A beszédpercepció helye a teljes megértési folyamatban

Mády Katalin

MTA Nyelvtudományi Intézet, Budapest

mady.katalin@nytud.mta.hu

Rövidített cím: Beszédpercepció

Absztrakt

A hangzó nyelvi feldolgozás alulról felfelé építkező modelljeinek bemenete az akusztikai jel, amely absztrakt fonológiai leképezés után kerül be a magasabb kognitív folyamatokba. Nem világos azonban, hogy hogyan lesz a fizikailag mérhető, folytonos, és erőteljes variabilitásnak kitett akusztikai jelből absztrakt, diszkrét, kis számú egységből álló fonémasorozat. A cikkben először a szegmentális és szupraszegmentális szint észlelését jellemezzük pszichoakusztikai szempontból. Ezután bemutatunk hat, napjainkban is meghatározó, egymásnak részben látszólag ellentmondó beszédpercepciós modellt, amelyek az akusztikai jelből vagy az artikulációs folyamatokból kiindulva jutnak el a megkülönböztető jegyeket hordozó fonémák azonosításához. Ilyenek a motoros elmélet, a kvantális elmélet és a LAFF, a közvetlen realista elmélet, a H&H elmélet, a példányelmélet és a nyommodell. Végül a modelleket összegezve tárgyaljuk a beszédpercepció és a magasabb szintű nyelvi folyamatok összefüggéseit.

Kulcsszavak: beszédpercepció, hallás, pszichoakusztika, percepciós modellek, kogníció.

Abstract

The input to models of spoken language processing is the acoustic signal that is projected on abstract phonological units which participate in higher cognitive processes. However, it is not clear in what way the physically measurable, continuous, and highly variable acoustic signal is transformed into a sequence of abstract and discrete units forming a closed set. The paper starts with a description of the perception of segmental and suprasegmental units based on psychoacoustics. In the next part six relevant models of speech perception are presented that seem to be partly contradictory at the first sight. They take either the acoustic signal or articulatory processes as

their input and result in the identification of phonemes with distinctive features as their output. The theories are: motor theory, quantal theory and LAFF, direct realist theory, H&H theory, exemplar theory and the trace model. Finally, the models are summarised and the relationship between speech perception and higher linguistic processes is discussed.

Key words: speech perception, hearing, psychoacoustics, perceptual models, cognition.

1. Bevezetés

A pszicholingvisztikai modellek többsége a beszédhangok szintjét tekinti a nyelvi feldolgozás legkisebb egységének. A modellek – a strukturalista nyelvészet szemléletének 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 percepciós láncban a bemenet utáni első, a produkciós 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, szemantika, 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 ítél egy anyanyelvi beszélő két hangot azonos osztályhoz tartozónak. Például az alveoláris, azaz a felső fogmeder érintésével képzett /n/[1] hang a környezet hatására jelentős változásokon megy keresztül: a *len* szóban alveoláris, a *kengyel* szóban palatális, a *ménkű* 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 lennénk 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. A beszédpercepció tehát a pszicholingvisztikai

1

A fonetikai leírásokban nem egységes a szögletes és ferde zárójel használata: egyes rendszerekben az /n/ jelölés a fonémára, az [n] pedig a kiejtett beszédhangra vonatkozik, mások viszont a ferde zárójelet használják a durvább, a szögleteset pedig a finomabb fonetikai átírásra. E dolgozat az utóbbi hagyományt követi. A fonetikai jeleket SAMPA-átírás szerint adjuk meg, ld. http://www.phon.ucl.ac.uk/home/sampa/hungaria.htm. (2013. május 10.) A magyar hangokra esetenként dőlten szedett betűjegyükkel utalunk.

modellek fonológiai és fonetikai szintje között található részfolyamatokat igyekszik leírni és modellezni.

A beszédészlelés a kommunikációs lánc harmadik nagy egységét képzi a beszédprodukció és a nyelvi jel után, következésképpen függvénye mind a beszélőszervekkel előállítható akusztikumnak, mind magának az akusztikai jelnek. Más szóval: a percepció arra a frekvenciasávra és azokra az időbeli felbontásokra van utalva, amit egy emberi beszélő képes produkálni, és amit a közvetítő közeg – többnyire a levegő – továbbítani tud. Mint látni fogjuk, az emberi fül különösen 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 működésben levő folyamatokról van szó. A beszédpercepció leírásakor tehát szem előtt kell tartanunk a beszédprodukció folyamatát is.

A következő fejezetben a fül anatómiájáról és a hallás fiziológiai alapjairól lesz szó. A leírás a percepciós folyamatok megértéséhez szükséges alapvető ismeretekre szorítkozik. A 3. részben az alapvető szegmentális és szupraszegmentális egységek felismerésére térünk ki, amelyeket a 4. részben különféle percepciós modellek fényében vizsgálunk meg újra. Végül az 5. részben a beszédpercepció és a magasabb kognitív műveletek együttműködési mechanizmusaira és neuroanatómiai beágyazására térünk ki.

2 Hallás és pszichoakusztika

Az emberi észleléssel általános értelemben a pszichofizika foglalkozik. E tudományág azt vizsgálja, hogy konkrét fizikai nagyságok és az ember által észlelt mértékük között milyen összefüggések vannak: például súly és szubjektív nehézségérzet, hangerő és hangosság. Ezen belül a pszichoakusztika a hanghullámok észlelésének részterülete. A percepciós 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 percepciós 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.

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

Mivel az észlelt jelhez közvetlenül nem férünk hozzá, a pszichoakusztika felismerései kísérletek eredményeire épülnek. A tudományág fő célkitűzése, hogy felfedje az akusztikai jel fizikai mérőszámai és az észlelés hozzájuk rendelt pszichikai mérőszámai közötti matematikai összefüggéseket.

2.1 Intenzitás, hangerő és hangosság

Az emberi hangészlelés alsó határa 20 Hz, felső határa 20000 Hz, azaz 20 kHz. Ez az adat fiatal felnőttek hallására vonatkozik. Idősebb korban a magas frekvenciák észlelése fokozatosan romlik, ezért előfordul, 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 logaritmikus egységben, decibelben (dB) adjuk meg. 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 3400 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 3400 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 személyek 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.

1. ábra kb. ide

1. 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 hangosságarányérzet 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 hangosságarány-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).

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 frekvenciaaránynak 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).

A belsőfül anatómiai sajátságaiból adódóan (ld. Szentágothai 1971, Bolla 1995, Pompino-Marschall 2003) az emberi hallószerv a hangmagasságot különböző frekvenciákon eltérő pontossággal észleli. Míg 20 Hz és 500 Hz között az észlelés nagyjából lineáris, efölött 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 észlelünk 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 1000 mel-ben állapították meg.[2] Míg az érzékelés 500 Hz-ig nagyjából lineáris, efölött a Hertz- és mel-értékek aránya logaritmikus. Így például 100 és 200 Hz távolságának 150 és 283 mel magasságú észlelet felel meg, az 1000 és 2000 Hz-es hangok között viszont már kisebb a távolság: 1 000 és 1 521 mel, a 10000 és 20000 Hz-nek megfelelő értékek aránya pedig még kisebb:

2 A mel-t egyes források a fentiektől eltérően a Bark-skála alapján definiálják: 1 Bark = 100 mel (ld. Pompino-Marschall 2003).

3079 és 3817 mel[3] (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 széles sávú 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 nagyobbnak, ill. a szinuszos hang intenzitásának kisebbnek kell lennie a maszkolási jelenség kiváltására.

A kritikus sávok módosító hatását szemlélteti a 2. ábra. A csendes környezetre vonatkozó hallásküszöb (szaggatott görbe) fölfelé módosul 60 dB-es fehér zaj bejátszásakor. Ha a zaj középső frekvenciája 250, 1000, ill. 4000 Hz, amint ezt az ábrán láthatjuk, akkor egy 50 dB-es hangerejű szinuszus 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).

2. ábra kb. ide

2. ábra: Folyamatos vonalak: 250 Hz, 1 kHz és 4 kHz körüli, 1 kritikus sáv szélességű, 60 dB intenzitású széles sávú 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 belsőfülben található csiga a magasabb frekvenciákra a bemenethez közelebb reagál, tehát ezek a frekvenciák nem, vagy csupán csekély mértékben befolyásolják a csiga bemenetétől távolabbra található alacsonyabb frekvenciák érzékelését. Fordított esetben viszont a hanghullám keresztülhalad a membrán magasabb frekvenciák érzékeléséért 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

3 Kerekített értékek.

fülünket egyszerre több hang éri, és ezek egy kritikus sávon belül vannak, akkor intenzitásuk összegződik, de nem észleljük őket különálló hangokként. A kritikus sávok 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 15500 Hz között (Tarnóczy 1982, Fastl–Zwicker 2006).

