket általában nem lehet egyetlen artikulációs vagy akusztikai mérőszámmal megelégedni, és ez különösen igaz a magánhangzókra. A hosszú magánhangzók akusztikailag mérhető tartama pl. nem feltétlenül hosszabb, mint a rövid magánhangzóké, azonosításuk ettől függetlenül – más kulcsok alapján – többnyire nem okoz problémát. Artikulációs szempontból sem igaz, hogy a hosszú magánhangzók esetén a hangképző szervek hosszabban maradnak ugyanabban a pozícióban, hiszen a magánhangzók képzésében igen kevés az állandó elem, ezen hangok fő jellemzője ugyanis épp dinamikus voltak. A hosszú és rövid magánhangzók közötti eltérés mibenlétét tehát csak nagyobb összefüggések vizsgálatával deríthetjük ki.

A magánhangzók jellemzésére leggyakrabban tartamukat (azaz mérhető hosszukat), az alaphangfrekvenciát (f_0), és az első két formánst (F_1 és F_2) szokás használni (ld. Vicsi: *A beszéd akusztikai fonetikai leírása* c. fejezet), továbbá a felsőbb formánsokat ($F_3 - F_5$), az intenzitást és egyéb akusztikai jegyeket. A mérések céljától függően ezek az értékek mérhetőek adott pontokon (leggyakrabban a magánhangzó középpontjában), vagy a hang teljes tartama

alatt, ha a dinamikus tulajdonságokra vagyunk kíváncsiak. Egyes jellemzők, mint a tartam, alapfrekvencia, F_1 és F_2 könnyen értelmezhető fonetikai jegyeknek felelnek meg: a tartam a hosszúságnak, az f_0 a hangmagasságnak, az F_1 a függőleges képzéshelynek, az F_2 pedig a vízszintes képzéshelynek. A magánhangzók felismerését vizsgáló kísérletek egy része azt példázza, hogy a fenti jegyek, vagy akár csak egy részük is elegendő az azonosításhoz, míg mások további jegyek (energiakontúr, alaphang és egyes formánsok közötti távolság) fontosságát hangsúlyozzák. A magánhangzó-felismerésre vonatkozó legismertebb kísérletet Peterson és Barney végezte. Tíz szót rögzítettek összesen 76 személy (férfiak, nők és gyerekek) kétszeri ejtésében: *heed, hid, head, had, hod, hawed, hood, who'd, hud, heard*⁷. A hívószavakat egy nyolc napon át tartó kísérletben összesen 70 kísérleti személy hallgatta meg, akik részben azonosak voltak a beszélőkkel (Peterson & Barney 1952).



8. ábra: amerikai angol magánhangzók formánseloszlása 76 beszélő ejtésében (x-tengely: F_2 , y-tengely: F_1). (Forrás: Kent 1996:337.)

A résztvevők az összesen 1520 stimulust 94%-ban helyesen ismerték fel, annak ellenére, hogy a magánhangzók 1. és 2. formánsukat tekintve számos átfedést mutattak (ld. 8. ábra). Tudvalevő, hogy a férfiak, nők és gyermekek toldaléksöve eltérő méretű, ezért az általuk ejtett magánhangzók abszolút formánsértékei is különböznek (arányukban viszont megegyeznek). Ezért megvizsgálták a felnőtt férfiak által ejtett magánhangzókat is, és kizárták azokat, amelyeket a kísérleti személyek nem tudtak egyöntetűen azonosítani. Még ekkor is voltak átfedések a gond nélkül azonosított magánhangzók között. A szerzők ezért feltételezik, hogy az abszolút formánsértékek nem elegendőek a magánhangzó osztályának meghatározásához.

A kísérlet két további fontos felismeréssel járt. Egyfelől igazolta a produkció és percepció összefonódását a hangok szintjén. Azok a kísérleti személyek ugyanis, akik két különböző fonémát egyforma hangként realizáltak, a percepció feladatban az ezekhez a fonémákhoz tartozó realizációkat nem tudták egymástól megkülönböztetni.

A kísérlet másfelől rámutat arra, hogy a magánhangzók azonosításának sikere függ az adott nyelv magánhangzórendszerének szerkezetétől. Peterson & Barney kísérletében az /i/ hangot 93%-ban helyesen azonosították a kísérleti alanyok, az /a/ hangot azonban csupán 6%-ban. Ezt egyrészt az /a/ megvalósulásainak nagy szórása magyarázza (az 1. formáns értéke 600 Hz-ről akár 1200 Hz-re nőhet a kategórián belül), másrészt hogy az /a/-nak több szomszédja van, mint pl. az /i/-nek, ezáltal nő a tévesztés veszélye is. (A magyar magánhangzórendszerben ugyanezen okból az /a:/ azonosítása a legkönnyebb, ld. Mádý &

⁷ Ezek a szavak az angolban egységesen /hVd/ szerkezetűek, azaz csak a magánhangzóban különböznek egymástól.

Reichel 2007.)

Későbbi kísérletek bizonyították, hogy a magánhangzó-felismerés nem abszolút, hanem relatív frekvenciaértékeken alapszik. Miller (1953) szintetizált hangokkal kimutatta, hogy a magánhangzó nyíltsági fokát az alapfrekvencia döntően befolyásolja (Miller 1953). Ha ugyanis az alapfrekvenciát egy oktávval felemelte, a magánhangzót a kísérleti személyek – azonos formánsértékek mellett – zártabb hangzóként érzékelték. További fontos felismerések adódtak Bark-alapú számításokból. Syrdal & Gopal Peterson & Barney kísérleti anyagát újraelemelve megállapította, hogy ha a magánhangzók feltérképezésekor Bark-értékeket vesznek alapul, valamint figyelembe veszik az f_0 és F_1 , valamint az F_1 és F_2 távolságát, akkor a hangok szórása nagyban csökkenthető, és a kategorizálás az alapfrekvencia és a két alsó formáns alapján elvégezhető (Syrdal & Gopal 1986). Traummüller rámutatott, hogy a nyíltsági fok észlelését döntően az f_0 és az F_1 Bark-ban mért távolsága befolyásolja (Traummüller 1981). Chistovitch szerint azokat a formánsokat, amelyeknek távolsága 3-3,5 Bark-nál kisebb, egyetlen formánsként érzékeljük, ennek értéke a két abszolút formánsérték spektrális súlypontjába esik (Chistovitch 1985).

Lehiste & Peterson (1961), valamint Strange (1989) a magánhangzó-azonosításban kiemelik a dinamikus jellemzők szerepét, különös tekintettel a rövid-hosszú, ill. feszes-laza opozícióra. Eszerint a magánhangzók azonosításában nemcsak a középső, viszonylag állandó szakasz formánsszerkezete játszik szerepet, hanem a magánhangzót megelőző, ill. azt követő mássalhangzó felé való átmenetek időbeli aránya is. Eszerint a hosszú magánhangzók állandó szakasza relatíve hosszabb, mint a rövid magánhangzóké. Strange kísérleti eredményei alapján amellett érvel, hogy a magánhangzó-hosszúság megállapításához elegendő a dinamikus szakaszok frekvenciaszerkezetének és tartamának ismerete.

A magánhangzók azonosításában egyes modellek az extrinzikus, azaz külső jegyek fontosságát hangsúlyozzák, amelyek elősegítik a képalkotást a beszélőről, s ezzel a jobb magánhangzó-felismerést. Ez lehet az f_0 (hiszen az alapfrekvenciából következtethetünk a beszélő nemére és korára), a egyéni sajátságokat hordozó F_3 , a kísérletben vizsgált magánhangzót körülvevő egyéb nyelvi információ (pl. a mondat, amelybe a hívószót beágyazták). A külső jegyek szerepét bizonyítja, hogy a célszó magánhangzójának kategorizálása változik, ha a hordozó mondat magánhangzóinak formánsértékét manipulálták, valamint hogy a magánhangzó besorolása befolyásolható az egyidejűleg vizuálisan prezentált „beszélő” nemén keresztül. (Az intrinzikus és extrinzikus modelleket átfogóan ismerteti Strange 1999b.)

2.1.1 Vizsgálatok a magyar magánhangzók percepciójáról

A beszédpercepció modellek főként angol nyelvű vizsgálatokra épülnek, amelyeket Gósy Mária számos munkájában adaptált magyar anyanyelvűek észlelésére. A következőkben az ő kísérleteiből említünk néhányat, a teljesség igénye nélkül.

Gósy eredményei szerint a magyar magánhangzók azonosításához nem feltétlenül szükséges a második formáns megléte, sőt esetenként még az elsőé sem (Gósy 1989). Ezt egy olyan vizsgálatban mutatta ki, amelyben különböző szűrőkön áteresztett, izolált magánhangzók felismerését vizsgálta. Bár a rövid /a/ első formánisa 600 Herz volt, a hangot a kísérleti alanyok akkor is 70%-os biztonsággal felismerték, ha a lejátszott hang kizárólag a 390 Hz alatti frekvenciatartományból állt. Más magánhangzók, mint pl. az /o/ hang hasonlóan jó felismeréséhez viszont a szűrt hangnak magasabb frekvenciákat is kellett tartalmaznia, jóval az /o/ első formánisa feletti tartományból. Gósy szerint a magánhangzók viszonylag biztos azonosításához szükséges frekvenciatartomány felső határa és a magánhangzó F_2 értéke közötti távolság specifikus, tehát magánhangzónként eltérő: az /a:/ esetében 100 Hz, az /i:/ esetében viszont 1 200 Hz. Ez utóbbi magánhangzó alapján Gósy arra a következtetésre jut, hogy az első formáns tartalmaz a második formánsra utaló felismerési kulcsokat, ill. hogy az első két formáns közötti – a formánsokénál alacsonyabb intenzitású – frekvenciasávok is

⁸ Gósy a rövid /a/ hang fonetikai átírására – számos magyar fonetikussal egyezően – az [O] jelet használja. Mivel az olvasók előtt ismert a magyar magánhangzók minősége, ebben a leírásban a hét osztály és a két hosszúság jelölésére szorítkozunk, függetlenül attól, hogy a rövid és hosszú magánhangzók akusztikailag megfelelnek-e egymásnak.

szerepet játszanak a magánhangzó felismerésében.

