

A sztochasztikus hallás és érzékelés: az akusztikai információ megragadásának szemlélete a hallás modellezésének figyelembevételével

WERSÉNYI GYÖRGY

Széchenyi István Egyetem, Győr
wersenyi@sze.hu

Lektorált

Kulcsszavak: akusztikai információ, hallórendszer, hallásmodellezés, beszédjelek

Az információ megragadása, rögzítése, továbbítása és végső feldolgozása a híradástechnika alapvető feladata. Az akusztikai információ kezelése nehéz feladat, hiszen azt a hanghullámok hordozzák, és rengeteg aspektusa van. Értelmezzük a hangforrás helyét, az átviteli út tulajdonságait, a forrás és a nyelő kapcsolatát. Beszédátvitel esetén pedig még természetes hibajavítással is találkozunk. Nehéznek tűnik az akusztikai információt számszerűsíteni és a megszokott információfeldolgozási és hírközléseméleti fogalmakkal kezelni. Az akusztikai információ tehát speciális szerepet tölt be az „információk között”, tekintettel a döntésképes vevőre, mely adott esetben más és más információt talál fontosnak. A cikk röviden taglalja a felmerülő problémákat, az akusztikai információ bizonyos részeinek objektív megragadhatóságát, leírását és tulajdonságait. Bevezeti a „sztochasztikus hallás” fogalmát és szerepét, végül pedig rámutat a manapság használatos hallásmodellelkel való kapcsolatára, miként kerül az információ a fül és a hallórendszer egyes lépcsőin feldolgozásra.

1. Bevezetés

Az információ létrehozása, rögzítése, átvitele és feldolgozása az akusztikában is a legfontosabb feladat. Hallórendszerünk fogadja és értékeli ki a hanghullámok által hordozott információt. Az információ mennyiség a hangforrás révén kerül bele a hanghullamba és mely az átvitel során módosulhat. A nyelő feladata a lehető legjobb kiértékelést elvégezni, akár zajos, hibákkal teli környezetben is. Az ismerős modell ellenére az akusztikai információ megragadása, számszerűsítése nem egyszerű feladat. A hírközléseméleti fogalmaink csak részben segítenek az akusztikai információ megragadása esetén. Hasznos lehet ezért áttekinteni az akusztikai információ megragadásának lehetőségeit, és azonosítani azokat a részeket, melyekhez objektív leírást (is) adhatunk.

Ebben a cikkben először az akusztikai információ szerepét mutatjuk meg, miben más ennek kezelése egyéb információk kezelésétől. Ennek lényege, hogy rámutatunk a nyelő kiemelt szerepére, a döntésképeségét hangsúlyozva. Felhívjuk a figyelmet, hogy az akusztikai információ modellezése többféleképpen is történhet, különös tekintettel a beszédjelek feldolgozására, mely kitüntetett szerepű az emberi kommunikációban. A hibajavítás szerepét így kiemelten kezeljük.

A következő részben a forrás helyének meghatározhatóságát, mint a legalapvetőbb akusztikai információ feldolgozását vizsgáljuk. A hangforrások helyének megtalálása az akusztikai kutatások legalapvetőbb és legfontosabb része, mely szoros kapcsolatban áll az objektíven mérhető fizikai mennyiségekkel is. Látni fogjuk, hogy ennek ellenére az agy, mint feldolgozó állomás szerepe fontosabb és kiterjedtebb, mint azt korábban gondoltuk. Ez a kérdés a legfontosabb akkor, amikor gyakorlati alkalmazásokra kerül sor és például vir-

tuális hangtér-szimuláció vagy hangszórós hangtérleképezés minőségét és lehetőségeit vizsgáljuk. A lokalizáció során a külső fül és annak átviteli függvénye kiemelt jelentőségű. Tény, hogy alapjában határozza meg az irányinformációt, mintegy ellátva a beérkező hanghullámot az iránytól függő szűrőhatással. Így logikusnak látszik, hogy a minél pontosabb megragadás, mérés és rekonstrukció helyes utat mutat az optimális információreprodukáláshoz. A mérések és a kísérletek azonban arra utalnak, hogy az agy szerepe itt is jelentős, az említett szűrés csak mint egy előzetes „bemenő szűrő” szerepel a feldolgozási láncban. Hasonlóan ahhoz, ahogy az analóg-digitális átalakítást is egy egyszerű aluláteresztő szűrővel kezdjük, mely szükséges de messze nem elégséges eleme az egész folyamatnak. Még ma is gyakran olvashatjuk e függvények kiemelt szerepét, mérési pontosságának növelését és egyénre szabásának fontosságát, holott elég eredmény bizonyítja, hogy képes a hallásunk felülkerekedni az átvitt hangjel apró változásain.

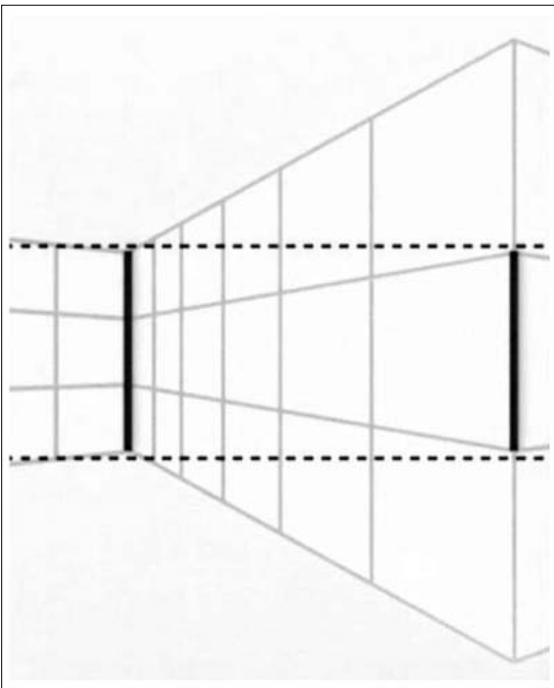
Gondoljunk csak arra, hogy ugyanaz a kimondott szó vagy mondat időtartománybeli képe soha nem egyforma, a sztochasztikus, kvázistacioner beszédjel realizációi mind különböznek a finomsztruktúrában és csak statisztikus jellemzőkkel kezeljük azt. Miért tenne a hallás másképpen? Ezért lépésenként mutatjuk be a lokalizációnál előforduló minőségi szinteket, megpróbálva felhívni a figyelmet arra, hogy az esetlegesen hiányzó és kieső információ a reprodukció során nem feltétlenül pótolható egyre pontosabb mérésekkel és a finomsztruktúra idő- vagy frekvencia-tartománybeli megragadásával. Ez kíván némi szemléletváltást a felfogásban.

A lokalizációs problémák, a fizikailag mérhető paraméterek és a szubjektív kiértékelés kapcsolatának bemutatása után egy manapság elfogadott és széles körben ismert hallásmodellt mutatunk be. A korábban felvá-

zolt információfeldolgozási lépések „anatómiai” megfeleltetése a modellezés feladata. A modell bemutatja, mely fontos feldolgozási lépcsők vannak jelen az akusztikai információ áramlása során, miközben a beérkezett hanghullám a fülektől a centrális, agyi feldolgozásig terjed. A modellezés fontos feladata a működés megismerésének, hiszen a modellek alapján épített számítógépes simulációk, kísérletek hivatottak eldönteni egy elmélet helyességét.

Az emberi érzékelés és információfeldolgozásban az agy igencsak furcsán viselkedik a hallás és a látás során. Néha nagyon gyors számítógépként, máskor viszont tehetlenséggel rendelkező és az időben kihagyásokkal dolgozó gépezetként. Nagyon jó példa erre az *időbeni elfedés*, mely hangtechnikában és képtechnikában is ismert. Előbbinél az idő és a frekvenciatarományban is megjelenik. Egy hangos hanghatás időben elfedi és észlelhetetlenné teszi az utána következő halkabb hangokat. Hasonlóan, az egymás után gyorsan felvillantott szimbólumok közül, az amelyik valami miatt kiemelkedik (például fényesebb, más színű, nagyobb stb.), fél másodpercre úgynevezett „funkcionális vakságot” okoz. Noha az ingerület továbbításra kerül, az agy nem képes azt feldolgozni, rövid idejű elfedést okozva az információfeldolgozásban.