A kritikus sávoknak, valamint a csiga tonotópiás felépítését tükröző hangmagasság-skáláknak igen fontos szerepe van az emberi beszédpercepciós folyamatok leírásában, hiszen – ellentétben a frekvencia- és félhangalapú mértékegységekkel – kifejezésre juttatják azt a tényt, hogy az emberi hallás érzékenysége frekvenciafüggő. Éppen ezért a kritikus sávok 500 Hz alatt nagyjából 100 Hz szélesek, efölött egyre nagyobb értéket vesznek fel. Ezt szemlélteti a 3. ábra: az *x* tengelyen jelölt értékek között egyre nagyobb a távolság, míg a nekik megfelelő Bark-értékek között nagyjából azonos a különbség: 500, 1000, 2000, 4000 és 10000 Hz hangmagasság-észlelete 4,9; 8,5; 13,0; 17,5 és 21,9 Bark.

3. ábra kb. ide

3. ábra: A frekvenviaérték (kHz) és a fül által észlelt hangmagasság (Bark) függvénye.

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, más frekvenciákon az észlelési küszöb akár 22 ms-t is elérhet.

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 esetén. Ez az ún. utóelfedés. A jelenség ellenkező irányban is megfigyelhető: ha egy gyengébb ingerre erősebb inger következik, bizonyos körülmények között csak a második ingert észleljük. Az előelfedés csak lényegesen kisebb időbeli eltérés esetén lép fel, és azzal magyarázható, hogy a nagyobb intenzitású második inger felismerési ideje rövidebb, ezért elnyomja a lassabban észlelt gyengébb ingert.[4] Ha nagyobb a hangingerek intenzitása, az elfedés nagyobb időbeli eltérés mellett is

4 Az utóelfedés angol megfelelője forward masking, az előelfedésé backward masking, az elnevezések tehát a magyarral ellentétes irányúak.

kimutatható. Az előelfedés esetén szerepet játszik az is, hogy a hangingert azonos vagy különböző fülön keresztül halljuk-e: az azonos fülön történő bejátszás esetében nagyobb időkülönbség is kiváltja az elfedést (ld. 4. ábra).

4. ábra kb. ide

4. ábra: Időbeli elfedés 40, 60 és 80 dB-es fehér zaj esetén. A: utóelfedés hatása, B_1: különböző fület érő hangok, B_2: azonos fület érő hangok (ld. bővebben Tarnóczy 1984.)

3 Szegmentális és szupraszegmentális elemek percepciója

A percepciós kísérletek alapjául természetes vagy szintetizált beszédet használunk 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 például idő-, frekvenciaszerkezetüket vagy intenzitásukat tekintve, így vonva részleges ellenőrzés alá akusztikai sajátosságaikat. Ez a módszer mindkét típus előnyeit és hátrányait egyesíti. A percepciós kísérletek örök dilemmája éppen ezért az, hogy a kísérletek eredményei mennyire érvényesek a természetes körülmények között zajló beszédérzékelésre. 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 kísérletek eredményeinek értelmezésekor ezt mindig szem előtt kell tartanunk.

3.1 Magánhangzók

3.1.1 A magánhangzók percepciója

A hangokat fonológiai szinten a generatív fonológiához kapcsolódva gyakran bináris szempontok alapján szokás jellemezni: hosszúság, ajakkerekítés, feszesség, zöngésség, zörejesség stb. Ezeket a megkülönböztető jegyeket általában nem lehet egyetlen artikulációs vagy akusztikai mérőszámnak megfeleltetni, és ez különösen igaz a magánhangzókra. A hosszú magánhangzók akusztikailag mérhető tartama például nem feltétlenül hosszabb, mint a rövid magánhangzóké, azonosításuk ettől függetlenül – más akusztikai 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 voltuk. A hosszú és rövid magánhangzók közötti eltérés mibenlétét tehát csak nagyobb összefüggések vizsgálatával deríthetjük ki.

A magánhangzók jellemzésére leggyakrabban tartamukat (azaz mérhető hosszukat), az alapfrekvenciát (f0) és az első két formánst (F1 és F2) szokás használni, továbbá a felsőbb formánsokat (F3–F5), az intenzitást és egyéb akusztikai mérőszámokat. A mérések céljától függően ezek az értékek mérhetőek adott pontokon (például a magánhangzó középpontjában), vagy a hang teljes tartama alatt, ha a dinamikus tulajdonságokra vagyunk kíváncsiak. Egyes jellemzők, mint a tartam, alapfrekvencia, F1 és F2 könnyen értelmezhető fonetikai jegyeknek felelnek meg, legalábbis a beszédre jellemző frekvenciatartományban: a tartam a hosszúságnak, az f0 a hangmagasságnak, az F1 a magánhangzó függőleges képzéshelyének, az F2 a vízszintes képzéshelynek. A magánhangzók felismerését vizsgáló kísérletek egy része azt példázza, hogy a fenti jellemzők, vagy akár csak egy részük is, elegendőek 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 gyermekek) kétszeri ejtésében: *heed*, *hid*, *head*, *had*, *hod*, *hawed*, *hood*, *who'd*, *hud*, *heard*[5]. A hívószavakat egy kísérletben összesen 70 kísérleti személy hallgatta meg, akik részben azonosak voltak a beszélőkkel (Peterson–Barney 1952).

5. ábra kb. ide

5. ábra: amerikai angol magánhangzók formánseloszlása 76 beszélő ejtésében
(*x* tengely: F2, *y* tengely: F1). (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. 5. ábra). A férfiak, nők és gyermekek toldalékcsöve eltérő méretű, ezért az általuk ejtett magánhangzók abszolút formánsértékei is különböznek (arányukban viszont megegyeznek). A kísérlet második részében csak felnőtt férfiak által ejtett magánhangzókat elemeztek, és kizárták azokat, amelyeket a kísérleti személyek nem tudtak egyöntetűen azonosítani. Még ekkor is voltak átfedések a könnyen azonosítható

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

magánhangzók frekvenciatartományai között. A szerzők ezért feltételezik, hogy az abszolút formánsértékek nem elegendőek a magánhangzók 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 percepciós feladatban nem tudták egymástól megkülönböztetni az ezen fonémákhoz tartozó realizációkat.

A kísérlet másfelől rámutat arra, hogy a magánhangzók azonosításának sikere függ az adott nyelv magánhangzórendszerének szerkezetétől. Peterson és 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 például az /i/-nek, ezáltal nő a tévesztés veszélye is.

Későbbi kísérletek bizonyították, hogy a magánhangzó-felismerés nem abszolút, hanem relatív frekvenciaértékeken alapszik. Miller (1953) szintetizált hangokkal kimutatta, hogy az alapfrekvencia döntően befolyásolja a magánhangzó nyíltsági fokának észlelését. Ha ugyanis az alapfrekvencia egy oktávval magasabb volt, a kísérleti személyek a magánhangzót – azonos formánsértékek mellett – zártabb hangzóként észlelték. További fontos felismerések adódtak Bark-alapú számításokból. Syrdal és Gopal Peterson és Barney kísérleti anyagát újraelemezve megállapította, hogy ha a magánhangzók feltérképezésekor Bark-értékeket vesznek alapul, valamint figyelembe veszik az f0 és F1, valamint az F1 és F2 távolságát, akkor a hangok szórása nagyban csökkenthető, és a kategorizálás az alapfrekvencia és a két alsó formáns alapján elvégezhető (Syrdal–Gopal 1986). Traunmüller (1981) rámutatott, hogy a nyíltsági fok észlelését döntően az f0 és az F1 Bark-ban mért távolsága befolyásolja. Chistovitch (1985) szerint azokat a formánsokat, amelyeknek távolsága 3–3,5 Bark-nál kisebb, egyetlen formánsként érzékeljük, ennek értéke a két abszolút formánsérték spektrális súlypontjába esik.