Hasonló eredmények születtek a kísérlet azon részéből, amelyben a sáváteresztő szűrő alsó határát 270 Hz és 2 200 Hz között mozgatták, felső határa pedig 2700 Hz volt. A magánhangzók azonosítása sok esetben akkor is sikeres volt, ha az első és a második formáns egyaránt az áteresztett tartomány alatt volt található. Valószínű tehát, hogy a frekvenciaeloszlás sajátosságai, valamint a felsőbb formánsok értéke jellemző az egyes magánhangzókra. Ezek jelentősége azonban csak akkor lesz nyilvánvaló, ha az elsődleges kulcsok (tehát az első két formáns) hiányoznak. Ha ugyanis az F_1 és F_2 megléte mellett az F_3 értékét manipuláljuk, a magánhangzó azonosításában nem történik változás (Gósy 1989). A redundáns jegyek, azaz a másodlagos felismerési kulcsok csak az elsődleges kulcsok hiánya esetén lépnek előre a felismerési hierarchiában.

A kísérlet harmadik részében a felül áteresztő szűrő alsó határa 1 000 Hz és 3300 Hz között mozgott. A legjobb az /i/ és /y/ hangok felismerése volt, hiszen ezen magánhangzók második formánisa 2 000 Hz felett található, ezért a felsőbb formánsok beleestek az áteresztett tartományba. Gósy eredményei arra utalnak, hogy a kerekített és kerekítetlen hangok közötti különbségtételt a részösszetevők eltérő intenzitása teszi lehetővé, hiszen a kerekítetlen /i/ intenzitása általában véve nagyobb, mint a kerekített /y/ hangé.

Gósy egy másik kísérletében (Gósy 1989) a magánhangzók hosszúsága és minősége közötti összefüggést vizsgálja. Ehhez szintetizált, 50 és 400 ms közötti hosszúságú hangokat használ, amelyek első és második formánisa egy skála mentén változik (az alaphfrekvencia és a felsőbb formánsok értéke változatlan). Eredményei szerint egy adott F_1 érték rövidebb magánhangzó esetén zártabb, hosszabb magánhangzó esetén nyíltabb kategóriába való sorolást eredményez, így pl. az 500 Hz körüli első formánsokat hol /o/-ként, hol /a/-ként hallották a kísérleti alanyok, azonban ha a hang rövid volt, nagyobb volt az /o/-válaszok száma, a hosszabb hangok esetében pedig az /a/ válaszok aránya. Ez azt mutatja, hogy a hallgató a percepció során beépíti azon ismeretét a kategorizálásba, hogy egy rövid magánhangzó képzésekor a beszélő nehezebben produkál nagyobb ajaknyílást, éppen ezért ergonómiai szempontból valószínűtlenebb, hogy egy /o/ képzésekor egy ilyen, viszonylag magas F_1 -értéket produkáljon. Ez a felismerés Liberman motoros elméletét támasztja alá (ld. 3.1).

2.2 Mássalhangzók

A magánhangzók artikulációs jellemzői egy-egy faktor mentén, folyamatosan változnak: az /o/ nyíltabb, mint az /u/, de kevésbé nyílt, mint az /a/. További nyíltsági fokok e három magyar magánhangzó között is léteznek, amint erre számos nyelvből ismerünk példát.

Ugyanez nem érvényes a mássalhangzókra. Bár a képzési helyek elvileg szintén kontinuumot képeznek, legalábbis a fogmedertől a lágy szájpaddlásig, a gyakorlatban a legtöbb nyelvben jól elkülöníthető képzési helyekkel találkozunk, a magyar zárhangok esetében pl. labiális (/p,b/), alveoláris (/t,d/), palatális (/c,/) és veláris (/k,g/) hangzókkal. E képzési helyeken felül ugyan elvileg elképzelhetőek köztes zárhangok is, ám ezekkel nem, vagy igen ritkán találkozunk más nyelvekben.

A mássalhangzók esetében felvetődik két további probléma is: egyrészt a mássalhangzók sok esetben nem vizsgálhatóak közvetlenül, csupán a szomszédos magánhangzókra keresztül (de ld. Vicsi 1981b). Másrészt míg a magánhangzók esetében az artikulációs és akusztikai jegyek között viszonylag egyértelmű összefüggéseket láttunk, a mássalhangzókat a szomszédos magánhangzók függvényében merőben eltérő megvalósulások jellemezhetik. Így pl. /di/ hangkapcsolatban a /d/ felpattanását követően a rá jellemző F_2 -tranzíció⁹ 2 kHz-től emelkedik az /i/ 2,4 kHz körüli formánsáig, /u/ előtt viszont az átmenet 1,2 kHz-nél kezdődik és 600 Hz-ig, az /u/ F_2 -jéig esik. Az F_2 -átmenet tehát sem menetében, sem kezdőértékében nem mutat hasonlóságot, mégis mindkét esetben /d/-t hallunk (ld. 6. ábra). A jelenség magyarázatára született a **lokuszelmélet**, amely szerint a CV-átmenet képzeletbeli kezdőpontja mindkét esetben ugyanaz, alveoláris hangok esetében 1800 Hz, de a formáns menete csak a magánhangzó kezdetével válik láthatóvá, ezért térnek el a mérhető kezdőpontok egymástól.¹⁰

⁹ A *tranzíció* elnevezés az angolban használatos *transition* kifejezés magyarítása. Magyarul átmeneti vagy tranziens fázisként is szokás emlegetni.

¹⁰ Többen kétségbe vonják, hogy a mássalhangzók képzési helye egyértelmű lokuszdefiniciókat tesz

Feltehetően ezzel függ össze az is, hogy az alveoláris /n/ hajlamosabb az asszimilációra, mint a más képzéshelyű nazálisok: ha egy alveoláris hang /E/-hez kapcsolódik, akkor a formánsátmenet stagnáló menetű, márpedig a változást (tehát az emelkedő vagy eső tranzíciót) az emberi agy könnyebben érzékeli, mint a változatlanúságot. Az /E/ sok nyelvben gyakori magánhangzó, az alveoláris mássalhangzók előfordulása általában szintén magas, így értelemszerűen a kapcsolódásuk is átlagon felül fordul elő. Greenberg ezzel magyarázza, hogy az alveoláris /t/, /d/, /n/ az amerikai angolban szótagvégi helyzetben igen gyakran nem realizálódik (Greenberg et al. 2003).

A mássalhangzók percepciójának leírását nehezíti, hogy fonológiai jegyeiket illetően meglehetősen eltérnek egymástól. Más jegyek szükségesek a nazálisok, likvidák vagy réshangok felismeréséhez, mint a zárhangokéhoz. Ezért a következőkben csak olyan kísérletek ismertetésére szorítkozunk, amelyek valamely jelentős percepciós elmélet kialakításához járultak hozzá, vagy azt cáfolják. A percepciós elméletek részletesebb ismertetésére a következő részben térünk vissza. További kísérletek leírását ld. Strange (1999a).

Az Egyesült Államok-beli Haskins laboratóriumban az 50-es évek óta számos kísérlet született a mássalhangzó-kategóriák észlelésével kapcsolatban. Úttörő szerepe volt Liberman et al. (1957) vizsgálatának, amelyben zöngés zárhang és /e/ magánhangzó kapcsolatát modellezték szintetikusán, úgy, hogy az F_2 -átmenet a /b/-re jellemző értéktől a /d/-n keresztül a /g/ értékéig fokozatosan emelték (120 Hz-es lépésekben), így összesen 14 F_2 -átmenetet kaptak. Az F_1 menetét nem változtatták. A kísérleti személyek a várához való megfelelés az alacsonyabb értékről kezdődő átmenetek kezdőhangjaként /b/-t, majd /d/-t, a felső értékek esetén pedig /g/-t adták meg. A kutatókat is meglepte, hogy az ítéletek igen egyöntetűek voltak, és a fonémák közötti váltás egyértelműen bizonyos jelek közötti tartományhoz kötődött. Egy diszkriminációs kísérletből az is kiderült, hogy a fonémakategóriák közötti különbségtétel lényegesen könnyebb feladat, mint az azonos kategóriához tartozó hangok azonosságának felismerése.

Hasonló eredményre jutott Lisker & Abramson (1967) a zárhangok zöngességét vizsgálva. CV-kapcsolatokban az alsó három formáns egyidejű megjelenését 0 ms-nak véve a zöngékezdési időt (*voice onset time*, VOT) szintetizált anyagon -150 ms-tól 150 ms-ig változtatták, 10 ms-os lépésekben. A kísérleti személyek viselkedése ebben az esetben is egyöntetű volt: a hangkapcsolatok mássalhangzóját -150 és +20 ms zöngékezdési idő között zöngésnek, +30 és +150 ms között zöngétlennek ítélték. Az előző kísérlethez hasonlóan a diszkriminációs tesztben sokkal nagyobb volt a helyes válaszok aránya 20 és 30 ms között (ahol az összehasonlítható jelek mássalhangzói különböző kategóriákhoz tartoztak), mint abban a tartományban, ahol a szomszédos VOT-értékek egyazon kategóriához tartozást sugalltak.