A látásnál is rendkívül fontos az agy információszűrő képessége: a túl sok egyszerre érkező vizuális információból csak a lényegest tudja és akarja megfejteni, így nem terheli túl a feldolgozó rendszert. A népszerű optikai csalódást okozó képek is arra a jelenségre hívják fel a figyelmet, hogy az agy becsapható – pontosabban az agyi feldolgozás csap be minket – a beérkező, egyébként helyes optikai, illetve hallóinger ellenére (1. ábra).



1. ábra

Optikai csalódás.

Agyunk magasabbnak érzékeli a bal oldalon látható vonalat a jobb oldalánál, holott azok egyforma magasak.

Ez az a feldolgozási lépcső és folyamat, melyet a modellezésből semmiképp nem lehet kihagyni, és nem korlátozhatjuk az érzékelést pusztán fizikai, mérhető és leírható mennyiségek feldolgozására. A gépek éppen ezért nem tudják megoldani az internetes személyellenőrzést, vagy más néven Turing-tesztet (2. ábra).



2. ábra

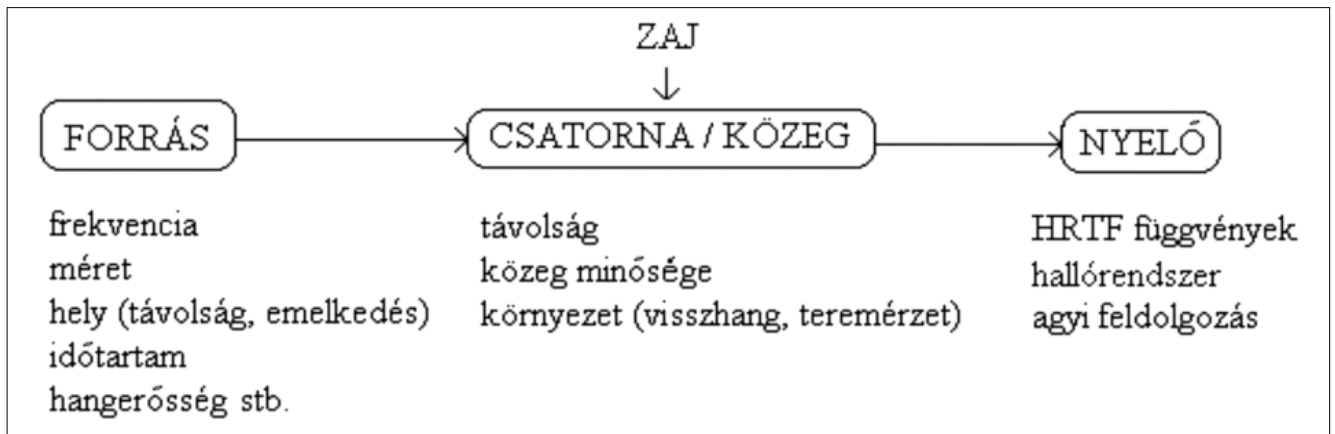
Személyellenőrző feladat számítógépes felismeréshez. Az ember könnyen felismeri a betűkombinációt ellentétben egy géppel, így kiszűrhető, hogy igazi ember vagy gépesített folyamat szeretne hozzáférni az alkalmazáshoz.

Az ember számára könnyen felismerhető és az agy számára feldolgozható betűfelismerés nem csupán optikai jellemzők alapján történik. Amennyiben tehát helyes információfeldolgozást szeretnénk, szükség lesz az agy sztochasztikus, statisztikai folyamatainak, átlagolási és információsúlyozási módszereinek felderítésére is. Ez azonban csak részben mérnöki feladat, erőteljesen összekapcsolódik pszichológiai folyamatokkal. A továbbiakban a hallás hasonló jelenségeivel foglalkozunk.

2. Az akusztikai információról

Meg kell különböztetnünk az információt az őt feldolgozó és értelmező rendszertől, mert az információ közlésének lehetősége függ a nyelőtől. A hanghullám, mint rendezett közeg-részecske halmaz (rendezett terjedési állapot), információt hordoz. A hangforrás éppúgy „rendezett állapotba” hozza a környezetét a részecskékre kényszerített rezgéssel, ahogy a részecskék is a dobhártyát. Az információ terjedése egyértelmű és egyirányú a nyelőben is: a külső fül felől terjed az agy felé. A hanghullámok nem csak láthatóan rendezett (szinuszos) alakúak lehetnek, hanem például beszéd és zene esetén, sztochasztikus jel formájában is hordozhatnak információt. Általánosságban elmondhatjuk, hogy a „valószínűbb” állapothoz, azaz, hogy a közeg részecskéi nem rendezett állapotban vannak, kevés információ és nagy entrópia tartozik. A magára hagyott rendszerek önmaguktól ebbe az irányba haladnak. Energiát kell közölni ahhoz, hogy a rendezettséget növeljük és információt tudjunk közölni.

Az információ tehát „bele van kódolva” a hanghullámokba. Ilyen formán a hang információt hordoz magáról a forrásról és az átviteli útról is, sőt bizonyos esetekben a nyelőről is (3. ábra). A továbbiakban feltételezzük, hogy a nyelő nem befolyásolja az átvitt információt (a saját fülünkkel hallunk), és az összes „dekódolási körülményt” a csatorna és a forrás határozza meg. Az



3. ábra Az információ terjedése a forrástól a nyelvőig

információt mennyiségét adott esetben függetlennek tekintjük attól, hogy a megjelenési formája az idő- vagy a frekvenciatartományban látható-e jobban.

Számunkra a nyelvő a hallórendszer, annak is az első nyomásérzékelő része, a dobhártya, ahová a $p(t)$ hangnyomás időfüggvény érkezik.

A kérdés, hogy milyen mértéket tudunk találni az információ mérésére, vagy legalábbis egy fontossági sorrend megállapítására. A szokásos információelméleti fogalom, az entrópia hasznunkra lehet, bár nehéznek tűnik számszerűsíteni az információt ebben az esetben. Egy fajta sorrend felállításában segíthet, mint az információ egyik leíró mennyisége. Ugyanakkor az akusztikai információ egyik sajátosságára is fény derül, mégpedig az intelligens (döntésképes) nyelvő miatt: az információ fontosságát a nyelvő dönti el (lásd később az E blokk szerepét a modellben). Ilyen módon pedig az entrópia és az információváltozás adott iránya esetén is tekinthetjük a „legtöbb információ” állapotát kevésbé fontosnak, mint a legalapvetőbbet. A nyelvő célirányosan működő, tehát feladatorientált multiinformációs rendszer.

Jó példa erre a beszédérthetőség és a lokalizáció összehasonlítása. Más szóval, az információ értelmének átvitele, illetve a forrás helyének meghatározhatósága. Tudjuk, hogy a forrás helyének ismerete a térben az egyik legalapvetőbb (tekinthetjük az egyik legfontosabbnak is) információ [1-4]. Ennek meghatározásához nem feltétlen szükséges „érteni” mit sugároz a forrás, fehérzajjal is elvégezhetjük a vizsgálatokat. Hasonlóan, egy tengerbe esett kiabáló ember esetén is a forrás megtalálása a legfontosabb, és nem az, hogy egész pontosan mit kiabál. Máskor viszont pusztán a beszédérthetőség a legfontosabb és a forrás helye lényegtelen, márpedig a beszédet anélkül is megérthetjük, hogy pontos lokalizációra képesek lennénk.