Lehiste és Peterson (1961), valamint Strange (1989) a magánhangzó-azonosításban kiemeli a dinamikus jellemzők szerepét, különös tekintettel a rövid–hosszú, ill. feszes–laza oppozícióra. 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. A hosszú magánhangzók állandó szakasza ugyanis relatíve hosszabb, mint a rövid magánhangzóké. Strange kísérleti eredmények 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 f0 (hiszen az alapfrekvenciából következtethetünk a beszélő nemére és korára), az egyéni sajátságokat hordozó F3, a kísérletben vizsgált magánhangzót körülvevő egyéb nyelvi információ (például a mondat, amelybe a hívószót beágyazták). A külső jegyek szerepét bizonyítja, hogy a célszó magánhangzójának kategorizálása változik, ha a hordozó mondat magánhangzóinak formánsértékét manipuláljuk, valamint hogy a magánhangzó besorolása befolyásolható az egyidejűleg vizuálisan prezentált „beszélő" nemén keresztül. (Az intrinzikus és extrinzikus modelleket átfogóan ismerteti Strange 1999b.)

3.1.2 Vizsgálatok a magyar magánhangzók percepciójáról

A beszédpercepciós 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 formánsa 600 Hz volt, a hangot a kísérleti alanyok akkor is 70%-os biztonsággal felismerték, ha a lejátszott hang kizárólag 390 Hz alatti frekvenciatartományból állt. Más magánhangzók, mint például az *o* hang hasonlóan jó felismeréséhez viszont a szűrt hangnak magasabb frekvenciákat is kellett tartalmaznia, jóval az *o* első formánsa feletti tartományból. Gósy szerint a magánhangzók viszonylag megbízható azonosításához szükséges frekvenciatartomány felső határa és a magánhangzó F2 értéke közötti távolság specifikus, tehát magánhangzónként eltérő: az *á* esetében 100 Hz, az *í* esetében viszont 1200 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 formánsoknál alacsonyabb intenzitású – frekvenciasávok is szerepet játszanak a magánhangzó felismerésében.

Hasonló eredmények születtek a kísérlet azon részéből, amelyben a sáváteresztő szűrő alsó határát 270 Hz és 2200 Hz között mozgatták, a felső határ 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 sajátságai, 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

F1 és az F2 megléte mellett az F3 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 1000 Hz és 3300 Hz között mozgott. A legjobb az *i* és *ü* hangok felismerése volt, hiszen ezen magánhangzók második formánsa 2000 Hz felett található, ezért a felsőbb formánsok az áteresztett tartományon belül voltak. 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 *ü* 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 formánsa egy skála mentén változik (az alapfrekvencia és a felsőbb formánsok értéke változatlan). Eredményei szerint egy adott F1 érték rövidebb magánhangzó esetén zártabb, hosszabb magánhangzó esetén nyíltabb kategóriába való sorolást eredményez, így például 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 viszonylag magas F1-értéket produkáljon. Ez a felismerés Liberman motoros elméletét támasztja alá (ld. **4.1**).

A magánhangzók hosszúsága és minősége közötti összefüggés fonológiai szempontból is érdekes. Úgy tűnik ugyanis, hogy a magyarban a fonológiailag hosszú és rövid magánhangzópárok több szempontból egy kontinuum mentén helyezkednek el, ahol a középső nyelvállású *o–ó* és *ö–ő* pár átmenetet képez: (1) a felső nyelvállású magánhangzók, mint *i–í*, *u–ú*, *ü–ű* funkcionális terheltsége alacsony, ritkán van jelentésmegkülönböztető szerepük, míg az alsó nyelvállású magánhangzók számos hosszúságalapú minimális párban vesznek részt (a *zug – zúg* típusú párból nagyon kevés van, a *hal – hál*, *fel – fél* típusúból sok). (2) A felső nyelvállású magánhangzók formánsértékei alig vagy nem térnek el egymástól, az alsó nyelvállású párok viszont fonetikai értelemben eltérő osztályokhoz sorolhatóak: az elölképzett *e–é* pár az állkapocs nyíltsági fokában, az *a–á* pár pedig mind a vízszintes képzéshelyben, mint az ajakműködésben eltér (az *a* hátsó képzésű és kerekített, az *á* középső és réses). A Gósy (2004, 119–120) alatt közölt formánstérképeken is látszik, hogy a rövid és hosszú felső nyelvállású magánhangzók formánsértékei jelentős átfedést mutatnak, a fonológiai értelemben alsóként kategorizált *e–é* és *a–á* formánsértékei viszont nem.

Mády és Reichel (2007) azt vizsgálta, hogy mennyire megbízható a magánhangzók azonosítása, ha a tartam nem nyújt támpontot a felismeréshez. A kísérletben 40 ms-ra rövidített magánhangzókat használtak, amelyek a magánhangzó középső, viszonylag statikus részét tartalmazták. Azt találták, hogy az alsó nyelvállású magánhangzók azonosítása ilyen rövid tartam mellett sem jelentett nehézséget, különösen az *a* és *á* esetében, szemben a felső nyelvállású magánhangzókkal, amelyeket a kísérleti személyek egyöntetűen rövidnek hallottak. Ez nem is meglepő, hiszen az alsó nyelvállású magánhangzók spektrális jellemzői eltérnek egymástól, a felsőké viszont nem. Érdekes azonban, hogy mind Gósy (2004), mind Mády és Reichel (2007) azt találják, hogy a rövid és hosszú középső nyelvállású magánhangzók formánsértékei sem fedik egymást – ennek ellenére a kísérleti személyek a 40 ms tartamú, eredetileg hosszú magánhangzókat szintén rövidnek hallják. Itt tehát a hosszúság mint fonológiai jegy gátolja az akusztikailag egyértelműen elkülöníthető spektrális jegyek észlelését.

3.2 Mássalhangzók

3.2.1 A mássalhangók percepciója

A magánhangzók artikulációs jellemzői nagyjából folytonosan változnak egy-egy faktor mentén: az /o/ nyíltabb, mint az /u/, de kevésbé nyílt, mint az /a/. Ugyanez nem érvényes a mássalhangzókra. Bár a képzési helyek elvileg szintén kontinuumot képeznek, legalábbis a fogmedertől a lágy szájpadlásig, a gyakorlatban a legtöbb nyelvben jól elkülöníthető képzési helyekkel találkozunk, a magyar zárhangok esetében például labiális (*p*, *b*), alveoláris (*t*, *d*), palatális (*ty*, *gy*) és veláris (*k*, *g*) hangzókkal. E képzési helyeken felül ugyan elvileg elképzelhetőek köztes zárhangok is, ám ezekkel nem, vagy igen ritkán találkozunk.

A mássalhangzók esetében felvetődik két további probléma is: egyrészt a mássalhangzók sok esetben nem vizsgálhatóak közvetlenül, csupán a szomszédos magánhangzókon 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 például /di/ hangkapcsolatban a /d/ felpattanását követően a rá jellemző F2-tranzíció[6] 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/ F2-jéig esik. Az F2-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-

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

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.[7] 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 észleli, mint a változatlanságot. Az /E/ sok nyelvben gyakori magánhangzó, az alveoláris mássalhangzók előfordulása általában szintén magas, így a kapcsolódásuk is átlagon felül fordul elő. Greenberg ezzel magyarázza azt a tényt, 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 (ld. **4**) kialakításához járultak hozzá. További kísérletek leírását ld. Strange (1999a).

Az Egyesült Államok-beli Haskins laboratóriumban az 1950-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 szintetikusan, úgy, hogy az F2-átmenetet a /b/-re jellemző értéktől a /d/-n keresztül a /g/ értékéig fokozatosan növelték 120 Hz-es lépésekben, így összesen 14 F2-átmenetet kaptak. Az F1 menetét nem változtatták. A kísérleti személyek a várakozásnak megfelelően az alacsonyabb értékről kezdődő átmenetek kezdőhangjaként /b/-t, majd /d/-t, a felső értékek esetén pedig /g/-t adtak 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 és Abramson (1967) a zárhangok zöngésségét vizsgálva. CV-kapcsolatokban az alsó három formáns egyidejű megjelenését 0 ms-nak véve a zöngekezdési 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

7 Többen kétségbe vonják, hogy a mássalhangzók képzési helye egyértelmű lokuszdefiníciókat tesz lehetővé, ld. pl. Lehiste és Peterson (1961), más kísérletekben viszont igazolták a lokuszelmélet helyességét. A kérdésben nincs egyértelmű konszenzus, bár a lokuszértékeket közelítő jelleggel mindenki elfogadja.

volt: a hangkapcsolatok mássalhangzóját –150 és +20 ms zöngekezdési 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 összehasonlítandó 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 (ld. **4.1**). A heterogén kísérleti eredmények fényében nem egyszerű sem a bizonyítás, sem a cáfolat. Egyrészt több kísérletben is igazolták, hogy a /b/–/d/–/g/ elkülönítéséül szolgáló F2-értékek nem-nyelvi jelek esetében nem váltanak ki 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.