A produkciót és percepciót összekötő motoros elmélet egyik fontos állítása, hogy a beszédpercepció hátterében egy speciális nyelvi modul áll. A heterogén kísérleti eredmények fényében nem egyszerű sem a bizonyítás, sem a cáfolat. Egyrészt több kísérletben is igazolták, hogy a *b-d-g* elkülönítését szolgáló F_2 -értékek nem-nyelvi jelek esetében nem hívják elő kategorizálást, ami a speciális nyelvi modul meglétét igazolja. Ezzel szemben Pisoni (1977) kimutatta, hogy a VOT kapcsán megfigyelt 20-30 ms határ nem-nyelvi jelek diszkriminációjában is megfigyelhető, ezen jelenség mögött tehát feltehetően általános pszichofizikai összefüggések állnak.

2.2.1 A magyar mássalhangzók percepciója

Az itt bemutatandó kísérletek arra keresik a választ, hogy az angol anyanyelvű hallgatók eredményei alapján kialakított elméletek mennyiben érvényesek magyar anyanyelvű kísérleti alanyokra, és alkalmasak-e magyar anyanyelvűek beszédértésének leírására is.

Gósy (1989) a magyar mássalhangzók felismerését szintetizált anyaggal tesztelte (9. ábra). Anyagában a /b/, /d/, /g/ mássalhangzók, valamint különféle magánhangzók szerepeltek CV-kapcsolatokként. A felpattanás pillanatában és a CV-átmenetben az F_2 értékét fokozatosan változott. Gósy eredményei szerint a /b/ felismerésének elsődleges kulcsa a második formáns

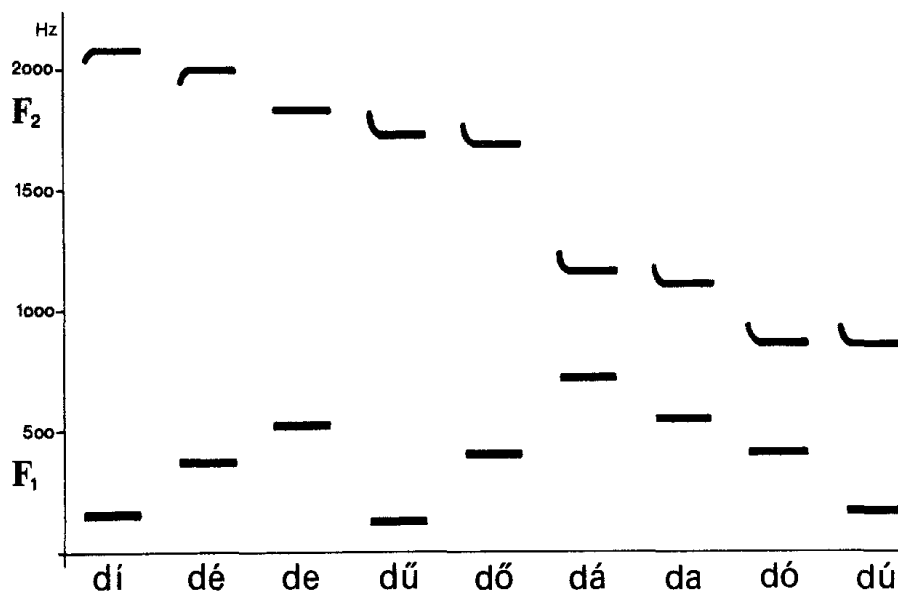
lehetővé, ld. pl. Lehiste & Peterson (1961), más kísérletekben viszont igazolták a lokuszelmélet helyességét. A kérdésben nincs egyértelmű konszenzus, bár a lokuszértékeket közelítő jelleggel mindenki elfogadja.

frekvenciájában rejlik, amely azonban egy viszonylag nagy, 1 000 Hz-t átölő tartományba eshet (800–1800 Hz). A klasszikus lokuszelméletnek megfelelően az F_2 kontúrja eltérő a palatális és veláris magánhangzók esetén: az itt közölt kísérleti adatokban veláris magánhangzók előtt emelkedő vagy eső is lehet, a palatális magánhangzók esetében pedig emelkedő. A magánhangzó és mássalhangzó F_2 értékének azonossága csak a hátul képzett magánhangzók esetében fordul elő.

/d/-t követően a magánhangzók átmenete lehet emelkedő és eső is (értéke 900 és 2200 Hz között mozog), és az átmenet itt is eltér a magánhangzó palatális vagy veláris voltától függően, mert a veláris magánhangzók esetében az F_2 frekvenciaértéke mellett az intenzitásvizonyok is szerepet játszanak a felismerésben. A CV kapcsolat tagjainak azonos F_2 értéke csak a kerekítetlen magánhangzók esetében vezet helyes felismeréshez.

A veláris /g/ akusztikai szerkezete összetettebb a másik két zárhangénál. Az F_2 értéke (ami 1740 Hz-nél nagyobb) csak az elől képzett magánhangzók esetében befolyásolja az azonosítást, együtt az intenzitásvizonyokkal, az F_1 kontúrával és az időviszonyokkal. A hátul képzett magánhangzók azonosítására csupán az utóbbi három jegy van hatással.

E rövid összefoglalásból kiderül, hogy a vizsgált három zöngés zárhang felismerési kulcsai nem azonosak: a /b/ azonosításában kizárólag az F_2 játszik szerepet, a /d/ felismeréséhez az intenzitásnak is megfelelő értékeket kell felvennie, míg a /g/ felismerési kulcsai komplexebbek: szerepet játszik a temporális szerkezet, valamint az F_1 mozgása is.



9. ábra: /dV/ formánsátmenetek a követő magánhangzó függvényében. (Forrás: Gósy 1989:62.)

A mássalhangzók képzésmódjuktól függően temporális szerkezetükben is eltérnek egymástól. Vicsi (1981a) és Gósy (1989) kimutatta, hogy a réshangok, affrikáták és zárhangok besorolása nagyban függ a hang relatív hosszától is. Ha az /s/ hangot rövidítjük, a környező hangok azonossága mellett, a percepció előbb /ts/-be, majd /t/ be csap át (hiszen ezen hangok képzéshelye és zöngéssége egyébként megegyezik). Érdekes, hogy Gósy (1989) tanúsága szerint az azonosítás függ attól, értelmes szót vagy hangkapcsolatot kell-e felismerni. Kísérleti alanyai egy kb. 100 ms-ra rövidített /s/ hangot 70%-ban /s/-ként azonosítottak, ha a *szél* szóban kellett a hangot felismerni, s csak az esetek 30%-ban hallottak *cél*-t. Az /s/+e:/ hangkapcsolatot viszont 80%-ban *cé*-ként azonosították. A felismerést nemcsak a magánhangzó hossza, hanem a szomszédos magánhangzó minősége is befolyásolja, méghozzá a magánhangzók intrinzikus, azaz rájuk jellemző tartamától függően. A magyarban a legelső nyelvvállású /a:/ a leghosszabb magánhangzó, ezért e hang mellett még egy viszonylag rövid /s/ hangot is /s/-ként érzékelünk, míg ugyane hangot más magánhangzó mellett /ts/-nek hallanánk.

Igen meglepő eredményre jutott Vicsi Klára, amikor a magyar zárhangok felismerését tesztelte (Vicsi 1981b). Azt találta ugyanis, hogy szemben a korábbi eredményekkel, a magyar beszélők számára nem a szomszédos magánhangzó tranzienst eleme a felismerési kulcs, hanem a felpattanással járó zörej szerkezete. A kísérleti személyek CV-kapcsolatokból kivágott, egységes tartamú szeleteket hallgattak meg. Ha a jel a magánhangzó zöngés periódusával kezdődött, tehát a teljes tranzíciót magában foglalta, de a mássalhangzót nem, a mássalhangzó felismerése igen bizonytalan volt. A mássalhangzó zörejét tartalmazó jel viszont ugrásszerűen megnövelte a helyes találatok számát. Ugyanez volt érvényes VC-kapcsolatokra is. Vicsi feltételezése szerint a korábbi vizsgálatokkal ellentétes eredmény vagy arra vezethető vissza, hogy a magyar zárhangok nem aspiráltak, ellentétben a korábbi kísérletek alapjául szolgáló nyelvekkel, vagy arra, hogy a kísérlet nem szintetizált, hanem természetes beszédjelle épült.

2.3 A szupraszegmentális jegyek percepciója

Míg a beszédhangok – absztrakt egységként – diszkrét, azaz egymástól elkülöníthető egységeket alkotnak, a prozódiai jegyek (ld. Vicsi: *A beszéd akusztikai fonetikai leírása c. fejezet*) csak a beszédfolyamat egészét tekintve elemezhetőek. Egy mondat hanglejtését például csak a mondat egészének ismeretében értelmezhetjük kérdő, óhajtó vagy egyéb modalitás kifejezésekként. Mivel a szupraszegmentális jegyekről nehéz általánosságban beszélni, hiszen a jelenségek nyelvenként nagyon eltérőek lehetnek, a következőkben néhány kísérlet ismertetésén keresztül mutatjuk be a percepció fonetika e szegletét.