Ei kell tehát döntenünk a nyelvőben, hogy az adott szituációban mi a legfontosabb információ: egy virtuális valóság szimulátorban, ahol a vadászpilóta gyakorol, az ellenséges gép helyének érzékelése épp olyan fontos, mint a vele kapcsolatban álló rádiós utasításai. E kétféle információ tehát külön (is) kezelhető, de mindenképpen hatással vannak egymásra: a lokalizáció képessége és pontossága erősen függ attól, miféle jelet bocsá-

tunk ki, függ a frekvenciától, a hangerősségtől, sebességtől [1,5]. Hasonlóan, feltételezhetjük, hogy egy mondat értelmének kinyerése sem független a forrás helyének ismeretétől. Ha a beszédfeldolgozás helyes lokalizációval jár együtt, az információ kinyerésének lehetősége bizonyosan nem fog csökkenni. Helyes lokalizáció esetén ugyanis a hangforrás feltehetőleg a hallás érzékeny tartományában van (vö. odafordulunk a beszélő felé), ahol jó a térbeli felbontás, nagy a jel-zaj viszony és a külső fül szűrő hatásai a legtöbb információ kinyerését teszik lehetővé, így optimális nyereségű rendszerrel vesszük az információt. Ne feledjük, hogy a fülkagyló alakja az evolúció során azért alakult ki ilyen formára, mert szemből érkező beszédhangok esetén ez biztosítja a legnagyobb erősítést a 3-4 kHz-es tartományban [6-10]. Kijelenthetjük tehát, hogy a hangforrás helyének megtalálása nagy valószínűséggel elősegíti az információ áramlását és helyes feldolgozását, így a bizonytalanság mértékének csökkenéséhez vezet.

A nyelvőbe jutó információ mennyiség egyik része a forrásra jellemző, a másik az átviteli útra, illetve a nyelvő helyzetére, helyére. Az akusztikában ez mindennapos jelenség, a hangforrásnak létezik rá jellemző paramétere a lehallgatási pozíciótól függetlenül (például a teljesítménye) és léteznek olyan mérhető akusztikai paraméterek is, melyek a nyelvő helyzetétől függenek (hangnyomás, amit a hangforrás adott távolságra produkál). A nyelvő helyzete ilyen értelemben számunkra a csatorna és a közeg tulajdonsága, hiszen a hangforrás távolsága a nyelvőtől nem más, mint az átviteli közeg hossza.

A 3. ábra szemlélteti, mely paraméterek tekinthetők a forrás sajátjának és melyek a csatornának. Utóbbiakat úgy tekintjük, mint a forrásinformációt zavaró egyéb kiegészítő információk zavaró halmazát. Ezek többségében (gondoljunk csak a termikus zajokra, zavarokra, reflexiókra) rossz hatással vannak az információra és rontják annak dekódolhatóságát.

Egyszerű digitális forrás és zajos, tévesztő csatorna esetén most azt mondanánk, hogy hibajavító kódolásra van szükség és a hibamentes átvitelhez redundanciát kell bevinnünk. Lehetőségünk lehet hibajelzésre, hibajavításra vagy éppen az információ megismétlésre hibás vétel esetén. Ezt az analógiát itt is követhetjük,

hiszen a beszédátvitel pontosan a fentiekben leírt redundáns, hibajavítással ellátott információközlő rendszer.

2.1. Az információk egymásra épülése

Tekintettel arra, hogy az információ fontossága is szerepet játszik, nem lehet egyértelműen egymásra építeni a feldolgozási lépcsőket. Annyit állíthatunk, hogy az információ mennyiségének növekedése együtt jár a rendezettség (mint absztrakt fogalom) növekedésével és egyben az entrópia csökkenéséhez vezet. Minél több információt tudunk kinyerni, annál kevesebb a bizonytalanságunk általános értelemben.

A vizsgálathoz meg kell különböztetnünk a beszéd és nem-beszéd jeleket. A beszéd, mint az emberi kommunikáció alapja, különleges szerepű. A fülünk, a hallásunk és az agyunk is elsődleges szerepet szán a beszéd megértésének. A beszéd egyik legfontosabb jellemzője a többi jellel szemben, hogy *hibajavítást* tartalmaz. Gondoljunk arra, hogy milyen mostoha körülmények között is képesek vagyunk az értelmezésre, akár hiányos szavakat, mondatokat is ki tudunk pótolni (keresztretjvény!). Ugyanez a tulajdonság egy szinuszos jelnél vagy zajoknál nem feltétlenül található meg. Egy lehetséges információs rendszer felépítést és szerkezetet láthatunk az alábbiakban.

Ez alapján azt tekintjük, amely minden jelforrásnál hasonlóan megtalálható. A legelső szint tehát a lokalizáció lesz ebben a modellben, mert a forrás helyének megtalálása, a térben elfoglalt helye jellemző bármilyen forrásra, tekinthetjük alapvető információnak [11-13]. A hallásunk az idő- és a frekvenciatartományban is korlátozott, valamint a térbeli felbontóképessége is véges (lokalizációs bizonytalanság). Éppen ezért elmondhatjuk, hogy végtelenül pontosan sosem fogjuk a forrás helyét és az összes tulajdonságot dekódolni tudni, így sosem állíthatjuk (elvi korlát) hogy „mindent tudunk”.

A kinyerhető információ maximális mértéke tehát korlátos, az entrópia sosem csökken nullára. Az entrópia ebben a modellben lefelé csökken és úgy tekintjük, hogy ha már a helyét is ismerjük a forrásnak (az értelmén túl), akkor szinte minden információval rendelkezünk, ez az optimális információáramlás útja. Ehhez a feldolgozáshoz a külső fülre van szükség, de a tényleges feldolgozás az agyban történik. Vegyük észre, hogy ez a modell bizonyos értelemben a feje tetején áll: a hallórendszer legkülső részét tekintjük a legalapvetőbb információfeldolgozásához szükségesnek, míg az értelem kinye-

rése a másik végén található. Ne feledjük, hogy a külső fül a szűrő hatásával csak „ellátja” a beérkező jelet az irányinformációval, annak tényleges kinyerése és feldolgozása az agyban történik.

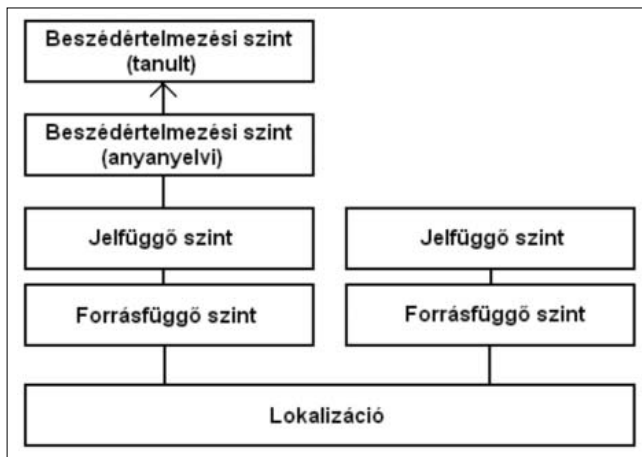
Itt egyik-másik paraméterhez rendelhetünk objektíven mérhető paramétert, mint az alaphang, formáns struktúra, hangerősség (amplitúdó), spektrum vagy a sebesség (1. táblázat). E paraméterek egy része a középfülben és a belső fülben kerülnek átalakításra az agy számára [14]. Ezek mérése, modellezése és leírása igencsak bonyolult feladat, a közép és belső fül mechanizmusaira gyakran csak a működés részleteit magyarázó részmodelleket tudunk felállítani [15-20].

A következő szint már forrásfüggő, és magát a forrást jellemzi. Ez nem-beszéd esetén például a hangszóró mérete, frekvenciatartománya, a forrás közvetlen környezete, anyaga. Érdekes, hogy a forrás méretére olyan paraméterekből (is) tudunk következtetni, mint a jelcsúcs és a burkoló. Beszéd esetén a beszélőre jellemző tulajdonságok tartoznak ide: neme, kora, a beszélőre jellemző formáns struktúrája.

A felette lévő szint már nem a forrástól, hanem a kisugárzott jeltől függ. E két réteg között nem lehet egyértelmű, éles határvonalat húzni. Ide tartozhat a kisugárzott hang frekvenciája, felharmonikus tartalma, alaphangja, esetleges torzítási jellemzői, fázisviszonyok, a jel egyéb paraméterei. Beszédnél ez teljesen hasonló, létezik a beszédnek típusa (normál, ének, torz stb.), hangsúlya és bizonyos érzelmi többlete. Akkor is képesek lehetünk az érzelmi töltet kinyerésére, ha magát a szöveget nem is értjük! Továbbá, ide vehetjük a szótagérthetőségi (logatom) vizsgálatokat, ahol már felismerjük, hogy a jel milyen (emberi beszéd), de az értelmetlen. Ennek a szintnek a legjellemzőbb tulajdonsága, hogy még nincs hibajavítás. Ilyet kísérletet gyakran az átviteli út, a berendezések vagy éppen az akusztikai környezet vizsgálatára végeznek (teremakusztika, teremjellemzők).