3.2.2 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 (6. ábra). Anyagában a /b/, /d/, /g/ mássalhangzók, valamint különféle magánhangzók szerepeltek CV-kapcsolatokként. A felpattanás pillanatában és a CV-átmenetben az F2 értékefokozatosan változott. Gósy eredményei szerint a /b/ felismerésének elsődleges kulcsa a második formáns frekvenciájában rejlik, amely azonban egy viszonylag nagy, 1000 Hz-t átölelő tartományba eshet (800–1800 Hz). A klasszikus lokuszelméletnek megfelelően az F2 kontúrja palatális és veláris magánhangzók esetén eltérő: az itt közölt kísérleti adatokban veláris magánhangzók előtt emelkedő és eső is lehet, a palatális magánhangzók esetében pedig emelkedő. A magánhangzó és mássalhangzó F2 értékének azonossága csak a hátul képzett magánhangzókra érvényes.

/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, hiszen a veláris magánhangzók esetében az F2 frekvenciaértéke mellett az intenzitásviszonyok is szerepet játszanak a felismerésben. A CV kapcsolat tagjainak azonos F2 értéke csak a kerekítetlen magánhangzók esetében vezet helyes felismeréshez.

A veláris /g/ akusztikai szerkezete összetettebb a másik két zárhangénál. Az F2 értéke (ami 1740 Hz-nél nagyobb) csak az elöl képzett magánhangzók esetében befolyásolja az azonosítást, együtt az intenzitásviszonyokkal, az F1 kontúrjával 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 F2 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 F1 mozgása is.

6. ábra kb. ide

6. 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 *sz* hangot rövidítjük, a környező hangok azonossága mellett a percepció előbb *cs*-be, majd *t*-be csap át (hiszen ezen hangok képzéshelye és zöngéssége egyébként megegyezik). Gósy (1989) tanúsága szerint az azonosítás függ attól, hogy értelmes szót vagy hangkapcsolatot kell-e felismerni. Kísérleti alanyai egy kb. 100 ms-ra rövidített *sz* hangot 70%-ban *sz*-ként azonosítottak, ha a *szél* szóban kellett a hangot felismerni, és csak az esetek 30%-ban hallottak *cél*-t. Az *sz+é* 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 legalsó nyelvállású *á* a leghosszabb magánhangzó, ezért e hang mellett még egy viszonylag rövid *sz* hangot is helyesen észlelünk, míg ugyane hangot egy rövidebb intrinzikus tartamú magánhangzó mellett *cs*-nek hallanánk.

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 tranziens 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 ez a kísérlet nem szintetizált, hanem természetes beszédjelre épült.

3.3 A szupraszegmentális jegyek percepciója

Míg a beszédhangok – absztrakt egységként – diszkrét, azaz egymástól elkülöníthető egységeket alkotnak, a prozódiai jegyek csak a beszédfolyamat egészét tekintve elemezhetőek. Egy mondat hanglejtését például csak a mondat egészének ismeretében értelmezhetjük kérdő, óhajtó vagy egyéb modalitás kifejezéseként. A szupraszegmentális jegyekről nehéz általánosságban beszélni, hiszen a jelenségek nyelvenként nagyon eltérőek lehetnek, ezért a következőkben néhány kísérlet ismertetésén keresztül mutatjuk be a percepciós fonetika e szegletét.

3.3.1 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 értékelni. A monoton hanglejtés gyors beszédként való értékelése nem minden esetben, csupán a normálisnál 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 (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 észleli. Ha viszont a csúcs a mondat elején helyezkedik el, annak észlelése 90% körüli pontosságú. A mondat közepén vagy végén található csúcsok észlelése szintén pontos. (5) A mondatvég intonációjának észlelése

ereszkedő dallam esetén megbízható, emelkedő dallam esetén viszont a mondat egészének hanglejtésétől függ, csakúgy, mint a lebegő dallamok felismerése: ha a mondat elején nagyobb csúcs található, valószínűbb az ereszkedőként való észlelés.

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 eltér, leírásával számos munka foglalkozik (ld. pl. Ladd 1996). Prototipikus megvalósulására az utolsó előtti szótag magasabb frekvenciája jellemző. 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 például *És erre a tanár?*) a dallam végén emelkedéssel fejeződtek be, ezeket 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 észlelték a kísérleti alanyok.

A magyarban a mondathangsúly fontos szerepet játszik a szintaktikai szerkezet közvetítésében. Azt várnánk, hogy a szintaktikai szerkezetnek nem megfelelő hangsúly zavarokat okoz a megértésben – MacWhinney et al. (1985) kísérlete azonban ennek az ellenkezőjét támasztja alá. A kísérleti személyeknek az alany és a tárgy szerepét betöltő főneveket kellett azonosítaniuk olyan mondatokban, mint *A macska átugorja a kutyát* vs. *A macskát átugorja a kutya*. A mondatokban változott az összetevők sorrendje, és a főnevek esetragja is – olyan mondatok is előfordultak, amelyekben mindkét főnév rag nélküli volt, tehát a mondatnak két alanya volt. Emellett változott az is, hogy az első vagy a második főnévre esett a mondat fő hangsúlya. A szerzők azt találták, hogy a hangsúly csak ebben az utobbi

esetben befolyásolja a mondat értelmét, vagyis ha a morfológiai jelölés alapján nem állapíthatóak meg a szerepek. Ha a második főnévre erősebb hangsúly esett, azaz kontrasztív jelentést sugallt, akkor a kísérleti személyek szignifikánsan gyakrabban választották azt a szituációt, amelyben a második főnév szerepelt ágensként.

3.3.2 Szóhangsúly

A magyar nyelvben a szóhangsúly kötött, mindig a szó első szótagjára esik. Ez nagyban befolyásolja a hangsúlyészlelést, 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 alapfrekvencia-változás és az intenzitás összefüggését vizsgálta (Gósy 1989). A kísérlet azonos részekből álló mondatokat vizsgált, amelyekben a hangsúly különböző szavakon volt: *A városban **dél**ben harangoznak. A városban délben **ha**rangoznak*, vagy amelyek modalitása eltért egymástól (például eldöntendő kérdés). A kísérlet anyaga egyrészt glottográffal rögzített mondatokból állt, amely kizárólag az f0 változásait, azaz a hanglejtést rögzíti, másrészt imitátorral készült mondatokból, amelyekben az intenzitás is megjelenik. A kísérletből kiderül, hogy a mondathangsúly észlelése 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.

Az észlelésben nemcsak a hangsúly, hanem a hangsúly hiánya is szerepet játszik. Ezt igazolja Honbolygó et al. (2004) EEG-vel végzett kísérlete, amelyben kétszótagú szavak helyes és a szabálysértő hangsúlyozású alakját vizsgálják, vagyis olyan szavakat, amelyekben az első helyett a másodikra esik a szóhangsúly. Eredményeik szerint a prozódiai szabálysértésekre jellemző kiváltott negatív potenciál nemcsak a helytelenül hangsúlyos második szótagon jelenik meg, hanem már az első szótagon is, ha az nem kap szóhangsúlyt. Érdekes módon ez csak felnőttekre igaz: a gyerekek csupán a hibás második szótagi hangsúlyra reagálnak, a hiányára nem.