2.3.1 Mondathangsúly

Gósy kísérletében azonos tartalmú és ejtésű mondat tempójának megítélését vizsgálja az artikulációs tempó (hang/másodperc) és a hanglejtés függvényében (Gósy 1989). A kísérletből kiderül, hogy a monoton (tehát sem emelkedő, sem ereszkedő hanglejtést nem tartalmazó) mondatokat a kísérleti alanyok gyorsabbnak "hallják", mint az intonációs mozgást tartalmazó mondatokat. A leglassabb besorolást az ereszkedő lejtésű mondatok kapták, míg az emelkedő és a szólameleji, valamint szólamvégi csúcsot tartalmazó mondatokat azonos sebesség esetén a kísérleti személyek inkább hajlamosak voltak gyorsként értékeln. A monoton hanglejtés gyors beszédként való értékelése nem minden esetben, csupán a normálnál gyorsabb mintánál volt megfigyelhető, tehát ott, ahol az értelmezési kulcsok a gyorsabb tempóból következően nehezebben voltak felismerhetőek.

Gósy egy további kísérletében 32 mondat dallamának felismerését vizsgálta egy percepció kísérletben (Gósy 1989). A résztvevők feladata először az volt, hogy rajzolják le a hallott mondat dallamát. A következő eredmények születtek: (1) A kísérleti alanyok pontosan ábrázolták a mondatok hosszát. (2) A csúcsok mondatbeli helyzete és frekvenciája esetleges az eredetihez képest. (3) A hangterjedelem jelölése következetes. (4) Ha a mondat eleje szökő és ereszkedő ágat egyaránt tartalmaz, az alanyok fele gyakran csak az ereszkedést érzékeli. Ha viszont a csúcs a mondat elején helyezkedik el, annak érzékelése 90% körüli pontosságú. A mondat közepén vagy végén található csúcsok érzékelése szintén pontos. (5) A mondatvég intonációjának érzékelése ereszkedő dallam esetén megbízható, emelkedő dallam esetén viszont a mondat egészének hanglejtésétől függ, csak úgy, mint a lebegő dallamok felismerése: ha a mondat elején nagyobb csúcs található, valószínűbb az ereszkedőként való érzékelés.

A kísérlet második részében a résztvevők feladata az volt, hogy határozzák meg a hallott mondat nyelvtani-pragmatikai funkcióját. Gósy a következőket állapítja meg: (1) A szólameleji csúccsal rendelkező mondatok megítélésében a kiegészítendő kérdésként és a kijelentésként való besorolás dominál, részben függetlenül a mondat eredeti tartalmától. (2) A kisebb frekvenciájú szólameleji csúcsot tartalmazó mondatokat elsősorban kiegészítendő kérdésként hallották a résztvevők.

Gósy kimutatta továbbá, hogy a kísérleti alanyok a dallamoknak akkor tulajdonítanak emocionális töltetet, azaz akkor hallják őket felkiáltásnak, ha a dallam csúcsa magasabb frekvenciájú, vagy ha a csúcs a mondat nagyobb hányadában magas frekvenciájú marad. A mondatmodalitás tehát relatív, nem pedig abszolút paraméterekben manifesztálódik.

A magyar eldöntendő kérdések intonációja jellegzetes, sok nyelvtől eltérő, leírásával

számos munka foglalkozik (ld. pl. Ladd 1996). Prototipikus megvalósulása az utolsó előtti szótag magasabb frekvenciáját foglalja magába. A Gósy által vizsgált anyagban az eldöntendő kérdések arról voltak felismerhetőek, hogy a mondat második felében különböző frekvenciájú csúcsok voltak találhatóak. Ezeket a mondatokat a kísérleti alanyok az esetek 100%-ában helyesen ismerték fel, tehát az eldöntendő kérdések intonációja igen erős felismerési kulcsokat hordoz magában. Egyes eldöntendő, illetőleg befejezetlen kérdések (mint pl. *És erre a tanár?*) a dallam végén emelkedéssel fejeződtek be, ezeket – a köznyelvi intonációnak megfelelően – nagyrészt kiegészítendő kérdésként értékelték a résztvevők.

A kísérlet összesítéséből kiderül, hogy a kiegészítendő kérdések helyes besorolása az esetek 52%-át, a kijelentések helyes besorolása az esetek 75%-át teszi ki. A felszólításokat nagyrészt kijelentésként, a felkiáltásokat pedig főként felkiáltásként vagy kiegészítendő kérdésként értékelték a kísérleti alanyok.

2.3.2 Szóhangsúly

A magyar nyelv jellegzetessége, hasonlóan a rokon finnugor¹¹ nyelvekhez, hogy a szóhangsúly kötött, mindig a szó első szótagjára esik. Ez nagyban befolyásolja a hangsúlyérzékelést, szemben az olyan nyelvekkel, amelyekben a hangsúly más szótagokra is eshet, és helyzete jelentésmegkülönböztető szerepet is betölthet. A magyar hangsúly első vizsgálata Fónagy Iván nevéhez kötődik (Fónagy 1958). Eszerint a hangsúlyrealizáció elsődleges kulcsa az intenzitás növekedése. Gósy egy későbbi kísérletében a mondathangsúly, az alapprofundum-változás és az intenzitás összefüggését vizsgálta (Gósy 1989). Kísérletében azonos részekből álló mondatokat vizsgált, amelyekben a hangsúly különböző szavakon volt: *A városban délben harangoznak. A városban délben harangoznak*, vagy amelyek modalitása eltért egymástól (pl. eldöntendő kérdés). A kísérlet anyaga egyrészt glottográffal rögzített mondatokból állt, amely kizárólag az f0 változásait, azaz a hangmagasságot rögzíti, másrészt imitátorral készült mondatokból, amelyekben az intenzitás is megjelenik. A kísérletből kiderül, hogy a mondathangsúly érzékelése kizárólag az f0 alapján is lehetséges, de lényegesen kevésbé megbízható, mint amikor az intenzitás is rendelkezésre áll. Az intenzitás hiánya ugyan bizonytalanságot eredményez, de nem vezet téves hangsúlyítéletekhez, legfeljebb a fő- és mellékhangsúly felcseréléséhez. Viszonylag kicsi az olyan esetek száma (20%), ahol a nagyobb intenzitás hiányzó frekvencianövekedés mellett is hangsúlyélményt idéz elő.

11 ...és a területi közelség miatt hasonló cseh és szlovák nyelvhez.

3 Percepció elméletek

Az utóbbi néhány évtizedben számos, a beszédpercepciót modellező elmélet született, amelyek részben igen eltérő irányzatot képviselnek, részben egymásra épülnek. Többségük a percepció alapvető kérdéseit igyekeznek megválaszolni:

- Hogyan képes az agy egymástól eltérő hangokat egy kategóriához tartozóként felismerni? Hogyan tud elvonatkoztatni egyénfüggő jellemzőktől, eltérő beszédhelyzetektől, a hangkörnyezettől?
- A hangok kódolása nem egyetlen hanghoz kötődik, hanem távolabbi szegmenseket is befolyásol. Hogyan vagyunk képesek az ebből adódó átfedések ellenére besorolni a hangokat?
- Mi a percepció alapegysége? A fonéma, a szó, a fonológiai megkülönböztető jegyek, vagy hangkapcsolatok?
- Megmagyarázható-e a beszédészlelés általános pszichoakusztikai ismeretek alapján, vagy létezik egy speciális nyelvi modul?
- Mi a feldolgozás iránya: a hangoktól a magasabb egységek felé (*bottom-up*), vagy a nagyobb egységek észlelésétől a hangok felé (*top-down*)?

Az elméletekről, ill. ezek egy részéről a jelenleginél részletesebb leírást közöl Gósy (1989, 2003), valamint Hawkins (ld. Pickett 1999).

3.1 Motoros elmélet

A motoros elméletet (*motor theory*, más néven motoros beszédmegértési hipotézis vagy motoros teória) fentebb már említettük a mássalhangzók felismerése kapcsán. Kialakulása az 50-es évekre tehető, a Haskins laboratóriumhoz, és elsősorban Liberman kutatásaihoz köthető (Liberman et al. 1967, később módosítása Liberman & Mattingly 1985).¹² Központi kérdése, hogy az akusztikumokban eltérő hangokat (főként a zárhangokat) hogyan vagyunk képesek invariáns kategóriákhoz kötni. Liberman és kollégái a megoldást abban látják, hogy a hallgató az akusztikai jelen keresztül észleli, hogy a beszélő milyen invariáns gesztust célt meg, és ebből az adott körülményeknek megfelelően (hangkörnyezet, beszédstílus stb.) milyen konkrét gesztus jött létre. Más szóval a percepció a produkcióról tárolt ismereteinken keresztül lehetséges, és ennek függvénye. A motoros elmélet ebből következően speciális nyelvi modult feltételez, hiszen a beszédészlelést a nyelvről alkotott ismeretekhez köti. A beszédpercepció és -produkció közös alapegységének az artikulációs gesztusokat tekinti (artikulációs cél, és az eléréséhez szükséges mozdulatok).

A motoros elmélet előnye amellett, hogy a produkciós és percepció folyamatok között szinte egyetlen elméletként átjárást teremt, az, hogy számos akusztikai jegyet egyetlen gesztusban foglal össze. Így pl. a zöngés–zöngétlen kontraszthoz kötődő akusztikai jegyeket, mint VOT, zörejenergia, f0-ingadozás és felpattanási energia, a gégefő zöngképzési gesztusára vezeti vissza.