Innentől fogva alapjában különböznek a szintek az alatta lévőkötől, hiszen csak beszéddel lehet feljebb lépni. Ennek oka pedig a hibajavítóképeség megjelenése (4. ábra). A beszédfüggő szinten értelmezett objektív mennyiség a szövegérthetőség. Itt már értelmes szavak és szöveg hangzik el, mely kiértékelése anyanyelven a legoptimálisabb. Ezen a lépcsőn kerül kiértékelésre az elhangzott szó vagy mondat *értelme*. Ehhez már nem csak a fülekre, hanem az agy alapvető beszéd-képességére is szükség van, más szóval az emberi értelem határa húzódik itt.

	Hallásfolyamat	Forrásinformáció
1. táblázat Néhány akusztikában használatos fogalom, mérték szerepe az információ hordozásában	Frekvenciaanalízis	Spektrális tartalom
	Binaurális feldolgozás, összegzés	Akusztikus tér leképezése
	Autokorreláció a fülek jeleiben	Aktivitás, mozgás
	Rövid idejű korreláció	Időbeni változások, dinamika
	Binaurális keresztkorreláció	Lokalizáció
	Interaurális koherencia	Térbeli kiterjedés, méretek
	Fejmozgás hatása	Pontosabb, kiterjesztett lokalizáció, stabilitás



4. ábra

A feldolgozási lépcsők egymásra épülése.

Balra beszédjelek esetén, jobbra nem-beszéd jelek esetén.

A lokalizáció minden jelfajtára alapvető feladat.

A forrásfüggő és a jelfüggő szintek erősen összefüggnek, jellemzőjük a hibajavítás nélküliség.

Csak beszédjelek esetén van hibajavítás, elsősorban alapvető anyanyelvi nyelvkészség szükséges hozzá.

A különböző nyelvtudások, lexikális ismeretek stb.

már tanult folyamatok eredménye,

erősen függ a nyelvtől (intelligencia stb.)

Vegyünk egy példamondatot: Péter tegnap moziba ment. Ez egy értelmes mondat, mely magyar nyelven beszélők számára több információt is hordoz: ki, mikor, mit csinált. Megtudjuk belőle, hogy egy ember (Péter) valamikor (tegnap) valamit tett (moziba ment). Egyértelmű, hogy az információ közlése során a nyelv által hordozott természetes hibajavítás működik, így a vevőbe érkeve egy hibajavítási algoritmuson keresztül kerül feldolgozásra. Ez a nyelvi hibajavítás abból áll, hogy a mondatban szereplő szavaknak a „vevőkészülék értelmezési tartományában” rendelkezésre kell álljanak: ismerni kell azokat a szavakat és eleme kell legyen az adó és a vevő szótárának is. Ha ez nincs így, akkor az értelmezés meghiúsulhat.

Ha minket csak ez az információ (vagy ennek egy része) érdekelt, akkor a beszélő egyéb információja nem releváns, pedig a hanghullám tartalmazza a forrás jellemzőit is (férfi, nő mondta-e, hangosan, halkán, milyen érzelmi töltéssel stb.) és az átviteli útról is kapunk információt (telefonon, hangosbeszélőn, messziről, közelről, víz alatt, milyen helységben stb.). Különösen igaz ez kohéziós szövegre. A mondatok bővítésével az információ tartalom nem csökken, mert ha információhoz jutunk, a bizonytalanságuk nem nő (kivéve, ha egymásnak ellentmondóak). Egy mondat információját hangsúllyal és bizonyos nyelvekben szórendcserével is kiemelhetjük: tegnap ment Péter moziba, vagy moziba Péter tegnap ment stb.

A további szintek már az emberi értelem és intelligencia, a tudatos tanulás és hibajavítás nehezen megragadható részére vezetnek. Természetesen, ugyanaz az információ többféleképpen is kódolható, így ezek már dekóderfüggő szintek. A Péter moziba ment, a Peter went to the cinema és a Peter ging ins Kino uga-

nazt az információt hordozzák, de ennek kinyerése a nyelő nyelvtudásától függ. A különböző nyelvjárások is ide tartozhatnak. Ugyanakkor a nyelvek egy bizonyos része biológiai fejlődést is igényel, a nyelvek egy jó részét az európai ember, ha nem is érti, meg tudja ismétetni hallás után, viszont a kínait nem. Nem véletlenül okoz nehézséget a nyelvfüggetlen gépi beszédfelismerés megvalósítása. A nyelv kutatásban is igyekeznek szétbontani a szöveg-beszélő-nyelv hármast. Így születnek meg az olyan objektív mennyiségek, mint a perplexity, a Minimum Discrimination Information (MDI), a Maximal Mutual Information (MMI), a Word Error Rate (WER), vagy Phone Error Rate (PER) és az olyan feldolgozási szintek, mint az akusztikai, lexikai, vagy a nyelvi (Acoustic level, Lexican level, Language level). Utóbbihoz hasonlatos a fent vázolt modell is.

A tudatos hibajavítás jól példázza a hibajavító kódolás megjelenését zajos csatornában. Gondoljunk csak a különböző ismétlési módokra, például mikor betűzzük a szavakat a telefonban (P, mint Péter, E, mint Emil...). Vagy az egyszerű ismétlésre, amikor ugyanazt a mondatot egymás után elismételjük. A hibajavítás és az információ ismétlése redundánssá teszi a jelet. Az átviteli sebesség csökken: ugyanazt az információt több idő alatt tudjuk csak átvinni. Ez a feldolgozás az agyban (cortex) történik, ezért is nevezzük „felsőbb feldolgozásnak” [21-24]. Ez a szint az emberi egészséges értelem meglétét megköveteli. A feldolgozás szintje itt nem állandó, tanulással, tudatos neveléssel javítható vagy éppen gátolható.

Hangsúlyozni kell, hogy ez a modell egy lehetséges elképzelés, amely legfontosabbnak a beszédérthetőséget tekinti. Nem beszédjel esetén pedig a kibocsátott jel általános értelemben vett értelmét, tulajdonságait, minőségét. A további paraméterek, mint a forrás tulajdonságai, másodlagosak. A legkevésbé fontos pedig a hangforrás lokalizációja, így a feldolgozásban felülről lefelé haladva szerzünk egyre több (és egyre kevésbé fontos) információt, mely a bizonytalanságot és az entrópiát csökkenti, az ismereteket növeli. Érdekes, hogy a modell fordítva is elképzelhető, ha a lokalizációt és a hangforrás helyét tekintjük a legfontosabb információnak, ugyanakkor a forrás egyéb paramétereit vagy „mondandójának” értelmét kevésbé lényegesnek.

2.2. Becsapható-e az agy?

A továbbiakban a hangforrás helyének meghatározhatóságát vizsgáljuk, melyet lokalizációnak hívunk. Ehhez a külső fül átviteli függvényére van szükség, melyet az angol rövidítésből HRTF (Head-Related Transfer Function) függvényeknek hívunk. Közöséges komplex átviteli függvények, melyek egy adott térirányhoz tartoznak, és egy egész készletre van szükségünk belőle. Az irányinformáció e függvények által kerül bele a hanghullamba: eltérő térirányokból eltérő HRTF függvény végzi a szűrést.

Virtuális szimulációban, fejhallgatót használva, ezeket a függvényeket mérni, tárolni kell és lejátszáskor

szűrést kell velük végezni. Ennek lehetőségei korlátozottak, elsősorban attól függ, hogy szimuláció során a hallgató a saját, úgynevezett individuális HRTF függvényeit kapja-e, vagy más emberi, illetve műfejen rögzítetteket-e (nem individuális készlet). Azt a folyamatot, amikor a két fül számára fejhallgató lejátszást biztosítunk korábban rögzített HRTF függvények és az átviteli út kiegyenlítésének segítségével, binaurális technológiának nevezzük.

Kutatásaink egy része arra irányul, hogy a már említett külső fül átviteli függvényeket mérjük és kiértékeljük. Ez a szakasz röviden értékeli a külső fül szerepét az információ feldolgozásában. Vizsgálataink során műfejes, süketszobai mérésekkel igazoltuk, hogy a fülelhez közeli akusztikai környezet hatása mérhető, szignifikáns és jelentősen befolyásolja az átviteli függvényeket, melyeknek fontos szerepe van az irányinformáció érzékelésében [25-28].