4 Az emberi beszédpercepció nagy kérdései

A hangképző szervek által létrehozott, a nyelvi kommunikáció

szempontjából fontos egységek akusztikai összetevőiket tekintve a 60 Hz és 8–10 kHz közötti frekvenciatartományba esnek. A 2. részben láttuk, hogy az emberi hallószerv különösen fogékony a néhány száz Hertz-től 10 kHz-ig terjedő frekvenciasávokra. Ennek alapján született meg a feltételezés, mely szerint a beszédprodukció és -percepció nem különálló egységként, hanem egymásra épülő és folyamatos kölcsönhatásban álló folyamatként magyarázható meg leginkább. Emellett szól, hogy a nyelvi, beszéden alapuló kommunikáció nem egyirányú, hanem többnyire folyamatosan változó szerepekkel játszódik le. Kézenfekvőnek tűnik tehát, hogy a hallgató a hallottak észlelésekor felhasználja a beszédprodukcióban felhasznált eszközökről tárolt ismereteit, a beszélő pedig alkalmazkodjon a hallgató – a beszélő saját tapasztalatai alapján feltételezett – rendelkezésére álló észlelési stratégiáihoz. A kölcsönös alkalmazkodást látványos formájában a mindennapi életben is tetten érhetjük: így például nagyothallókhoz hangosabban, a magyart nem anyanyelvi szinten beszélőkhöz pontosabb artikulációval szólunk. Ugyanígy a hallottak értelmezésekor igyekszünk a beszélő esetleges sajátságait is figyelembe venni. Ha egy kétéves gyermektől ezt halljuk: *Ellomlott tlaktó!*, akkor különösebb nehézség nélkül rájövünk, hogy a hallott /l/ hangok egy része a beszélő szándéka szerint /r/, ám a gyermek beszédkészsége még nem teszi lehetővé a komplex képzésű vibráns ejtését. A modern percepciós elméletek többsége szerint a kommunikációban részt vevők a fentiekhez hasonlóan alkalmazkodnak kommunikációs partnerükhöz, bár ennek többnyire nincsenek tudatában.

Az emberi beszédpercepció kutatása a következő kérésekre összpontosít:

1. Az emberi agy egyes részei, így például a bal félteke, a nyelvi jelek észlelésére specializálódott. Érvényes-e ez a beszédre is? Elmondható, hogy a beszédészlelés kognitív folyamatai érzékenyebbek az emberi beszéd akusztikumára, mint más hangingerekre? Azaz: létezik-e külön **beszédészlelő modul**?
2. Hogyan lesznek az észlelés során a folytonos akusztikai jelből kategóriák? Szükséges-e egyáltalán a percepcióban kategorizálást feltételezni? Más szóval: a beszédészlelés **folytonos vagy kategorikus egységeken** alapul?
3. Az emberi beszédet nagyfokú variabilitás jellemzi. Léteznek mindennek ellenére állandó, azaz **invariáns egységek**, amelyek a percepció alapjául szolgálnak, és ha igen, melyek ezek?
4. Hogyan kapcsolódik a beszédpercepció a nyelvi feldolgozás magasabb kognitív egységeihez?

A következőkben bemutatjuk a főbb percepciós elméleteket, amelyek a mai

kutatást meghatározzák. Az elméletek mindegyike empirikus adatokra támaszkodik, és meggyőző válaszokat nyújt a fent vázolt kérdésekre egyes részterületek vizsgálata alapján. Ez gyakran ahhoz vezet, hogy a különféle elméletek végkövetkeztetései szöges ellentétben állnak egymással, holott külön-külön minden elmélet hiteles bizonyítékokat sorakoztat fel a maga igaza mellett.

4.1 Speciális beszédmodul: a motoros elmélet

Amennyiben elfogadjuk egy különálló, csak az emberre jellemző beszédpercepciós modul létezését, kézenfekvő, hogy ez a modul tartalmazza a beszédprodukcióról tárolt ismereteket is. Ezen a feltételezésen alapulnak a gesztus alapú percepciós modellek.

A motoros elméletetről (*motor theory*, más nevén motoros beszédmegértési hipotézis vagy motoros teória) korábban már szó esett a mássalhangzók felismerése kapcsán. Kialakulása az 1950-es évekre tehető, a Haskins laboratóriumhoz, ezen belül elsősorban Liberman kutatásaihoz köthető (Liberman et al. 1967, később módosítása Liberman–Mattingly 1985).[8] Központi kérdése, hogy az akusztikumukban eltérő hangokat (főként a zárhangokat) hogyan vagyunk képesek invariáns kategóriákhoz kötni. Liberman és kollégái abban látják a megoldást, hogy a hallgató az akusztikai jelen keresztül észleli, hogy a beszélő milyen invariáns gesztust célzott 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 célpont és az eléréséhez szükséges mozdulatok).

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

8 Gósy (1989) felhívja rá a figyelmet, hogy Stricker Salamon már 1880-ban, majd Sarbó Artúr 1906-ban hasonló nézeteket vallott.

elkülöníteni, amint ezt Peterson és Barney kísérletében láttuk, és amint ezt saját magunkon is tapasztaljuk, ha idegen nyelvet tanulunk. Ellene szól viszont, hogy a gyermeki beszédfejlődésben a percepciós 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.[9] A tapasztalati ellenpéldákon túl megkérdőjelezhető az alapegységként meghatározott invariáns artikulációs gesztus is, hiszen az artikuláció közel sem invariáns, sem végeredményét, sem a neuromuszkuláris folyamatokat tekintve. Egyesek az artikulációs gesztus fogalmát az absztrakcióba 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.

Galantucci et al. (2009) szerint a motoros elmélet nem tudja egyértelműen megválaszolni a kérdést, hogy létezik-e speciális beszédmodul, mert nem egyértelmű, hogy a kérdés mire vonatkozik, illetve hogy mit értsünk speciális beszédmodul alatt: kizárólag a beszédfeldolgozás számára elkülönített idegpályákat, vagy hogy a beszéd azért speciális, mert a percepció magasabb szinten nem az akusztikai jelre, hanem az artikulációra támaszkodik. A motoros elmélet állítását, miszerint a produkcióért felelős idegpályák részt vesznek a percepcióban, a szerzők a tükörneuronok működésére vezetik vissza.

4.2 Folytonos és kategoriális észlelés: 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 találhatók, amelyek megkönnyítik ezek percepcióját, az észlelés tehát kategoriális (ld 7. ábra).

7. ábra kb. ide

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

7. ábra: Egy adott artikulációs és akusztikai paraméter összefüggése. Az I. és III. szakaszban a beszélőszervek 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/ képzéshelyét tekintve közel áll egymáshoz, akusztikai szerkezetük viszont meglehetősen eltérő: míg az /s/ és a /z/ spektrumára 5 000 Hz körül kezdődő, nagy intenzitású frekvenciasáv jellemző, addig az /S/ és a /Z/ frekvenciaképét 2000 és 4000 Hz körüli nagy intenzitású frekvenciakomponensek határozzák meg. Perkell et al. (1979) a következő kísérletet írják le: a nyelvperem az /s/ képzésekor a fogmedret érinti, ekkor az akusztikumot 5000 Hz körül kezdődő, nagy intenzitású frekvenciasáv jellemzi. A nyelvperem lassú hátrafelé mozgatásával a jellemző frekvenciasávok először egyenletesen, majd egy ponton hirtelen csökkennek, így az /S/ szűkülete, amely az /s/-hez képest csupán néhány milliméterrel hátrébb képződik, jóval alacsonyabb, 2000 és 4000 Hz körüli magas intenzitású frekvenciasávot eredményez.[10]

A kvantális elméletből nőtt ki a LAFF-elmélet (*Lexical Access from Features*, lexikális hozzáférési jegyek alapján), amely legújabb változatában (Stevens 2002) az észlelés alapjaként fonémák helyett absztrakt bináris megkülönböztető jegyeket feltételez, amelyek egy vagy több akusztikai kulcsban (*cue*) manifesztálódnak (például zöngés hangok esetén az alacsony frekvenciatartományban jelen lévő energia). Az akusztikai kulcsok, ill. az általuk reprezentált bináris jegyek felismerése az ugrásszerű változások létrehozta invariancián alapul. Stevens az 1980-as években kidolgozta a Chomsky és Halle (1965) által megadott artikulációs alapú bináris megkülönböztető jegyekhez tartozó akusztikai sablonokat.[11]

A LAFF-elmélet szerint a mentális lexikon a lexikális egységeket 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

10 Az /s/ magas frekvenciáinak létrejöttét az okozza, hogy a kiáramló levegő visszaverődik a felső, mások szerint az alsó fogsorról.