A motoros elmélet befolyása máig igen nagy, de az elmélet éppen annyira vitatott is. Mellette szól, hogy ha egy beszélő a képzésben nem tesz különbséget két hang között, akkor a percepcióban sem tudja őket elkülöníteni, amint ezt Peterson & Barney kísérletében láttuk, és amint ezt saját magunkon is tapasztaljuk, ha idegen nyelvet tanulunk. Ellene szól viszont, hogy a gyermeki beszédfejlődésben a percepció képességek fejlettsége sokszor előbbre jár, mint a produkciós készség (*Nem Tabi, hanem Tabi!* — mondja a /tr/ hangkapcsolatot képezni még nem tudó gyermek). Másrészt a papagáj beszédét is megértjük, holott a madár más mechanizmusokat használ a beszédhangok előállítására, mint az ember, ezért itt nem beszélhetünk saját tapasztalatról.¹³ A tapasztalati ellenpéldákon túl megkérdőjelezhető az alapegységként meghatározott invariáns artikulációs gesztus is, hiszen az artikuláció közel sem

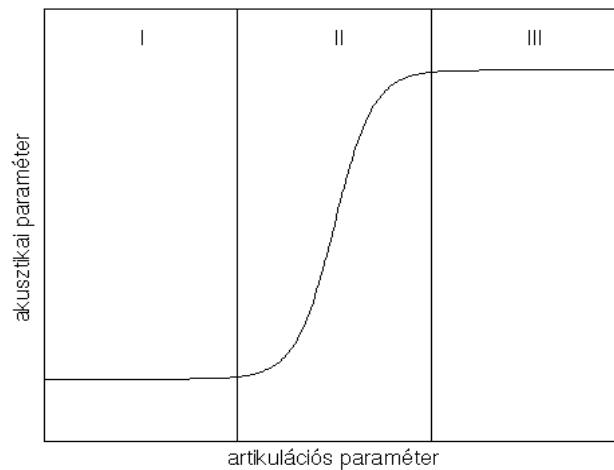
¹² Gósy (1989) szerint Stricker Salamon 1880-ban, majd Sarbó Artúr 1906-ban hasonló nézeteket vallott már.

¹³ A papagájnak két pár hangszalagja van, ezzel imitálja a magánhangzók első és második formánsát.

invariáns, sem végeredményét, sem a neuromuszkuláris folyamatokat tekintve. Egyesek az artikulációs gesztus fogalmát az absztraktba való menekülésnek tartják, hiszen a gesztusok kísérleti úton nem tesztelhetők. Az sem tisztázott, hogy a hallgató az általa közvetlenül észlelt variáns akusztikai jelből hogyan tud következtetni a megcélzott invariáns gesztusra.

3.2 Kvantális elmélet és LAFF

A motoros elmélettel egyidőben jött létre az Egyesült Államok-beli Massachusetts Institute of Technology (MIT) fonetikai laboratóriumában az ún. kvantális elmélet, amely Stevens nevéhez fűződik (Stevens 1989). Az elmélet szerint az artikuláció és az akusztika, valamint az akusztika és percepció közötti összefüggés nem lineáris, hanem kvantális (ugrásszerű), ennek köszönhetően az akusztikai jelben invariáns elemek találhatóak, amelyek megkönnyítik ezek percepcióját, az észlelés tehát kategoriális (10. ábra).



10. ábra: Egy adott artikulációs és akusztikai paraméter összefüggése. Az I. és III. szakaszban a beszélszervek helyzetének változása lassú változást idéz elő az akusztikumban, míg a II., átmeneti szakaszban hirtelen változást vált ki. Az összefüggésben az x- és y-tengelyen ábrázolt paraméterek felcserélhetőek.

Erre példaként szolgálhatnak a szibilánsok: az /s/ és az /S/ képzéshelyüket tekintve közel állnak egymáshoz, akusztikai szerkezetük viszont jelentősen eltér egymástól: míg az /s/ és a /z/ spektrumára 5 000 Hz körül kezdődő, nagy intenzitású frekvenciasáv jellemző, addig az /S/ és a /Z/ frekvenciaképét 2 000 és 4 000 Hz körüli nagy intenzitású frekvenciakomponensek határozzák meg. Perkell, Boyce, & Stevens (1979) a következő kísérletet írják le: a nyelvperem az /s/ képzésekor a fogmedret érinti, ekkor az akusztikumot 5 000 Hz körül kezdődő, nagy intenzitású frekvenciasáv jellemzi. A nyelvperem lassú hátrafelé mozgásával a jellemző frekvenciasávok először egyenletesen, majd egy ponton hirtelen csökkennek, így az /S/ szükülete, amely az /s/-hez képest csupán néhány milliméterrel hátrébb képződik, jóval alacsonyabb, 2 000 és 4 000 Hz körüli intenzív frekvenciasávot eredményez.¹⁵

A kvantális elméletből nőtt ki a LAFF-elmélet (*Lexical Access from Features*, lexikális hozzáférés jegyek alapján), amely legújabb változatában (Stevens 2002) az észlelés alapjaként fonémák helyett absztrakt szinten bináris megkülönböztető jegyeket feltételez, amelyek egy vagy több akusztikai jellemzőben (*cue*) manifestálódnak (pl. zöngés hangok esetén az alacsony frekvenciatartományban jelenlévő energia). Az akusztikai kulcsok, ill. az általuk reprezentált bináris jegyek felismerése az ugrásszerű változások létrehozta invariancián alapul. Stevens a 80-as években kidolgozta a Chomsky & Halle (1965) által megnevezett, artikulációs alapú bináris megkülönböztető jegyekhez tartozó sablonokat.¹⁶

14 SAMPA-átírás, IPA megfelelőit ld. Vicsi: *A beszéd akusztikai fonetikai leírása* c. fejezetben.

15 Az /s/ magas frekvenciáinak létrejöttét feltehetőleg az okozza, hogy a kiáramló levegő visszaverődik a felső, mások szerint az alsó fogsorról, amelyek így akadályként működnek közre az /s/ képzésében.

16 A bináris jegyeket elsőként bevezető Roman Jakobson részben akusztikai alapú meghatározásokat

A LAFF-elmélet szerint a mentális lexikon a szavakat bináris jegyek sorozataként tárolja, azaz az akusztikai jel auditív percepciója közvetlenül a lexikonhoz kapcsolódik. A LAFF-elmélet (és a kvantális elmélet) tehát az akusztikumban és a percepcióban látja az invarianciát, nem pedig az artikulációban. A megkülönböztető jegyek akusztikai megfelelői az elmélet korábbi változata szerint az ún. sablonok (*template*), a későbbi változat szerint pedig a határjelzők (*landmark*), azaz a szegmensek határán, egyes esetekben magában a szegmensben található jellegzetes invariáns spektrális minták.¹⁷ A határjelzők az akusztikai jel egyes frekvenciatartományaiiban észlelt hirtelen, jelentős változások, mint pl. egy zárfelpattanás. Az akusztikai átmenetből meghatározható, hogy a hang mássalhangzó ([+mássalhangzó]), ezen belül pedig zárhang ([–folyamatos]), amivel egyszersmind számos, pl. a magánhangzókra vagy a szonoránsokra jellemző jegy ([±kerek], [±laterális]) értelmét veszti, azaz a további elemzésből kizárható. Egy határjelző sikeres felismerése után az észlelés során ennek közvetlen környezetében további megkülönböztető jegyre utaló akusztikai kulcsokat keresünk. A bináris jegyekből összeálló mintát folyamatosan összevetjük a mentális lexikonban tárolt mintákkal. Végül az a szó aktiválódik, amelyik legjobban hasonlít az észlelt jegysorozathoz. Fontos megjegyeznünk, hogy a LAFF, hasonlóan a generatív fonológiához, a fonémákat megkülönböztető jegyek kötegének tartja, amelyek az időtengelyen egymás után, illetve részleges vagy teljes átfedésben léphetnek fel.

A LAFF-modell előnye, hogy a megszünteti a fonémaegység és koartikuláció között húzódo ellentmondást, és alkalmas a koartikulációs variabilitás leírására, mivel a megkülönböztető jegyek figyelembe veszik az időbeli változást. Mellette szól továbbá, hogy egyszerű, jól megfogható jegyeken alapszik, nem feltételez nehezen igazolható absztrakt szinteket vagy egységeket, mégis jól magyarázza a percepció megfigyeléseket. Az elmélet további jelentős előnye, hogy empirikus úton jól tesztelhető, egyrészt mert lehetővé teszi a számítógépes modellezést, másrészt mert számos, akusztikailag ellenőrizhető megfigyelésen alapul, és ezáltal jelentősen támogatja a gépi beszédfelismerést is (lásd Böhm: *Nem-nyelvi beszédfeldolgozás* c. fejezet).

Az elmélet hátránya azonban, hogy az eredmények szinte kizárólag mesterségesen (azaz laborban) előhívott észlelésen alapulnak, ezért kérdéses, hogy a spontán beszédre is érvényesek-e. Ellene szól továbbá az az összefüggés, hogy a jól felismerhető akusztikai jellemzőkkel rendelkező hangokat a nyelvek nem feltétlenül részesítik előnyben a kevésbé jól felismerhető hangokkal szemben (az /y/ például CV átmenete alapján könnyen azonosítható, mégsem fordul elő gyakran a világ nyelveiben, mert komplex artikulációs gesztus áll mögötte). Végül, a többi modellhez hasonlóan, a LAFF-modell sem írja le részletesen, milyen lépésekben történik a megkülönböztető jegyek összehasonlítása a mentális lexikonban tárolt egységekkel: mi szolgál az összehasonlítás alapjául? Mekkora súllyal esnek latba az egyes jegyek?