A valóságban, a szabadtéri hallás során, amikor nem fejhallgatóval dolgozunk, az információ átvitele nem sérül. A hallórendszer – mely a külső fültől az agyi feldolgozásig tart – képes felülkerekedni akár 15 dB-es ingadozásokon is anélkül, hogy lényegében megváltoztatná a hangtér érzékelését és az akusztikai hullámban kódolt információ dekódolhatóságát. Ez az a pont, ahol a sztochasztikus szemléletmódot be kell vezetnünk, hiszen nyilvánvaló, hogy a hallórendszer az egyes realizációk finomsztruktúrájával itt nem foglalkozik. Összehasonlítva két kimondott „a” betű időtartománybeli képét, hasonlóságot láthatunk, de egyezést aligha. A hasonlóságok azok, melyeket a sztochasztikus leírásmóddal matematikailag is kezelhetünk. Sztochasztikus hallásnak nevezhetjük azt a felfogást, ahol a hallórendszert komplex információfeldolgozó állomásnak tekintjük, mely alapállapotában a jelek részleteit elhanyagolva végzi a feldolgozást és csak akkor kapcsol finomsztruktúra kiértékelésére, ha az elsődleges működése valamilyen oknál fogva sérül.

Ismert tény azonban a másik oldalról, hogy ha fejhallgatót használunk és mért átviteli függvényekkel dolgozunk, azok „minősége” erősen befolyásolja a kísérletek, elsősorban a lokalizációs problémák eredményét [3, 11, 29-32]. Úgy tűnik tehát, hogy ebben az esetben a kisebb változások a függvényekben zavarként kerülnek kiértékelésre és a mért eredmények rosszabbodásához vezetnek. Magyarán, ha normál környezetben teszem fel a napszemüvegem vagy vágatom le a hajam, nem fogok változást érzékelni, ha viszont ezt virtuális környezetben teszem meg (szimulálom ezek hatását az átviteli függvényekben), akkor ez érzékelhető romláshoz vezethet. Ennek magyarázata az lehet, hogy a hallórendszer képes (egyéb paraméterek figyelembevételével) érzékelni azt, hogy virtuális vagy valóságos térben létezik-e.

Feladatunk tehát azon paraméterek megtalálása, amelyeket az agy még felhasznál e folyamatok során, mert az nyilvánvaló, hogy nem csupán a külső fül szűrőhatása vesz részt a lokalizációban. Ilyen lehet például a fejmozgás által szerzett információ, a teremakusztikai

visszaverődések szerepe, illetve a legfontosabbnak tűnő: a fejhallgató szerepe [33].

A fejhallgató lejátszás hibái, a fejközép-lokalizáció, az elől-hátul döntések hibái, de akár a fejhallgató viselésének kényelmi szempontjai is befolyásolhatják az érzékelést. Amíg a kísérletben résztvevő alanyok több, mint a fele beszámol ilyen jellegű hibákról, nincs értelme a problémát a szimulált paraméterekben keresni [34-36]. Lehet, hogy a probléma szimulációs eszközben rejlik.

A fejlődés útja az egyre jobb virtuális szimulációhoz nem úgy vezet, hogy egyre jobb, pontosabb műfejeket építünk, modellezünk, és egyre pontosabb HRTF függvényeket nyerünk méréseinkkel, hanem további paraméterek meghatározására. Ezek a csak részben meghatározott, nem túl pontosan mért „rosszabb minőségű” függvényeket használva is képesek a szimuláció minőségét javítani. Az agyat nem olyan egyszerű becsapni!

2.3. Minőségi szintek

lokalizációs feladatok megoldásakor

Nézzük meg röviden, milyen minőségi szintek és lépések vannak a hallásfolyamat során, amikor lokalizációs kérdéseket vizsgálunk!

1. A legjobb eredményeket szabadtéri hallás során érzük el, fejhallgató nélkül, a megszokott saját HRTF függvényeink használatával. Individuális HRTF készlet, végtelen térbeli felbontás, apró fejmozgások segítik a folyamatot. Habár a reflexiók és a terem paramétereitől függetlenül szoktuk a méréseket végezni, a valóságban ezek is rendelkezésre állnak és egyelőre nem teljesen tisztázott, hogy jelenlétük segíti-e vagy gátolja a lokalizáció képességét [1, 4]. Az nyilvánvaló, hogy a túl sok reflexió (zengő terek) gátolják, de „valamekkora” utözengés hasznos lehet, hogy ne érezzük idegennek a süket környezetet. Továbbá rendelkezésünkre állhat a látás (amely erősen befolyásolja a hangtér érzékelését, mert képes a vizuális információ felülírni a hallott információt).

2. Mesterségesen előállított szabadtérben, ahol hangszórókkal dolgozunk, csak véges számú forrást tudunk létrehozni. Süketszobában, egy vagy néhány hangszóró által keltett szintetizált hangtér akkor is korlátozott, és okoz romlást a lehetőségekben, ha egyébként saját hallásunkat használjuk. Ilyenkor a lokalizációt elsősorban a hangszórók száma, elhelyezkedése, mérete, műszaki paraméterei (minősége), valamint a kibocsátott jel paraméterei határozzák meg (hangerősség, spektrum, időbeni hosszúság és változások). Eddig a pontig a saját HRTF készletünket használtuk, és a hallórendszer képes az abban létrejövő változásokat részben elhanyagolni.

3. A virtuális szimuláció és a fejhallgató bevezetése a kísérletben lényegében változtatja meg a vizsgálatot és az eredményeket. Általánosságban igaz, hogy az ilyen szimuláció során rosszabbak a lokalizációs feladatok elvégzésének eredményei. A szimulált HRTF függvények száma limitált, amely magával hozza a véges

térbeli felbontást. Sokszor a hiányzó méréseket interpolált függvényekkel pótolják. Ebben az esetben az agynak nagyobb kapacitására van szüksége, és több energiát fordít a lokalizációra, mert ezt a feladatot egy idegen, nem természetes környezetben kell elvégeznie. A HRTF-ek változása, az individuálistól való eltérése rontja a lokalizációt így igyekezünk mindent elkövetni ennek minimalizálására. A legjobb eredményeket individuális, nagy mennyiségű HRTF-ekkel és kiegyenlített fejhallgatóval érhetünk el.

4. Amennyiben a kiegyenlítést és a szimulációt nem komplex módon végezzük, hanem a fázisinformáció elhagyásával pusztán az amplitúdó menetét vesszük figyelembe, romolhatnak az eredmények. Külön kérdés, hogy ezt mikor tehetjük meg, milyen feltételek mellett minimálfázisú a HRTF szűrőrendszer és hogyan vesszük be az időbeni késleltetések szerepét [37, 38].