11 Már a bináris jegyeket elsőként bevezető Roman Jakobson is részben akusztikai alapú meghatározásokat használt.

spektrális minták.[12] A határjelzők az akusztikai jel egyes frekvenciatartományaiban észlelt hirtelen, jelentős változások, mint például egy zárfelpattanás. Az akusztikai átmenetből meghatározható, hogy egy hang például mássalhangzó ([+mássalhangzó]), ezen belül pedig zárhang ([–folyamatos]), amivel egyszersmind számos, például a magánhangzókra vagy a szonoránsokra jellemző jegy ([±kerek], [±laterális]) értelmét veszti és 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 a 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 megszünteti a fonémaegység és koartikuláció között húzódó ellentmondást, és alkalmas a koartikulációs variabilitás leírására, mivel a megkülönböztető jegyek temporális információt is hordoznak. Mellette szól továbbá, hogy egyszerű, jól megfogható jegyeken alapszik, nem feltételez nehezen igazolható absztrakt szinteket vagy egységeket, mégis jól magyarázza a percepciós 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.

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, hogy milyen lépésekben történik a megkülönböztető jegyek összehasonlítása a mentális lexikonban tárolt egységekkel: mi szolgál az összehasonlítás alapjául? Mekkora súllyal

12 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.

esnek latba az egyes jegyek?

4.3 Változékonyság és invariancia

Az európai ember számára első hallásra értelmetlen kérdésnek tűnik, hogy van-e értelme a beszédhangok állandó kategóriáiról beszélni, hiszen az Európában használatos helyesírások többségükben fonémaalapúak, azaz a hanghullámokat 30–50 betűvel vagy betűkombinációval jelzik. Amint azonban korábban láttuk, az akusztikai jel egyes esetekben igen, másokban azonban egyáltalán nem sugallja a hangok közötti határok meglétét.

Eldöntetlen kérdés, hogy az emberi percepció invariáns egységekre, vagy éppen a változékonyságra, vagyis a varianciára támaszkodik. A következőkben három további percepciós elméletet mutatunk be, amelyek az egyik, illetve a másik oldal mellett érvelnek.

4.3.1 Közvetlen realista elmélet

A közvetlen realista elmélet (*direct realist theory*) általános percepciós 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 húzódó 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 megkülönböztetni egymástól, ha azokat eltérő fonetikai kontextusban hallották (Fowler 1984).

Az elmélet jelentősége egyrészt az általános észlelési modellekbe való beágyazottság, másrészt az, hogy összekapcsolja az artikulációt, percepciót és fonológiát. Ellene szól azonban, hogy az észlelést kiváltó történések ismerete számos esetben nem szükséges a percepcióhoz: így például olyan zenét is képesek vagyunk észlelni, amit számunkra ismeretlen hangszeren 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.

4.3.2 H&H-elmélet

Lindblom *Hyper and Hypo Speech* („túlartikulált és alulartikulált") elmélete gyakorlati irányultságú: a beszélő és a hallgató között létrejövő kooperatív kommunikációra épül, amely meghatározza a beszédprodukciós és -percepciós folyamatokat (Lindblom 1990). A beszédhelyzettől és a kommunikációs partnerek közötti viszonytól függően a beszélő gondosabban vagy lazábban artikulál, ami alapvetően befolyásolja az akusztikumot, éppen ezért az nem is tartalmazhat invariáns jegyeket. A hallgató a percepciós 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 (például a beszélő anyanyelvét rosszul ismerő külföldivel szemben, fontos információ közlésekor, vagy ha zajban 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 magánhangzórendszerében is, ahol a hosszú magánhangzók egyben feszesek, a rövidek pedig lazák, azaz az artikulációs szervek utóbbi esetben nem érik el célkonfigurációjukat.

Az akusztikai jel tehát a hiper- és hipoartikuláltság 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. Nem ad viszont választ arra, hogy a leírt jelenségek mögött milyen neuropszichológiai mechanizmusok húzódnak meg.

4.3.3 Példányelmélet

Az invariancia-alapú elméletek ellenpontja egy újabb elmélet, az ún. példányelmélet. Alapjául az 1980-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 például Goldinger 1996-os kísérletében. Itt tíz beszélőtől rögzítettek összesen 300 egyszótagú szót. Először a beszélők hangja közötti hasonlóságot határozták meg egy diszkriminációs tesztben.[13] A kísérleti személyek egy 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 bemondó 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 például 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 megvalósulásaiból 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

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

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 alapaktiváltság, az aktiváltsági 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.

5 Beszédpercepció, kogníció és neuroanatómia

Míg a kognitív nyelvi modellek magukba foglalják a fonológiai, és részben a fonetikai szintet is, a bemutatott elméletek többsége nem tért ki a beszéd és a magasabb nyelvi kognitív szintek összjátékára. Egyelőre tisztázatlan, hogy a beszédpercepció milyen mértékben használja fel a nyelvi és a világról szerzett tágabb ismereteket. Nem tűnik azonban ésszerűnek, hogy a beszédészlelés során ne támaszkodjunk a nyelvi és a tágabb kommunikációs kontextusra, hiszen ezek az ismeretek minden bizonnyal támogatják a hangsorok felismerését. Az összjátékot valahogy így képzelhetjük el: ha egy nehezen olvasható kézírással írt cédula akad a kezünkbe, jóval könnyebb elolvasnunk az írottakat, ha vannak bizonyos előfeltevéseink az üzenet tartalmával kapcsolatban.

Kísérletek részben azt mutatják, hogy a magasabb kognitív szintek (szintaxis, szemantika, pragmatika) nemcsak támogatják a beszédpercepciós folyamatokat, hanem akár el is fedhetik a beszéd bemeneteként szolgáló hangsorok hiányosságait.

Warren (1970) kísérletében egy hosszabb mondat egyik /s/ hangját köhögéssel helyettesítette, és megkérte a kísérlet résztvevőit, hogy azonosítsák a köhögés helyét. Érdekes módon a kísérleti személyeknek nemcsak a pontos helyet nem sikerült meghatározniuk, hanem azt sem vették észre, hogy az /s/ hang hiányzott. Az is kiderült, hogy a résztvevők az elfedett hangot a mondat összefüggésének függvényében "azonosítják": a *Mindenki tudta, hogy a _ár…* utolsó szavának első hangját másként hallották, ha a mondat folytatása *milliókra rúg* volt, mint ha *sikeresen ellenált a török ostromnak* követte a hiányos szót. Samuel (1981) későbbi kísérletekkel igazolta, hogy a fehér zajjal elfedett mássalhangzó felismerésében a felülről lefelé ható feldolgozási folyamatok is részt vesznek: a mássalhangzó azonosítása függ a résztvevők előzetes várakozásától, a szó hosszától, azaz a rendelkezésre álló magasabb szintű információtól is. Ezzel magyarázható, hogy az elfedett hangok

behelyettesítése jóval ritkán következett be álszavak, azaz önálló jelentéssel nem bíró szavak esetén.

Samuel (1981) egyszersmind az alulról felfelé ható folyamatok jelenlétét is kimutatta: a hallgatók a fehér zaj helyén gyakrabban hallottak olyan hangokat, amelyek akusztikai jellemzőikben hasonlítanak a fehér zajhoz (például frikatívákat). Ez arra utal, hogy a felülről lefelé (*top-down*) és alulról felfelé terjedő (*bottom-up*) folyamatok párhuzamosan hatnak a beszédészlelés folyamán.

Erre utal Kazanina et al. (2006) magnetoencefalográfiás kísérlete is, amelyben orosz és koreai anyanyelvi beszélők fonémadiszkriminációs képességét vizsgálták. A /t/ és a /d/ hang mindkét nyelvben előfordul, de csak az oroszban bír jelentésmegkülönböztető szereppel, a koreaiban a /d/ intervokális, a /t/ pedig szókezdő pozícióban fordul elő, tehát egy fonéma allofónjai. A kísérletből kiderül, hogy a különböző VOT-jú, azaz eltérő zöngésségű zárhangok az orosz kísérleti személyekben nagyon korai reakciókat váltottak ki, míg a koreai beszélőkben nem – hiszen az ő beszédészlelésük szempontjából e megkülönböztetés nem releváns. A szerzők érvelése szerint a beszédhang szintű észlelést tehát befolyásolja a szavak jelentésmegkülönbözető szerepe, azaz tágabb értelemben a jelentés.

5.1 Nyommodell

A neurális pályák működését modellezi a nyommodell (*trace model*), amely a pszichológiából és a mesterséges intelligencia kutatásából 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 8. ábra egy szótalálási folyamatot ábrázol. Az egyes szinten található egységek össze vannak kötve a saját szintjükön és a szomszédos szint(ek)en található minden más egységgel. A modell bemenetét a hangszínképhez hasonló elemzés képezi, amely 5 ms-nyi szeletekre építkezik. 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.

