

# Tartalom

<i>Előszó</i>	2
<b>Teleki Csaba, Vicsi Klára</b> Többnyelvű európai híryanag-adatbázis gyűjtése és feldolgozási módszereinek kutatása multimédiás műsorok automatikus feldolgozásához	3
<b>Bóhm Tamás, Németh Géza</b> Algoritmus formánsok követésére, módosítására és szintézisére	11
<b>Németh Géza</b> Az akusztikai arculat szerepe az infokommunikációs szolgáltatók megítélésében	17
<b>Takács György, Tihanyi Attila, Bárdi Tamás, Feldhoffer Gergely, Srancsik Bálint</b> MPEG-4 modell alkalmazása szájmozgás megjelenítésére	22
<b>Pintér István</b> Beszédjelek pillanatnyi jellemzőinek becslése a Teager-operátorral és a Hilbert-Huang-transzformációval	28
<b>Standeisky István, Dósa György, Balla Éva</b> A kibővített Jászberény Rádióállomás új sugárzási rendszere és az optimális üzemi frekvenciaterv meghatározása	38
<b>Dósa György</b> A hazai rádiózás születése és kezdetei	46
<b>Kardkovács Zsolt, Tikk Domonkos, Bánsághi Zoltán</b> A 2005-ös KDD kupa feladatának megoldása a Fürkész algoritmussal	50
<b>Lajtha György</b> A WTC 2006 konferencia szubjektív értékelése	59
Az üvegszál-as átviteli hálózat változó világunkban (x)	62

*Címlap: A Magyar Tudományos Akadémia épülete*

---

## Védnökök

SALLAI GYULA a HTE elnöke és DETREKŐI ÁKOS az NHIT elnöke

---

## Főszerkesztő

SZABÓ CSABA ATTILA

## Szerkesztőbizottság

Elnök: ZOMBORY LÁSZLÓ

BARTOLITS ISTVÁN  
BÁRSONY ISTVÁN  
BUTTYÁN LEVENTE  
GYŐRI ERZSÉBET

IMRE SÁNDOR  
KÁNTOR CSABA  
LOIS LÁSZLÓ  
NÉMETH GÉZA  
PAKSY GÉZA

PRAZSÁK GERGŐ  
TÉTÉNYI ISTVÁN  
VESZELY GYULA  
VONDERVISZT LAJOS

# Előszó

szabo@hit.bme.hu

Jelen számunk első részében a beszédtechnológia további eredményeit szeretnénk bemutatni. Emlékeztetjük az Olvasót, hogy ez év márciusi célszámunkat teljes egészében ennek a témakörnek szenteltük, s a téma jelentőségére és a hazai kutatás-fejlesztési aktivitásra jellemző, hogy jóval több cikk érkezett be, mint amennyi az akkori célszámunkban elfért, így a második csokrot ebben a számunkban adjuk közre.

Elsőként Teleki Csaba és Vicsi Klára cikkét mutatjuk be, amely a többnyelvű híryanag-adatbázisok gyűjtése és ezek egységes elvű feldolgozási módszereinek kidolgozása tárgyú nemzetközi projekt keretein belül végzett munkáról számol be. A munka célja egyfelől a magyar híryanag-adatbázis létrehozása és feldolgozása volt, másrészt pedig különböző nyelvfüggetlen kiértékelő eszközök kifejlesztése a beszéd akusztikai paramétereire támaszkodva.

Böhm Tamás és Németh Géza egy olyan, a lineáris predikció módszerén alapuló eljárást mutatnak be, mely alkalmas a formánsmenetek követésére és módosítására beszédjelben, valamint hatékony eszköz lehet a beszéd spektrális tartalmának elemzésére és megváltoztatására, lehetőséget adva a hangminőség és a hangszínezet módosítására

Németh Géza a jól ismert „vállalati arculat” kifejezésnek az akusztikai jelenségekkel kapcsolatos kiterjesztésére tesz kísérletet az „akusztikai arculat” bevezetésével, amelynek tudatos kialakítása a külvilág felé a szervezetek részéről a hangjelenségekkel történő megjelenés jellemzőinek és minőségének rendszerezett tervezését és megvalósítását igényli.

Takács György és szerzőtársai az MPEG-4 szabványnak a fej és az emberi test alakjának és mozgásával foglalkozó részének kódolási alapelveit mutatják be és annak egy speciális alkalmazását. Ebben beszédjelből előállított jellemzők vezérlik a fejmodellt, amely siketek számára szolgál kommunikációs segédeszközként.

A beszédtechnológiai cikksokrot egy elméleti munka zárja: Pintér István cikke beszédjelek nemlineáris és nemstacionárius jellemzőinek meghatározásával foglalkozik és a Teager-operátort és a Hilbert-Huang-transzformációt (HHT) ismerteti, mint a pillanatnyi amplitúdó és a pillanatnyi frekvencia becslésére alkalmazható jel-feldolgozási eljárást.

Ezt követően két, rádió-műsorszórással foglalkozó cikket ajánlunk olvasóink figyelmébe.