5. Nem individuális HRTF-ek használata tovább rontja az eredményeket. Mivel ezek mérése körülményes,

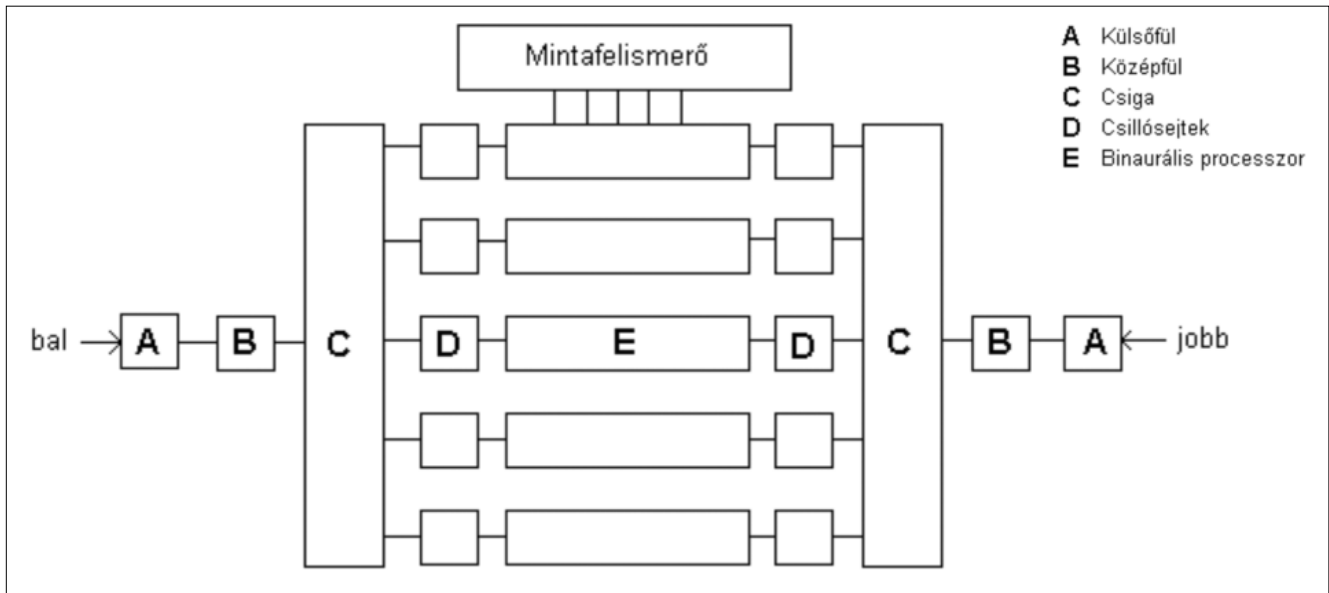
időigényes és drága, a legtöbbször előre megmért emberi HRTF-eket használunk. Ezek ugyan emberek, de nem individuálisak, használatuk általános. Léteznek egyszerű módszerek ezek „továbbindividualizálására”, a fejméret beállítására stb. A legoptimálisabb megoldásnak az tűnik, ha egy alapjában jó lokalizációs képességekkel rendelkező alany pontos, nagy mennyiségben megmért HRTF függvényeit használjuk fel.

6. A HRTF függvények számának csökkenése egyértelműen káros, mert csökkenti a térbeli felbontóképességet. Ez ahhoz vezet, hogy egyre több köztes iránynak megfelelő függvényt kell interpolációval vagy számításokkal meghatározni, melyek csak közelítők lehetnek.

7. További romláshoz vezet a véletlen módon kiválasztott, tetszőleges emberi fejen mért függvények bevetése, majd az átlagos emberi fej, végül pedig a műfejes HRTF-ek következnek. Elmondhatjuk tehát, hogy a legjobb az individuális szűrőkészlet, aztán a véletlen

2. táblázat A szabadtéri hallás és a virtuális, fejhallgató környezet összehasonlítása [28]

	Szabadtéri hallás	Virtuális szimuláció
Individuális HRTF függvények használata	mindig	ritkán
HRTF használata más alanyoktól, műfejtől	—	gyakran
Fejhallgató átviteli függvényének kiegyenlítése	—	mindig
Végtelen számú (végtelen finom térbeli felbontású) HRTF függvénysereg használata	mindig	soha
Véges számú HRTF használata (véges térbeli felbontóképesség)	—	mindig
Interpoláció használata hiányzó HRTF függvények kiszámítására	—	néha
Dinamikus HRTF függvény változtatás a fej kis mozgásainak megfelelően	mindig (kivéve rögzített fejnél)	nagyon ritkán (szimulációval segíthető)
Fejközép lokalizáció (hibaforrás)	soha	gyakran
Elöl-hátul döntések hibája	ritkán	gyakran
Emelkedés eltolódás és a forrás túl közeli érzékelése	nagyon ritkán	gyakran
Visszaverődések jelenléte	nem (süketszobában) igen (normál szobában)	nem (szimuláció nélkül) ritkán (szimulálás esetén)
A használt HRTF függvények finomstruktúrája és változásának fontossága	nem	igen
Egyéb lokalizációt befolyásoló paraméterek szerepe (fej forgatása, teremakusztikai paraméterek stb.)	igen	igen



6. ábra Blokkdiagram a hallás feldolgozási lépcsőjéhez [42]

emberi fej és az átlagos emberi fej, végül pedig a műfejek függvényei.

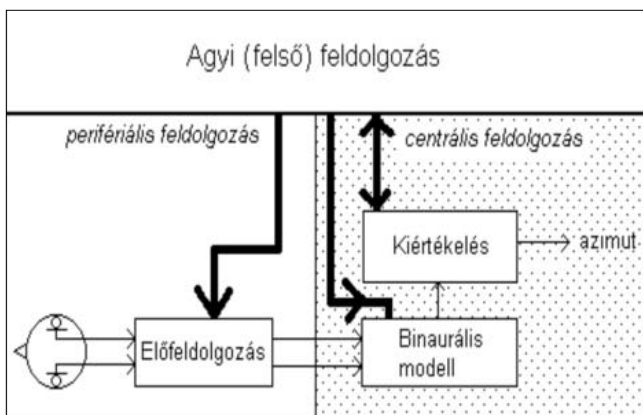
8. Bármilyen kalkulált, szimulált, elméleti megközelítésből vagy modellezésből és nem mérésből származó HRTF közelítések elégtelenek bizonyulnak. Próbálkozta gömbként modellezett fej esetén kiszámított függvények és hangtér szimulációjával, de ezek eredményei elmaradnak a fentiekől, elsősorban a másodlagos és harmadlagos csúcok-leszívások hiányának következtében [39]. Ez a HRTF függvények finomstruktúrájának szerepére hívja fel a figyelmet, egy olyan határ létezésére, amelyet el kell érni ahhoz, hogy a HRTF függvény betölthesse a szimulációban a szerepét.

9. Végezetül, a hiányzó fejhallgató kiegyenlítés, a rossz minőségű eszközök használata elfogadhatatlan tudományos kísérletekhez. Ne feledjük azonban, hogy a nem tudományos, hanem mindennapi alkalmazásokban ezek meglehetősen ritka. A felhasználók többsége otthon nem tudományos minőségnek megfelelő fej-

hallgatót és eszközöket használ, és bizonyosan nincs kiegyenlítve az átviteli út. Ennek ellenére korlátozott lehetőségekkel hozhatunk létre irányinformációt és hangtér szimulációt.

Látható, hogy igazi „letérés” a minőségben a második pont után történik, amikor áttérünk a saját fülünkről a fejhallgató szimulációra, így jogosnak látszik a felvetés, hogy a hibát ebben az eszközben (is) keressük. A hallórendszer elveszíti azt a képességét, hogy az akusztikai irányinformációt a HRTF függvények nagy és gyors változása ellenére is sikeresen feldolgozza, és hirtelen számítani kezd a HRTF-ek minősége, származási helye. Talán szerepe lehet az elvesztett levegőbeni terjedés hatásának és a fejhallgató rendszerben valamilyenre mégis részt vevő fülkagyló hatásnak. A fejhallgatók többsége supra-aurális, azaz a fejre illeszkedik, körülveszi, illetve ráfekszik a fülkagylóra, mely kapcsolódás befolyásolhatja az átvitelt [40].

Ma már tudjuk, hogy egyéb paramétereket is figyelembe kell vennünk, elsősorban a fej mozgatót és a reflexiók szimulációját ahhoz, hogy jobb eredményeket érjünk el fejhallgató lejátszás során (2. táblázat).



5. ábra

A hallórendszer feldolgozási lépései.

A perifériális (fülek szintjén) lévő feldolgozást éppúgy felülegyeli az agy, mint a centrális lépéseket, melyek során a két füljel binaurális és interaurális kiértékelése történik [42].

3. A hallás egyszerű modellje

A hallás modellezésekor a külső, a belső, a középfül és az agy szerepét próbáljuk meghatározni és modellezni. Ennek része a számítógépes szimuláció is, melynek során a géppel próbáljuk elvégezteni a két hangjel ismeretében azt a feladatot, melyet az agy is elvégez [41-43].

A klasszikus felépítés szerint a külső fül feladata az irányinformáció kinyerése, a közép és a belső fül végzi a frekvenciatartománybeli analízist, és az agy a többi kiértékelést. Ezt a modellt már revidiáltuk, rámutattunk arra, hogy az agyi feldolgozás nem e „soros” lánc végén található, és a feldolgozás során nem egymásnak adják az információt a fül részei. Sokkal inkább párhuzamosan történik a feldolgozás a fül részei között.

zamos a feldolgozás, az agy „mindent lát” és felügyel. Ennek értelmében létezik centrális és perifériális feldolgozási lépések, ahogy azt az 5. ábra is mutatja [44-48]. A perifériális előfeldolgozás a HRTF szűrésnek felel meg, míg a tényleges helyzetmeghatározás (a forrás kitérése a fej szimmetriasíkjából, az azimút) már centrális agyi folyamat.