8. ábra kb. ide

8. ábra: A nyommodell szintjei és kapcsolatuk. Példa a zöngésségi 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 például 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 bele van építve az egy szinthez tartozó szomszédos elemek gátlásának gondolata. Megmagyarázza azt a jelenséget is, hogy ha felismertünk egy szót, akkor utólag „belehalljuk" azokat a fonémákat is, amelyek ugyan a szóhoz tartoznak, de a jelből hiányoztak. Gósy (1989) felhívja rá a figyelmet, hogy ezt a jelenséget már Kempelen Farkas is leírta beszélőgépe kapcsán: amennyiben ismerjük a beszélő (ez esetben 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).

5.2 Lateralizáció

A korábbiakban láttuk, hogy a hangingerek túlnyomórészt kontralaterálisan továbbítódnak, hasonlóan a többi észlelési folyamathoz. Ennek megfelelően a beszédszegmentumok feldolgozásában főként a bal agyfélteke játszik szerepet, azonban a jobb fül elsőbbrendűsége (*right ear advantage*) nem

minden hang esetén egyforma, így például a zárhangok esetén jelentős, a magánhangzók állandó szakasza esetén csekély. A dichotikus hallást vizsgáló kísérletekben többféle fúziós jelenségről számoltak be: ha az egyik fülbe /b/-re jellemző, azaz emelkedő tranzíciójú szintetizált hangot játszanak be, a másikba /g/-re jellemző, azaz ereszkedő tranzíciójút, 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 egyértelmű formánsátmenet, a másik fül pedig változó alapfrekvenciájú 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 érjük 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).

Dehaene-Lambertz et al. (2005) szerint a percepció erősebben lateralizált, ha tudatában vagyunk annak, hogy nyelvi stimuluszoknak vagyunk kitéve. Kiváltott potenciállal és fMRI-vel végzett kísérleteikben azt találták, hogy ha a kísérleti személyek előbb szinuszzörejeket hallgattak meg, majd rávezették őket, hogy ezeket nyelvi jelként is lehet észlelni, akkor jelentősen felgyorsult a feldolgozás, és javult a kategóriahatárok észlelése – egyszersmind a kategórián belüli különbségek felismerése visszaesett. A fonémaalapú észlelés során a bal agyfélteke aktiválása jóval erősebb volt, mint a nem-nyelvi jelként észlelt zörejdiszkrimináció esetén.

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

Eckstein és Friederici (2006) arra mutat rá, hogy bár a szintaktikai szerkezet feldolgozása a bal, a prozódiai szerkezeté pedig a jobb féltekében lokalizálható, ezek a folyamatok kölcsönhatásban vannak egymással. Kiváltott potenciálokkal végzett kísérletükben kimutatták, hogy a szintaktikai szerkezettől való eltérések csak akkor jelentkeztek a bal féltekében, ha a mondat prozódiája helyes volt. Azt is megállapították, hogy ha a prozódiai és a szintaktikai szerkezet hibás, akkor a jobb féltekében mérhető a deviáns prozódiai szerkezetekre jellemző eltérési negativitás (EN), ám ha csak a szintaktikai szerkezet tér el a szabályostól, akkor a jobb

agyféltekében nem jelenik meg az EN. Ez a prozódia és a szintaxis szoros együttműködésére utal, legalábbis a szerzők által vizsgált német anyanyelvű kísérleti személyeknél.

5.3 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. Fontos szerepet játszik a szájról olvasásban, főként zajos környezetben, vagy ha egy általunk kevéssé 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 fokozott mértékben érvényes az értelmes nyelvi egységek azonosítására.

A beszédészlelést azonban nemcsak támogathatja, hanem meg is zavarhatja a vizuális információ, amint McGurk és MacDonald híres kísérletéből kiderül. Ha egy kísérleti személy fejhallgatón a /ba/ szótagot hallja, ezzel egyidejűleg pedig a képernyőn egy artikuláló személyt lát, aki a /ga/ szótagot ejti ki, az észlelt szótag /da/ lesz. Ez a McGurk-effektus,[14] 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 fúziókhoz (Roberts–Summerfield 1981).

5.4 A percepció neuroanatómiai leképezése

A bemutatott kísérletek mindegyike meggyőző bizonyítékokkal támasztja alá az egyes percepciós elméleteket, holott ezek részben meglehetősen ellentmondó nézeteket képviselnek. Valószínűnek látszik, hogy a percepció során mindazokat a folyamatokat felhasználjuk, amelyekre az egyes kísérletek rámutatnak. Ezt a feltételezést támasztják alá a hangingerek agyi feldolgozásáról szerzett eddigi ismereteink is (Péter 1991, Scott– Johnsrude 2003).

A jelenlegi kutatások alapján bizonyosnak látszik, hogy mind a hallópálya, mind az elsődleges hallókéreg ingerspecifikusan működik, a feldolgozás hierarchikus és párhuzamos szerkezetű, vagyis az idegpálya

14 Az interneten számos illusztrációt lehet találni a „McGurk effect” keresőszó beadásával.

egyes szakaszai bizonyos ingereket preferáltan kezelnek más típusú ingerekhez képest. Az agyalap és a thalamus magjai feltehetően még nem tesznek különbséget a hangforrások között, a beszédspecifikus ingerek megkülönböztetett feldolgozása valószínűleg csak az agykéregben kezdődik el. Sajnos ez utóbbi működéséről mindmáig igen kevés információ áll rendelkezésre.

Az emlősök hallókérgében az elsődleges hallókéreg több más áreával közvetlen kapcsolatban áll. Így például a tonotópiás ingerfeldolgozás az elsődleges hallókéregben történik, míg a sávok szerinti frekvenciafelismerés (amely a Bark-skála alapját is képezi) az asszociált területekre tehető (Scott–Johnsrude 2003).

A hangok észlelésében fontos szerepet játszik a Heschl-tekervény (gyrus temporalis transversus), amely a sulcus lateralisban található. Funkcionális MRI-vel végzett kísérletek azt mutatják, hogy a hanginger megjelenése, illetve az akusztikus paraméterek változása a Heschl-tekervény aktiválását váltja ki. Az aktivitás bármely, tehát nem csupán beszédspecifikus, auditív inger hatására kimutatható. A központi területhez képest anterolaterális elhelyezkedésű asszociált kéregterületek egyrészt a komplex spektrális tulajdonságú, másrészt az időbeli információ észlelését lehetővé tevő széles sávú és változó amplitúdójú hangingerek észlelésére specializálódtak. A Heschl-tekervényhez képest laterális helyzetű gyrus temporalis superior aktiválható a beszédre jellemző akusztikus ingerekkel és értelmes emberi beszéddel, de egyszersmind nyelvi információt nem tartalmazó komplex periodikus hangokkal és változó frekvenciájú ingerekkel is (Scott–Johnsrude 2003).

A nyelvi információt hordozó beszéd feldolgozása a magasabb nyelvi szintekhez hasonlóan a bal féltekében történik. Kimondottan beszédre jellemző hangingerek feldolgozására hivatott területként a baloldali sulcus temporalis superior azonosítható. Erre a területre az emlősöknél az agykéreg több pontjáról érkeznek ingerek, így a halló-, a látó-, és a szomatoszenzoros kéreg felől. Valószínű, hogy az agykéreg e pontján kapcsolódik össze a hanginger a jelentéssel, mivel a szemantikus demenciában szenvedő betegekre éppen e terület sorvadása jellemző.

Egyes kísérleti eredmények arra utalnak, hogy az általános auditív és a nyelvi információban részt vevő folyamatok a planum temporale területén kapcsolódnak össze. A planum temporale anterolaterális széle mind a beszéd szempontjából releváns, mind nem-nyelvi auditív ingerek észlelése közben aktív, a vele szomszédos, hozzá képest posterior helyzetű terület viszont artikuláció közben aktiválódik, még hangtalan beszéd közben is.

Az itt vázolt kutatási eredmények azt jelzik, hogy a hallórendszer

területi és funkcionális szempontból két részre osztható (Scott–Johnsrude 2003). A rendszer első (anterior) részében a beszédakusztikai és a lexikális információ összekapcsolódása figyelhető meg, a hátsó (posterior) rendszer pedig a hangingerek és az artikuláció kapcsolatáról árulkodik. Ez a kutatás egy későbbi pontján megteremtheti annak az alapját, hogy a különféle percepciós elméletek egymásnak részben ellentmondó szemléletét egy neurofiziológiai alapú közös rendszerben foglaljuk össze.