3.3 Közvetlen realista elmélet

A közvetlen realista elmélet (*direct realist theory*) általános percepció elméletbe ágyazódik, és az artikulációs fonológia alapjául szolgáló artikulációs gesztusokra épít. Eszerint a beszéd egymást követő, elvileg különálló (diszkrét) artikulációs gesztusokból, azaz a beszédképző szervek egyes pontjainak összehangolt mozgásából áll, ami a folyamat végére egymást részlegesen fedő gesztusokat eredményez (Fowler 1986, Browman & Goldstein 1992). A közvetlen realista elmélet szerint a percepció invariáns eleme a gesztusok elkülönítésére irányuló képesség.

Az elmélet mögött húzódo általános felfogás lényege, hogy az akusztikai jel önmagában elegendő információt hordoz az általa megjelenített történés felismeréséhez. Az észlelés tehát közvetlenül kapcsolódik a történéshez, a bejövő jel elemzése nem szükséges a feldolgozáshoz — ezzel ebben az elméletben az akusztikai feldolgozás háttérbe szorul az artikulációs folyamatokkal szemben. Ezt olyan kísérletekkel igazolták, amelyekben a kísérleti alanyok fonológiailag megegyező, tehát azonos artikulációs gesztusokból álló hangsorokat nem tudtak

használt.

17 Míg a határjelzők jellegzetesen a hangátmenetekhez kapcsolódnak, pl. egy laterális mássalhangó és egy magánhangzó között a formánsok intenzitásának hirtelen növekedésében és adott esetben frekvenciaváltozásában is manifesztálódnak, a felpattanó orális zárhangok esetében a felpattanás szolgál határjelzőként, amely nem érintkezik közvetlenül a megelőző vagy követő magánhangzó akusztikai leképezésével.

megkülönböztetni egymástól, ha azokat eltérő fonetikai kontextusban hallották (Fowler 1984).

Az elmélet jelentősége egyrészt az általános észlelési modellekbe való beágyazottság, másrészt az, hogy az artikulációt, percepciót és fonológiát összekapcsolja. Ellene szól azonban, hogy az észlelést kiváltó történések ismerete számos esetben nem szükséges a percepcióhoz: így pl. olyan zenét is képesek vagyunk észlelni, amit számunkra ismeretlen hangszerezen játszanak, olyan szagokat is érzékelünk, amelyek forrását nem ismerjük. Ez az elmélet sem ad pontos felvilágosítást arról, az akusztikai jelben milyen módon vannak kódolva a diszkrét artikulációs gesztusok, és hogyan történik ezek felismerése.

3.4 H&H-elmélet

Lindblom *Hyper and Hypo Speech* („túlargikulált és alulargikulált”) elmélete gyakorlati irányultságú: a beszélő és a hallgató között létrejövő, kooperatív kommunikációra épül, amely meghatározza a képzési és értési folyamatokat (Lindblom 1990). A beszédhelyzettől és a kommunikációs partnerek közötti viszonytól függően a beszélő pontosabban vagy lazábban artikulál, ami alapvetően befolyásolja az akusztikumot, éppen ezért az nem is tartalmazhat invariáns jegyeket. A hallgató a percepció folyamatban a jel értelmezése közben folyamatosan felhasználja a rendelkezésére álló egyéb információt is (a beszédhelyzetről, a beszélőről, a világról).

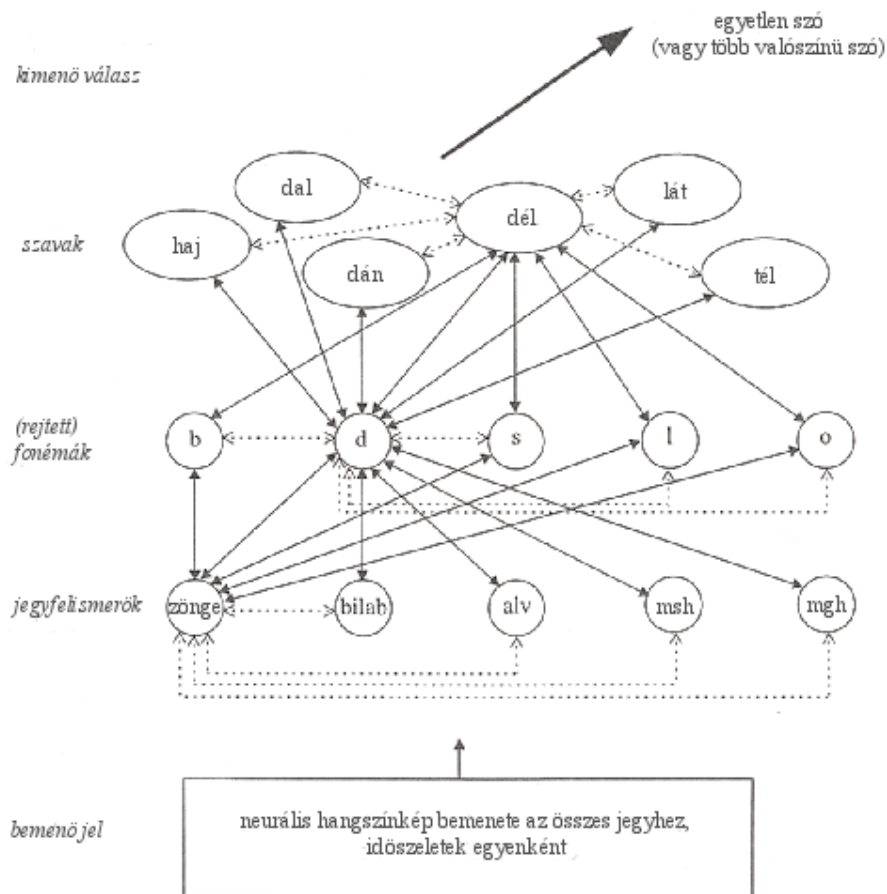
A beszélő a kommunikációs helyzetben kétféle célt követ. Egyrészt érthetővé akarja tenni beszédét a hallgató számára, ezért igyekszik a szegmensek megkülönböztetéséhez elegendő jegyet produkálni (*sufficient contrast*). Amennyiben ez a cél dominál, hiperartikulált beszéddel találkozunk (pl. a beszélő anyanyelvét rosszul ismerő külföldivel szemben, fontos információ közlésekor, vagy ha zajos háttér ellenére kell kommunikálnunk). A beszélő másrészt törekszik a gazdaságos produkcióra, azaz az artikulációs befektetés minimalizálására, ezért az egyes hangokra jellemző **célkonfigurációt** nem mindig éri el – vagy mert az értés szempontjából redundánsnak ítéli, vagy mert a szegmens tartama alatt ez túlságosan nagy befektetést igényelne tőle. (Célkonfiguráció alatt az adott beszédhangra jellemző idealizált artikulációs és akusztikai mintát értjük.) Ezt megfigyelhetjük az angol vagy a német nyelv magánhangzó-rendszerében is, ahol a hosszú magánhangzók egyben fészesekek, a rövidek pedig lazák, azaz az artikulációs szervek utóbbi esetben nem érik el célkonfigurációjukat.

Az akusztikai jel tehát a hiper- és hiperartikuláltság közötti kontinuumon helyezkedik el. A hallgató a jel feldolgozásakor képet alkot arról, hogy a beszélő az adott körülmények között milyen fokú artikuláltságot valósít meg, és ettől függően értelmezi a hallottakat. Az előző három elmélettel szemben a H&H-elmélet figyelembe veszi a beszédhangok szintjénél magasabb feldolgozási szinteket, valamint a hallgató világról tárolt tudását is, viszont Lindblom sem ad választ arra, hogy a leírt jelenségek mögött milyen neuropszichológiai mechanizmusok húzódnak meg.

3.5 Nyommodell

A neurális pályák működését modellezi a nyommodell (*trace model*), amely a pszichológiából és mesterséges intelligencia kutatásból ismert konnekcionista modellek hagyományát követi (McClelland & Elman 1986). A modell alapja egy háromszintű neuronális háló, amely megkülönböztető jegyekből, fonémákból és szavakból épül. A modell kétirányú, azaz lehetővé teszi mind a magasabb egységektől az alacsonyabbak felé történő feldolgozást, mind a fordított irányút.

A 11. ábra egy szótalálási folyamatot ábrázol. Az egyes szinten található egységek össze vannak kötve a saját szintükön és a szomszédos szint(ek)en található minden más egységgel. A modell bemenetét a hangszínekhez hasonló elemzés képezi, amely 5 ms-nyi szeletekre épül. Az alsó szint egységei képesek az akusztikai jegyek kiszűrésére, és ha a bemenettel egyező mintát tárolnak, akkor második lépésként a többi egység is aktiválódik a kódolt mintának megfelelő mértékben. A harmadik szinten a folyamat végén egyetlen szó aktiválódik, és ez szolgál a modell kimeneteként.



11. ábra: A nyommodell szintjei és kapcsolatuk. Példa a zöngésség jegy, a /d/ fonéma és a *dél* szó összeköttetéseire. Az azonos szinten működő inhibitorikus kapcsolatot szaggatott vonal, a szintek közötti excitatorikus kapcsolatot folyamatos vonal jelzi. (Hawkins 1999:272 nyomán.)