A hallás részletesebb modellezésére az alábbi ábra segítségével mutathatunk rá. Ez jól mutatja az egyes blokkok szerepét és kapcsolatát (6. ábra). A modell számítógépes szimuláció során a két fül jelének kereszt-korrelációs függvényét számítja ki, a koherens részeket azonosítva a két jel között.

A modell szerint a külső fül az irányfüggő HRTF szűrés nyomait helyezi el a beérkező jelben. Ez egy előfeldolgozó állomás. A forrás oldalirányú kitérései miatt interaurális időkülönbség lép fel a két fül jele között, az árnyékolás és az elhajlások okán pedig további szinteltérés is. Ez az irányhallás, a sztereó leképezés alapja is. A középfül már irányfüggetlen, lineáris átvitelrel rendelkezik.

A C blokkban, a csigában történik a hely-frekvencia transzformáció: a baziláris membránon terjedő haladó hullám maximumai ingerlik a csillósejteket, melyek az ingerületet közvetítik. A maximumok kialakulása a frekvencia függvénye, ezért ez tulajdonképpen egy „biológiai FFT” feldolgozás. Ezt a hatást sáváteresztő szűrőkkel modellezhetjük (melyek kapcsolatban állnak az úgynevezett kritikus sávokkal). 24 darab aszimmetrikus szűrővel modellezve, ahol az alsó rész negyedrendű felüláteresztő, míg a felső rész tizedrendű aluláteresztő. Ez egyben az „analóg-digitális átalakítás” helye is: az analóg kitérésekből sztochasztikus, időben diszkrét ingerület sorozat keletkezik. Ezt a feladatot a D blokkban a csillósejtek (szőrsejtek) végzik, melyek a kitéréssel arányos elektromos ingerületet állítanak elő és a hallóidegen át az agyba továbbítják azokat.

A binaurális mintafelismerő feladata az ingerület kiértékelése, a korreláció kiszámítása és a hangforrás helyének meghatározása. Ehhez a hallás – jelenlegi ismereteink szerint – többek között felhasználja a két fül jel korrelációját, koherencia fokát, időbeli és szintbeli eltéréseit, melyek az iránytól függően speciális koincidencia mintát rajzolnak ki a frekvencia-késleltetési idő síkon (innen a mintafelismerő elnevezés) [42,45,49-51].

Az E blokk végül egy súlyozást ad a modellhez, mely egy minősítő függvénnyel látja el a kialakult mintát a felismerő számára. A súlyozás szerepe, hogy a hasonló minták között különbséget tegyen, a „nem túl valószínű” hangforrás helyeket lepontozza, míg a nagyobb valószínűségűeket felértékelje. Tulajdonképpen egy valószínűségű függvény által történő súlyozás ez, mely függvénynek van egy maximuma, amelytől jobbra és balra monoton csökken.

A mintafelismerő egy függvényhalmaz, ahol a D és az E blokk információja együttesen kerül feldolgozásra, és ahol a tényleges binaurális lokalizáció és jelfelismerés megtörténik. Minden egyéb felső (centrális agyi) és tanult feldolgozás innen ered.

Ez a modell nem univerzális, csak némely aspektust magyaráz. Így a monaurális feldolgozást nem. A monaurális, azaz egyfülű hallás általában egyszerűbben és a binaurálissal párhuzamosan kezelhető. Egy füllel ugyanis elsősorban a források távolságát (annak hangereósége alapján) és csekély irányinformációt tudunk csak kinyerni, rossz térbeli felbontás mellett. Ezért ezt a működést gyakran csak távolsághallásra redukáljuk és a fenti modellel párhuzamos működésként szoktuk figyelembe venni.

Látható, hogy ebben a modellben – helyesen – egyszerre szerepel a perifériális előfeldolgozás, a centrális felügyelet és tartalmaz objektíven számolható mennyiségeket (pl. keresztkorreláció). Ugyanakkor sztochasztikus jellemzőkkel és sztochasztikus működéssel keveset foglalkozik. Sztochasztikus modell hozzáadása, a hallás azon képessége, amikor képes gyorsan változó környezetben statisztikus jellemzőkkel dolgozva a helyes információ kinyerésére a jövő kutatásainak feladata.

4. Összefoglalás

Az emberi hallórendszer a dobhártyákon fellépő hangnyomás időfüggvényéből nyeri ki az akusztikai információt. Ennek első, előfeldolgozást végző része a külső fül szűrőhatása, mely alapvető irányinformációval látja el a beérkező hanghullámot. A perifériális feldolgozást és a centrális kiértékelést egyaránt az agy felügyeli, természetes hangtérben a sztochasztikus jellemzőkre építve és a finomstruktúra elhanyagolásával. Ugyanakkor, virtuális hangterekben, fejhallgató és mesterségesen előállított szűrők, szimuláció segítségével rosszabb eredményeket kapunk. Nem sikerül az agyat becsapni és így az képtelen lesz a sztochasztikus feldolgozást végezni.

Ennek javítása azonban nem az egyre pontosabb szimulációban, hanem a módszer, a szemlélet megváltoztatásában és egyéb paraméterek bekapcsolásában rejlik. Az információáramlás útján végigvonul a részparaméterek finom megváltozását felfogó, elemző, majd döntő agyi funkció, tehát ezen a fokon multifunkcionális paraméter-vezérelt rendszerként működik a hangjel feldolgozás. Ha elég bátrak vagyunk, odáig is elmehetünk, hogy elfogadjuk a virtuális szimuláció korlátait, annak csökkentett lehetőségeit a szabadtéri halláshoz képest. Szétválaszthatjuk az információ feldolgozásának kérdését és a lokalizáció problémáját virtuális valóságban. A kísérletek irányulhatnak arra, hogy egy adott virtuális szimuláció lehetőségeit megvizsgáljuk, illetve arra, hogy a hallás ma ismert modelljét kiterjesszük.

A bemutatott modell képes az információ feldolgozásának legfontosabb lépéseit – jelenlegi tudásunk szerint – az anatómiai részekhez hozzárendelni, és a modell alapján számítógépre bízni a feladatok megoldását. Aki hallásmodellezésre adja a fejét, olyan számítógépes szimulációkkal fog találkozni, mely a két fül jelét használja bemenetnek, és amely a kimenetén a hallásélménnyel azonos információt tud szolgáltatni.

Ez a bizonyítéka annak, hogy a modell helyes: pusztán számításokkal eljutni olyan következtetéshez és döntéshez, mely megegyezik a pszichoakusztikai tapasztalatokkal. A sztochasztikus hallásmodellezés az egyik legjobban járható útnak tűnik ebben a feladatban.