6 Zárszó

A fentiekből láthattuk, hogy a beszédpercepciós 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 a beszédfeldolgozás kimenetét a fonéma, a megkülönböztető jegy vagy a szó képezi-e.

Az elméletek sokféleségének hátterében az a probléma áll, hogy a percepcióban nem léteznek egy-az-egyhez típusú kapcsolatok, sőt, a több-a-többhöz 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 ismertetett kísérletek túlnyomórészt mesterségesen előállított vagy manipulált hangmintákra épültek. Ezért, bár a fentiekben a pszicholingvisztikai modellek percepciós oldalának csupán két szintjét, a fonetikai és fonológiai szintet kíséreltük meg összekötni, még ez sem járhat sikerrel a hangingerek feldolgozásának neurofiziológiai alapjainak pontosabb ismerete nélkül.

Köszönetnyilvánítás

Köszönöm mindazok segítségét, akik hozzájárultak e tanulmány elkészüléséhez: Vicsi Klára, Uwe Reichel, Hartmut Pfitzinger, Bőhm Tamás és Markó Alexandra. Az írás ideje alatt az Alexander von Humboldt Alapítvány és a Deutsche Forschungsgemeinschaft, majd az OTKA támogatását élveztem (PD 101050 sz. pályázat).

Irodalom

Bolla, K. (1995). Magyar fonetikai atlasz: a szegmentális hangszerkezet elemei. Budapest:
Nemzeti Tankönyvkiadó.

Browman, Catherine P. & Goldstein, Louis M. (1992). Articulatory phonology: an overview. *Phonetica* **49**, 155–180.

Chistovitch, Ludmilla A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America* **77**, 789–805.

Chomsky, Noam & Halle, Morris (1968). *The sound pattern of English*. New York: Harper & Row.

Dehaene-Lambertz, Ghislaine, Pallier, Christophe, Serniclaes, Willy, Sprenger-Charolles, Liliane, Jobert, Antoinette & Dehaene, Stanislas (2005): Neural correlates of switching from auditory to speech perception. *NeuroImage* **24**, 21–33.

Eckstein, Korinna & Friederici, Angela D. (2006): It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience* **18** (10), 1696–1711.

Fastl, Hugo & Zwicker, Eberhard (2006). *Psychoacoustics: facts and models, 3. kiadás*. Berlin et al.: Springer.

Fletcher, Harvey (1940). Auditory patterns. *Reviews of Modern Physics* **12**, 47–65.

Fónagy, Iván (1958). A hangsúlyról. *Nyelvtudományi Értekezések* 18.

Fowler, Carol (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics* **36**, 359–368.

Fowler, Carol (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* **14**, 3–28.

Galantucci, Bruno, Fowler, Carol A., & Turvey, M. T. (2009): The motor theory of speech perception reviewed. *Psychon Bull Rev.* ***13*** *(3)*, 361–377.

Goldinger, Stephen 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, Mária (1989). *Beszédészlelés*. Budapest: Akadémiai Kiadó.

Gósy, Mária (2004). *Fonetika: a beszéd tudománya*. Budapest: Osiris.

Greenberg, Steven, Carvey, Hanna, Hitchcock, Lea, & Chang, Shuangyu (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics* ***31***, 465–485.

Hawkins, Sarah (1999). Reevaluating assumptions about speech perception: interactive and integrative theories. In Pickett 1999, pp. 232–288.

Honbolygó, Ferenc, Csépe, Valéria, & Ragó, Anett (2004): Suprasegmental speech cues are automatically processed by the human brain: a mismatch-negativity study. *Neuroscience Letters* **361** (1), 84–88.

ISO Szabvány R 226-1961, Normal Equal-Loudness Contour for Pure Tones and Treshold of Hearing under Free Field Listening Condition.

Johnson, Keith (1997). Speech perception without speaker normalization: an exemplar model. In: Keith Johnson, & John W. Mullennix (szerk.), *Talker variability in speech perception* (pp. 145–166). San Diego: Academic Press.

Kazanina, Nina, Phillips, Colin, & Idsardi, William (2006): The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences USA* **103**, 11381–11386.

Kempelen, Farkas (1791). *Az emberi beszéd mechanizmusa, valamint a szerző beszélőgépének leírása*. Bécs: J. V. Degen.

Kent, Raymond D. (1996). *The speech sciences.* San Diego & London: Singular Publishing Group.

Ladd, Robert D. (1996). *Intonational phonology*. Cambridge Studies in Linguistics 79. Cambridge: Cambridge University Press.
Lehiste, Ilse & Peterson, Gordon E. (1961). Transitions, glides and diphthongs. *Journal of the Acoustical Society of America **33***, 268–277.
Liberman, Alvin M., Cooper, Franklin S., Shankweiler, Donald P., & Studdert-Kennedy, Michael (1967). Perception of the speech code. *Psychological Review* **74**, 431–461.
Liberman, Alvin M., & Mattingly, Ignotius G. (1985). The motor theory of speech perception revised. *Cognition* **21**, 1–36.
Lindblom, Björn (1990). Explaining phonetic variation: a sketch of the H&H theory. In William Hardcastle & Alan Marchal (szerk.), *Speech production and speech modelling* (pp. 403–439). Dordrecht: Kluwer.
Lisker, Leigh & Abramson, Arthur (1967). Some effects of context on voice onset time in English stops. *Language and Speech* **10**, 1–28.
MacWhinney, Brian, Pléh, Csaba, & Bates, Elisabeth (1985): The development of sentence interpretation in Hungarian. *Cognitive Neuropsychology* **17**, 178–209.
Mády, Katalin & Reichel, Uwe D. (2007). Quantity distinction in the hungarian vowel system—just theory or also reality? *Proc. 16. ICPhS* (pp. 1053–1056). Saarbrücken, Germany.
McClelland, James L., & Elman, Jeffrey L. (1986). The TRACE model of speech perception. *Cognitive Psychology* **18**, 1–86.
Miller, Roger L. (1953). Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America* **25**, 114–121.
Nosofsky, Robert M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* **42A**, 39–57.
Perkell, Joseph, Boyce, Suzanne E., & Stevens, Kenneth 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, Gordon E., & Barney, Harold L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* **24**, 175–184.
Péter, Ágnes (1991). *Neurológia – neuropszichológia*, 4. kiadás. Tankönyvkiadó, Budapest.
Pickett, James M. (szerk.). (1999). *The acoustics of speech communication*. Boston et al.: Allyn and Bacon.
Pisoni, David 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, Bernd (2003). *Einführung in die Phonetik, 2. kiadás*. Berlin & New York: de Gruyter.
Roberts, Martin & Summerfield, Quentin (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception and Psychophysics* **30**, 309–314.
Samuel, Arthur G. (1981), 'Phonemic restauration: insights from a new methodology', *Journal of Experimental Psychology: General* **110**, 474–494.
Scott, Sophie K. & Johnsrude, Ingrid S. (2003), 'The neuroanatomical and functional organization of speech sperception', *Trends in Neurosciences* **26** *(2)*, 100–107.
Stevens, Kenneth N. (1989). On the quantal theory of speech. *Journal of Phonetics* **17**, 3–45.
Stevens, Kenneth 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, Winifred (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America* **85**, 2207–2217.
Strange, Winifred (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, Ann 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). Functionalis anatomia I-III. Budapest: Medicina.

Tarnóczy, Tamás (1982). Zenei akusztika. Budapest: Zeneműkiadó.
Tarnóczy, Tamás (1984). Hangnyomás, hangosság, zajosság. Budapest: Akadémiai Kiadó.
Traunmüller, Hartmut (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America* **69**, 1465–1475.
Vicsi, Klára (1981a). Az időtartam szerepe néhány mássalhangzótípus hallás alapján történő megkülönböztetésében. *Magyar Fonetikai Füzetek* **7**, 59–66.
Vicsi, Klára (1981b). The most relevant acoustical micro-segment and its duration necessary for the recognition of unvoiced stops. *Acustica* **48**, 53–58.
Warren, Richard M. (1970), ‘Perceptual restoration of missing speech sounds’, *Science* **167**, 392–393.