Az egységek közötti kapcsolat lehet excitatorikus vagy inhibitorikus: előbbi növeli a másik egység aktivációs szintjét, utóbbi csökkenti. A szintek közötti kapcsolatok excitatorikusak, azaz mindkét irányban segítik az aktiváltság továbbterjedését. Így pl. egy magasabb szintű egység aktiválódása támogatja az alsóbb szinteken lévő releváns egységek aktiválását, és természetesen fordítva is, interaktív hálót alkotva. Egy bizonyos egység aktiválódása egyidejűleg gátolja saját szintje konkurens egységeit.

A nyommodell számos előnyt foglal magában. Egyrészt megoldást kínál a koartikuláció okozta varianciaproblémára, mert az egyes szeletekben kódolt információ elősegíti a későbbi szeletek jobb értelmezését, a koartikuláció itt tehát nem gátló tényezőként jelenik meg. Összhangba hozható a kategorikus észleléssel is (kvantális elmélet), mert lehetővé teszi az egy szinthez tartozó szomszédos elemek gátlását. Megmagyarázza azt a jelenséget is, hogy ha felismertünk egy szót, akkor utólag „belehalljuk” azokat a fonémákat is, amelyek ugyan a szóhoz tartoznak, de a jeltől hiányoztak. Gósy (1989) felhívja rá a figyelmet, hogy ezt a jelenséget már Kempelen Farkas is leírta beszélőgépe kapcsán: amennyiben ismerjük a beszélő (Kempelen esetében a beszélőgép) közlésének tartalmát, úgy azt könnyebben megértjük, és hajlamosak vagyunk azt hinni, hogy azt helyes formában hallottuk (Kempelen 1791).

A modell hiányosságai közé tartozik, hogy nem ad kellő indoklást arra, miért éppen e három felismerési szintet nevezi meg. Nem alkalmas továbbá a mentális lexikon elemei között fennálló szemantikai kapcsolatok modellezésére sem, valamint kizárja azt a lehetőséget, hogy a bemenet nem vezet létező szó aktiválásához (például mert a szó nem a lexikon része, vagy a rendelkezésre álló információ alapján nem tudjuk azonosítani).

3.6 Példányelmélet

Végül említést kell tennünk egy olyan elmületről, amely csak napjainkban kezd ismertté válni, de máris jelentősen befolyásolja az aktuális fonetikai kutatást. Az elmélet alapjául a 80-as években kialakított példány-alapú (angolul *exemplar*) elméletek szolgálnak (Nosofsky 1986), amelyeket Johnson 1997-ben adaptált a nyelvi viselkedésre. Kiindulási pontként olyan kísérleti eredmények szolgálnak, melyek szerint a felismerésben meglepően fontos szerepet játszanak a hívószavakhoz köthető redundáns információk, így pl. a következő kísérletben (Goldinger 1996):

Tíz beszélőtől rögzítettek összesen 300 egyszótagú szót. Először a beszélők hangja közötti hasonlóságot határozták meg egy diszkriminációs tesztben¹⁸. A kísérleti személyek először tanuló fázisban vettek részt: le kellett írniuk a hallott szavakat. A második szakasz kétféle vizsgálatból állt, amelyek a tanuló fázis után öt perccel, egy nappal, és egy héttel következtek. Ebben a kísérletben szavakat hallottak, amelyeknek egy részét már a tanuló szakaszban is hallották, másokat nem. Azt kellett eldönteniük, szerepelt-e a szó az első részben (explicit felismerés). A második feladatban a szavakat fehér zajjal maszkolták, és a kísérleti személyeknek így kellett felismerniük őket (implicit felismerés). Az első kísérlethez hasonlóan a szavak részben ismertek voltak a tanuló szakaszból, és ezt a kísérletet is különböző időpontokban végezték el.

Az explicit feladatban biztosabban felismerték azokat a szavakat, amelyeket hasonló hangú bemozdó ejtésében hallottak a tanuló szakaszban, ez az effektus azonban csak egy napig tartott. Az implicit feladatban a kísérleti személyek még egy hét után is profitáltak a hangok hasonlóságából: a hasonló hangú beszélő ejtette szavakat még ilyen távolságban is könnyebben felismerték az akusztikai elfedés ellenére, mint amelyeket eltérő hangú beszélőtől hallottak.

A kísérlet legfontosabb megállapítása az, hogy a percepció során nem egyszerűen felismerjük a nagyobb nyelvi egységeket, hanem látszólag redundáns információt is tárolunk (mint amilyen a beszélő hangszíne). Ez ellene szól az absztrakt egységekre épülő modelleknek. Nosofsky (1986) szerint a kategóriák ugyanis nem absztrakt elemekből, hanem ezen elemek általunk ismert megvalósulásaiából állnak össze, azaz a példányokból. Egy példány a kategória egy konkrét képviselője, a rá jellemző külső tulajdonságokkal és a hozzá tartozó kategóriák címkéivel (hallott fonéma, beszélő neme, hangszíne stb.). A kategorizálás úgy történik, hogy a hallott elemet összevetjük az eltárolt példányokkal, és a rá leginkább illő kategóriához rendeljük hozzá. A modell működését leíró algoritmus része az euklideszi távolságon nyugvó hasonlóság indexszáma, az ún. figyelmi súly (mennyire hangsúlyos az észlelésben egy bizonyos paraméter), az alapaktíváltság, az aktíváltsági fok, valamint a hozzárendelési tendencia.

A példányelmélet legnagyobb nyitott kérdése, hogy hogyan lehetséges az életünk során észlelt példányok hosszú távú tárolása. Fennáll továbbá a túlkomplikálás veszélye, hiszen a modellbe számos paraméter beépíthető, amelyeket igen nehéz kontrollálni.

4 A beszédészlelés pszicholingvisztikai vetületei

4.1 Lateralizáció

A korábbiakban (1.1.3) láttuk, hogy a hangingerek túlnyomórészt kontralaterálisan továbbítódnak, hasonlóan a többi észlelési folyamathoz. Ennek megfelelően a beszédsegmentumok feldolgozásában főként a bal agyfélteke játszik szerepet, azonban a jobb fül elsőbrendűsége (*right ear advantage*, REA) nem minden hang esetén egyforma, így pl. a zárhangok esetén jelentős, a magánhangzók állandó szakasza esetén csekély. A dichotikus hallást vizsgáló kísérletekben többféle fúziós jelenségről számoltak be: ha pl. az egyik fülbe /b/-re jellemző, azaz emelkedő tranzíciójú szintetizált hangot játszanak be, a másikra /g/-re jellemző, azaz eső tranzícióját, akkor a formánsmenetek összeadódnak, és a kísérleti alanyok /d/-t hallanak, amelynek stagnáló átmenete van (pszichoakusztikai fúzió). Hasonlóan, ha az egyik fül réshang-magánhangzóra emlékeztető kapcsolatot hall, amelyből hiányzik az

¹⁸ A kísérleti személyek feladata az volt, döntsék el, azonos-e két beszélő. A hasonlósági index alapjául a válaszadás reakcióideje szolgált.

egyértelmű formánsátmenet, a másik fül pedig változó alaphangfrekvenciájú hangot, amely megegyezik a hiányzó formánsátmenettel, akkor a kísérleti személy egyértelmű CV-kapcsolatot hall, miközben a változó frekvenciájú hangot is el tudja különíteni (spektrális fúzió). Ugyanez a jelenség megfigyelhető a magasabb feldolgozási egységek szintjén is: a zöngés és labiális megkülönböztető jegyek fúzióját érzük el a /ba/ és /ta/ szótag egyidejű lejátszásával, ekkor a kísérleti személy /pa/ szótagot hall, a /tabi/ és /rabi/ szavak első fonémái pedig /trabi/-vá egyesülnek (fonológiai fúzió) (Pompino-Marschall 2003).

A bal agyfélteke domináns szerepe természetesen jelenti azt, hogy a jobb agyfélteke ne venne részt a beszédpercepcióban. Egyes kísérletek tanúsága szerint a jobb agyfélteke képes zöngéség, ill. aspiráció érzékelésére, valamint szavak és egyszerű mondatok felismerésére. A jobb agyfélteke kikapcsolásával sérül továbbá a prozódiai jegyek felismerése is (Gósy 1989).

4.2 Heteromodális beszédészlelés

A hangingerek feldolgozása összefügg a többi érzékszerv működésével, így a vizuális ingerek feldolgozásával is. Ennek egyrészt a fonéma-graféma feldolgozásban van szerepe, de magát a beszédészlelést is befolyásolja. Egyrészt fontos szerepet játszik a szájról olvasás, főként zajos környezetben, vagy ha egy általunk kevésbé ismert idegen nyelvet kell megértenünk.

Gósy kimutatta, hogy az értelmetlen hangsorok felismerése 10%-kal javítható, ha az auditív információ kiegészül a hangsornak megfelelő ajakmozgás látványával (Gósy 1989). Ha viszont a hangsorokat más hangok artikulációjának megfelelő vizuális inger kíséri, a felismerés 20%-kal romlik. Ugyanez az értelmes nyelvi egységek azonosítására fokozott mértékben érvényes.