Irodalom

- [1] J. Blauert:
Spatial Hearing.
The MIT Press, MA, 1983.
- [2] W. M. Hartmann:
How we localize sound.
Physics Today, pp.24–29., November 1999.
- [3] J. C. Middlebrooks, D. M. Green:
Sound localization by human listeners.
Ann. Rev. Psychol. 42, pp.135–159., 1991.
- [4] M. Kleiner, B. I. Dalenbäck, P. Svensson:
Auralization – an overview.
Journal Audio Eng. Soc. 41, pp.861–875., 1993.
- [5] Gy. Békésy: Experiments in hearing.
McGraw-Hill Book Co., New York, 1960.
- [6] P. Berényi, Gy. Wersényi:
A külső fül fejre vonatkoztatott átviteli függvényeinek vizsgálata. Akusztikai Szemle, IV. évf., 1-4.szám, pp.35–41., Budapest, 1999.
- [7] T. Tarnóczy:
Az egyfülű hallás síkbeli iránytól függő szinképi torzításai. Beszédkutatás 1995, pp.4–19., Budapest, 1995.
- [8] C. I. Cheng, G. H. Wakefield:
Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency and Space.
Journal Audio Eng. Soc. 49, pp.231–249., 2001.
- [9] H. Møller, M. F. Sorensen, D. Hammershøi, C. B. Jensen:
Head-Related Transfer Functions of human subjects.
Journal Audio Eng. Soc. 43(5), pp.300–321., 1995.
- [10] D. Hammershøi, H. Møller:
Sound transmission to and within the human ear canal.
Journal Acoust. Soc. Am. 100(1), pp.408–427., 1996.
- [11] P. Minnaar, S. K. Olesen, F. Christensen, H. Møller:
Localization with Binaural Recordings from Artificial and Human Heads.
Journal Audio Eng. Soc. 49(5), pp.323–336., 2001.
- [12] J. Kawaura, Y. Suzuki, F. Asano, T. Sone:
Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear.
Journal Acoust. Soc. Japan E12, pp.203–215., 1991.
- [13] Gy. Wersényi:
Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications. Audio Engineering Society (AES) Convention Preprint Paper, Nr. 5902, Pres. at the 115th Convention, New York, 2003.
- [14] A. Illényi:
Decoding the source information by dichotic hearing. Proc. of Inter-Noise 97, Vol. II., pp.1147–1152., Budapest, 1997.
- [15] K. Hartung:
Modellalgorithmen zum Richtungshören, basierend auf den Ergebnissen psychoakustischer und neurophysiologischer Experimente mit virtuellen Scallquellen.
Dissertation, Ruhr-Universität, Bochum (1997), Shaker Verlag, Aachen, 1999.
- [16] K. Genuit:
Eine systemtheoretische Beschreibung des Aussenohres. DAGA'85, pp.459–462., Stuttgart, 1985.
- [17] K. Genuit:
Ein Modell zur Beschreibung der Außenohreigenschaften. Dissertation, TH Aachen, 1984.
- [18] C. Jin, M. Schenkel, S. Carlile:
Neural system identification model of human sound localization.
J. Acoust. Soc. Am. 108(3), pp.1215–1235., 2000.
- [19] W. Gaik:
Untersuchungen zur binauralen Verarbeitung Kopfbezogener Signale.
VDI Fortschrittberichte, Reihe 17, Biotechnik, Nr. 63, VDI Verlag, Düsseldorf, 1990.
- [20] M. Bodden, G. Canévet, J. Grabke, K. Hartung, T. Takahashi:
Raumliches Hören in komplexen akustischen Umgebungen, DAGA 94, 1994. pp.1137–1140.
- [21] K. Hartung, M. Bodden, J. Grabke, K. Rateitschek:
Eine physiologienahe Strategie zur Modellierung des binauralen Hörens, DAGA 94, pp.1141–1144.
- [22] J. Grabke:
Modellierung des Praezedenzeffektes, DAGA 94, pp.1145–1148.
- [23] M. Bodden:
Verifikation der räumlichen Abbildung bei Verwendung von Signalaufzeichnungssys. mit Datenreduktionsverfahren mittels eines binauralen Modells, DAGA 94, pp.1153–1155.
- [24] M. Bodden, K. Hartung, K. Rateitdchek, O. Dubielzig:
Psychoakustische Untersuchungen zur Lokalisation in reflexionsbehafeter Umgebung, DAGA 94, pp.1169–1172.
- [25] Gy. Wersényi, A. Illényi:
Differences in Dummy-Head HRTFs Caused by the Acoustical Environment Near the Head. Electronic Journal of „Technical Acoustics” (EJTA), Russia, January 2005 (1), pp.1–15. <http://ejta.org/en/wersenyi1>
- [26] A. Illényi, Gy. Wersényi:
Evaluation of HRTF data using the Head-Related Transfer Function Differences.
Proc. of the Forum Acusticum 2005, Budapest, pp.2475–2479.

- [27] A. Illényi, Gy. Wersényi:
Environmental Influence on the fine Structure of Dummy-head HRTFs.
Proc. of the Forum Acusticum 2005, Budapest, pp. 2529–2534.
- [28] Gy. Wersényi:
On the amplification-damping-effect of the pinnae and of the head. *Acoustical Review*, Vol. VI., Nr.2, pp.23–28., Budapest, 2005.
- [29] J.C. Makous, J.C. Middlebrooks:
Two-dimensional sound localization by human listeners. *Journal Acoust. Soc. Am.*, Vol. 87(5), May 1990. pp.2188–2200.
- [30] H. Møller:
Fundamentals of binaural technology. *Applied Acoustics* 36, (1992), pp. 171–218.
- [31] A. Schmitz:
Diskussion verschiedener Verfahren zur Wiedergabe kopfbezogener Signale. *DAGA 94*, pp.277–280.
- [32] J. Blauert, H. Lehnert:
The auditory representation in virtual reality. Proc. of the 15th Int. Congress on Acoustics, Trondheim, Norway, 26-30 June 1995, pp.207–210.
- [33] D.R. Begault, E. Wenzel, M. Anderson:
Direct Comparison of the Impact of Head Tracking Reverberation and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *Journal Audio Eng. Soc.* 49(10), pp.904–917., 2001.
- [34] P. Laws:
Entfernungshören und das Problem der Im-Kopf-Lokalisierung von Hörerignissen. *Acoustica* 29, pp.243–259., 1973.
- [35] G. Plenge:
Über das Problem der Im-Kopf-Lokalisation. *Acoustica* 26, pp.241–252., 1972.
- [36] Gy. Wersényi:
HRTFs in Human Localization: Measurement, Spectral Evaluation and Practical Use in Virtual Audio Environment. Ph.D. doctoral thesis, BTU Cottbus, 2002.
- [37] P. Berényi, A. Illényi:
What does it mean for an HRTF not to have the minimal phase property? Proc. of Inter-Noise 96, pp.2127–2130., Liverpool, 1996.
- [38] D.J. Kistler, F.L. Wightman:
A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.* 91, pp.1637–1647., 1991.
- [39] V.R. Algazi, C. Avendano, R.O. Duda:
Estimation of a spherical-head model from anthropometry. *Journal Audio Eng. Soc.* 49(6), pp.472–479., 2001.
- [40] H. Møller, D. Hammershøi, C.B. Jensen, M. Sorensen:
Transfer Characteristics of Headphones Measured on Human Ears. *Journal Audio Eng. Soc.* 43(4), pp.203–216., 1995.
- [41] E. Zwicker, R. Feldtkeller:
Das Ohr als Nachrichtenempfänger. S.Hirzel Verlag, Stuttgart, pp.181., 1967.
- [42] M. Bodden:
Binaurale Signalverarbeitung: Modellierung der Richtungserkennung und des Cocktail-Party-Effektes. VDI Fortschrittberichte, Reihe 17, Biotechnik, Nr.85, VDI Verlag, Düsseldorf, 1992.
- [43] L. A. Jeffress:
A place theory of sound localization. *J. Comp. Physiol. Psychol.* 61, pp.468–486., 1948.
- [44] W. Lindemann:
Extension of a Binaural Cross-Correlation Model by Contralateral Inhibition I-II., *J. Acoust. Soc. Am.*, Vol. 80, pp.1608–1630., 1986.
- [45] W. Lindemann:
Die Erweiterung eines Kreuzkorrelationsmodells der binauralen Signalverarbeitung durch kontralaterale Inhibitionmechanismen. Dissertation, Ruhr-Universität, Bochum, 1985.
- [46] S. Wolf:
Ein probabilistisches Modell zur Simulation binauraler Phänomene. *DAGA 87*, pp.533–536.
- [47] S. Wolf:
Untersuchungen zur Lokalisation von Schallquellen in geschlossenen Räumen. Dissertation, Ruhr-Universität, Bochum, 1991.
- [48] E. Osman:
A correlation model of binaural masking level differences. *J. Acoust. Soc. Am.*, Vol.50, (1971), pp. 1494–1511
- [49] H.S. Colburn, N.I. Durlach:
Models of binaural interaction. *Handbook of Perception*, Ed. by Cartarette/Friedman, Academic New York, Vol. IV., pp.467–518., 1978.
- [50] H. Gaskell:
The precedence effect. *Hear. Res.* 11, pp.277–303., 1983.
- [51] J. Gonzales-Rodriguez, S. Cruz-Llanas, J. Ortega-Garcia:
Concurrent speakers separation through binaural processing of stereo recordings, Proc. Eurospeech'99, Vol. 6., Budapest, 1999. pp.2595–2598.