A beszédészlelést azonban nemcsak támogathatja, hanem meg is zavarhatja a vizuális információ, amint McGurk és MacDonald híres kísérletéből kiderül. Ha egy kísérleti személy fülhallgatón a /ba/ szótagot hallja, ezzel egyidejűleg pedig képernyőn egy artikuláló személyt lát, aki a /ga/ szótagot ejti ki, az észlelt szótag /da/ lesz. Ez a McGurk-effektus,¹⁹ amelyet számos nyelvre igazoltak, és amely meglehetősen robusztusnak mutatkozik: ha az auditív inger 180 ms késéssel éri a fület a vizuális ingerhez képest, még mindig megtörténik az egybeolvadás. Egyes vizsgálatok szerint a jelenség auditív alapú, azaz a /ba/ szótag váltja ki, ezért nem hasonlítható a fúziókhöz (Roberts & Summerfield 1981).

4.3 Gyermek perzepciójának fejlődése

A hallásérzékenység jelentős mértékben változik az egyedfejlődés során. Selnes & Whitaker megállapította, hogy a hallás öt éves kortól egyre jobb, az érzékenység 12–13 éves korban a legnagyobb. 14 éves korra kialakul a felnőttkorra jellemző hallásérzékenység. A fejlődés a beszéd szempontjából releváns, 100 Hz és 4 000 Hz közötti tartományban a legjelentősebb (Selnes & Whitaker 1976).

Gósy (1989) kísérletben igazolta, hogy a szófelismerés hatékonysága erősen függ az életkortól. Kísérletében a gyermekek mindennapi életében előforduló szavakat rögzített és fedett el fehér zajjal. Míg az iskola előtti korosztály (3 és 5 évesek) e szavakat legfeljebb felerészben ismerték fel helyesen, az iskolába járó 7 évesek már 63%-ban, a 40 év körüli felnőttek pedig 84%-ban (a 10, 14 és 20 éves korosztályok teljesítménye ezen értékek között volt, növekvő tendenciával). A kétszótagú szavak felismerése minden korcsoportban jobb, mint az egyszótagúaké, ami a nagyobb akusztikus információadagnak köszönhető (Büki, Egyed & Pléh 1984). A vizsgált mondatok felismerése (amelyek egy három éves gyermeket esetenként szókincsük, jelentésük és nyelvtani szerkezetük alapján elfedés nélkül is nehéz feladat elé állítottak) növekvő életkorral szintén egyre javult. Ugrásszerű változást a 3. és 5., valamint a 14. és 20. életév között lehetett megfigyelni.

4.4 Zárszó

A fentiekből láthattuk, hogy a beszédpercepció kutatások egyelőre nem tudtak kielégítően válaszolni arra a kérdésre, hogyan absztrahálja és kapcsolja az agy a beérkező beszédjelet magasabb nyelvi egységekké, sőt abban sincs egyetértés, hogy bemenetként az akusztikai jel szolgál-e, vagy az artikulációs gesztus (mint a motoros és a direkt realista elméletben), és hogy vajon a beszédfeldolgozás kimenete a fonéma, a megkülönböztető jegy vagy a szó-e.

¹⁹ Az interneten számos példát lehet találni a „McGurk effect” keresőszó beadásával.

Az elméletek sokféleségének hátterében az a probléma áll, hogy a percepcióban nem léteznek egy az egyhez típusú kapcsolatok, sőt, a több a többhez mintázat a jellemző. Ezért ha egy kísérlet eredményei alátámasztják valamelyik modell helyességét, az nem jelenti azt, hogy az a modell – és csak az a modell – térképezte fel helyesen a beszédfeldolgozás folyamatát, hiszen az eredmények egyrészt más modellekben is értelmezhetőek lennének, másrészt lehet, hogy a neuropszichológiai valóság egy szeletét képezik csak – főként ha emlékezetünkbe idézzük, hogy az ismertett kísérletek túlnyomórészt mesterségesen előállított vagy manipulált hangmintákra épültek. Ezért, bár a fentiekben a pszicholingvisztikai modellek percepció oldalának csupán két szintjét, a fonetikai és fonológiai szintet kíséreltük meg összekötni, még ez sem járhat sikerrel a hangingerek feldolgozásának neurofiziológiai alapjainak jobb ismerete nélkül.

Köszönetnyilvánítás

Köszönöm mindazok segítségét, akik hozzájárultak e fejezet elkészüléséhez: Vicsi Klára, Uwe Reichel, Böhm Tamás, Hartmut Pfitzinger és Markó Alexandra. Az írás ideje alatt az Alexander von Humboldt Alapítvány támogatását élveztem.

Irodalom

- Bolla, K. (1995). Magyar fonetikai atlasz: a szegmentális hangszerkezet elemei. Budapest: Nemzeti Tankönyvkiadó.
- Browman, C. P., & Goldstein, L. M. (1992). Articulatory phonology: an overview. *Phonetica*, 49, 155–180.
- Büki, B., Egyed, A., & Pléh, Cs. (szerk.). (1984). *Nyelvi képességek – fogalomkincs – megértés*. Budapest: Tankönyvkiadó.
- Chistovitch, L. A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America*, 77, 789–805.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Fastl, H. & Zwicker, E. (2006). *Psychoacoustics: facts and models*, 3. kiadás. Berlin et al.: Springer.
- Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, 12, 47–65.
- Fónagy, I. (1958). A hangsúlyról. *Nyelvtudományi Értekezések*, 18.
- Fowler, C. (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics*, 36, 359–368.
- Fowler, C. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3–28.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 22, 1166–1183.
- Gósy, M. (1989). *Beszédészlelés*. Budapest: Akadémiai Kiadó.
- Gósy, M. (2003). *Fonetika: a beszéd tudománya*. Budapest: Osiris.
- Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics*, 31, 465–485.
- Hawkins, S. (1999). Reevaluating assumptions about speech perception: interactive and integrative theories. In Pickett 1999, pp. 232–288.
- ISO Szabvány R 226-1961, Normal Equal-Loudness Contour for Pure Tones and Threshold of Hearing under Free Field Listening Condition.
- Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In: K. Johnson, & J. W. Mullennix (szerk.), *Talker variability in speech perception* (pp. 145–166). San Diego: Academic Press.
- Kempelen, F. (1791). *Az emberi beszéd mechanizmusa, valamint a szerző beszélőgépezet leírása*. Bécs: J. V. Degen.
- Kent, R. D. (1996). *The speech sciences*. San Diego & London: Singular Publishing Group.
- Ladd, R. D. (1996). *Intonational phonology*. Cambridge Studies in Linguistics 79. Cambridge: Cambridge University Press.
- Lehiste, I., & Peterson, G. E. (1961). Transitions, glides and diphthongs. *Journal of the Acoustical Society of America*, 33, 268–277.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W. Hardcastle, & A. Marchal (szerk.), *Speech production and speech modelling* (pp. 403–439). Dordrecht: Kluwer.
- Lisker, L., & Abramson, A. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, 10, 1–28.
- Mády, K., & Reichel, U. D. (2007). Quantity distinction in the Hungarian vowel system—just theory or also reality? *Proc. 16. ICPhS* (pp. 1053–1056). Saarbrücken, Germany.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- Miller, R. L. (1953). Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America*, 25, 114–121.

- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 42A, 39–57.
- Perkell, J., Boyce, S. E., & Stevens, K. N. (1979). Articulatory and acoustic correlates of the [s-sh] distinction. *Speech Communication Papers, 97th Meeting of the Acoustical Society of America* (pp. 109–113). Cambridge, Massachusetts.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24, 175–184.
- Pétursson, M., & Neppert, J. (1991). *Elementarbuch der Phonetik*. Hamburg: Buske.
- Pfützinger, H. R. (2006). *Einführung in die Perzeptive Phonetik*. Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilians-Universität München.
- Pickett, J. M. (szerk.). (1999). *The acoustics of speech communication*. Boston et al.: Allyn and Bacon.
- Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America*, 61, 1352–1361.
- Pompino-Marschall, B. (2003). *Einführung in die Phonetik, 2. kiadás*. Berlin & New York: de Gruyter.
- Reichel, U. D. (2007). *Perzeption für Fortgeschrittene*. Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilians-Universität München.
- Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception and Psychophysics*, 30, 309–314.
- Selnes, O., & Whitaker, H. (1976). Morphological development of the auditory system. In R. Rieber (Ed.), *The neuropsychology of language* (pp. 125–156). New York & London: Plenum.
- Stevens, K. N. (1989). On the quantal theory of speech. *Journal of Phonetics*, 17, 3–45.
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111, 1872–1891.
- Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America*, 85, 2207–2217.
- Strange, W. (1999a). Perception of consonants: from variance to invariance. In Pickett 1999, pp. 166–182.
- Strange, W. (1999b). Perception of vowels: dynamic consistency. In Pickett 1999, pp. 153–165.
- Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, 79, 1086–1100.
- Szentágothai, J. (1971). *Functional anatomy I-III*. Budapest: Medicina.
- Tarnóczy, T. (1982). *Zenei akusztika*. Budapest: Zeneműkiadó.
- Tarnóczy, T. (1984). *Hangnyomás, hangosság, zajosság*. Budapest: Akadémiai Kiadó.
- Trautmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America*, 69, 1465–1475.
- Vicsi, K. (1981a). Az időtartam szerepe néhány mássalhangzótípus hallás alapján történő megkülönböztetésében. *Magyar Fonetikai Füzetek*, 7, 59–66.
- Vicsi, K. (1981b). The most relevant acoustical micro-segment and its duration necessary for the recognition of unvoiced stops. *Acustica*, 48, 53–58.