Standeisky István, Dósa György és Balla Éva cikkében a 2004-2005 folyamán kibővített Jászberény Rádióállomás történetével ismertetnek meg bennünket, majd ezt követően a cikk bemutatja az optimális üzemi frekvencia meghatározásának folyamatát különböző célterületekre, három különböző módszerrel.

Dósa György történeti írása a magyar rádiózás születésével foglalkozik, az érdekes mondanivalót korabeli fényképekkel illusztrálva. E cikknek az is aktualitást ad, hogy lapszámunk megjelenését követően, szeptember 7-én délután kerül sor a HTE rendezésében „A hazai rádiózás születésének 100. évfordulója” című tudományos emlékülésre a Magyar Telekom Rt. székházában.

E számunkban adtunk helyet egy érdekes, már régóta közlésre váró cikknek, amely egy rangos verseny feladatának díjnyertes megoldását mutatja be, amelyvel az ACM (Association of Computing Machinery) KDD (Knowledge Discovery and Data Mining) kupáján nyertek díjat a szerzők: Kardkovács Zsolt, Tikk Domonkos és Bánsághi Zoltán.

Végül a nemrég a HTE társrendezésében sikerrel lezajlott budapesti WTC-konferenciáról olvashatunk beszámolót Lajtha György tollából.

*Szabó Csaba Attila,  
főszerkesztő*

# Többnyelvű európai híryanag-adatbázis gyűjtése és feldolgozási módszereinek kutatása multimédiás műsorok automatikus feldolgozásához

TELEKI CSABA, VICSI KLÁRA

BME Távközlési és Médiainformatica Tanszék, Beszédakusztikai Kutatólaboratórium  
{vicsi, teleki}@tmit.bme.hu

Lektorált

**Kulcsszavak:** digitális jelfeldolgozás, beszédfeldolgozás, beszédatadabázisok

Többnyelvű híryanag-adatbázisok (Broadcast News – BN) gyűjtése és ezek egységes elvű feldolgozási módszereinek kidolgozására nemzetközi munkacsoport jött létre a COST278 EU projekt keretein belül. A BME TMIT Beszédakusztikai Kutatólaboratóriuma a csoport tagjaként magyar híryanag-adatbázist hozott létre, amely 3 óra és 30 percnyi kép- és hanganyagot tartalmaz. Az adatbázis feldolgozásához a BN munkacsoport által kidolgozott módszereket és előírásokat használta fel, ilyen például az átíró és annotáló szoftver, amely a NIST (National Institute of Standards and Technology) ajánlásai alapján készült. Az átíratok egységes formátumra hozása érdekében, a NIST ajánlásai mellett, pontos címkézési módszereket, szabályokat hoztunk létre. Kutatócsoportunk másik célkitűzése az volt, hogy a beszéd akusztikai paramétereire támaszkodva különböző nyelvfüggetlen, kiértékelő eszközöket fejlesszen ki (beszéddetektálás, beszélő nemének meghatározása stb.). E tanulmányban laboratóriumunk magyar nyelvre vonatkozó feldolgozási módszereit mutatjuk be, valamint tesztelési eredményeinket hasonlítjuk össze a munkacsoport tagjai által elért eredményekkel.

## 1. Bevezetés

Köztudott, hogy napjainkban a média egyre nagyobb teret hódít és talán nem is tudatosul bennünk, fogyasztókban, hogy a rádiós, televíziós műsorok, híradások egy akusztikus számára is új kutatási területek szinte kiaknázhatatlannak tűnő tárházát jelenti. Egyre több kereskedelmi csatornát hallgathatunk a rádióban, nézhetünk a televízióban, és ezek mindegyike megpróbálja egyéni arculatát megteremteni, amely a képi ábrázolás mellett egyfajta „akusztikai arculatot” is definiál. A televíziós híradások világának akusztikai vizsgálatát céloztuk meg, támaszkodva a képi anyagra is, bár kisebb mértékben.

A COST278 BN munkacsoportban létrehozott európai többnyelvű (flamand, portugál, gall, cseh, szlovén, szlovák, görög, horvát és magyar) híryanag-adatbázis alapul fog szolgálni a multimédiás műsorok automatikus feldolgozásához, például a híryanag automatikus lejegyzéséhez, reklámsugárzás számlálásához, stb. A beszédtechnológiai kutatások egyik kiemelkedő területe a híryanagok automatikus lejegyzése. Ez a lejegyzés annál pontosabb, minél nagyobb és minél jobban feldolgozott adatbázissal történik a használt beszédfelismerő rendszer betanítása, tesztelése [1]. Éppen ezért igen nagy fontossága van annak, hogy milyen módszerrel történik az adatbázis feldolgozása. A munkacsoport célul tűzte ki, hogy az Amerikai Egyesült Államokban folyó kutatásokat is figyelembe véve (HUB4 amerikai híryanag korpusz [2]) egy európai feldolgozási és értékelési módszert dolgozzon ki [3].

A kifejlesztett algoritmusokat a többnyelvű BN adatbázisokon teszteltük, kiértékelve a kidolgozott algoritmusok hibáit és előnyeit, így érve el egyre jobb feldolgozási eljárásokat, amelyeket a munkacsoport minden

tagja használ. Például az akusztikai és nyelvi feldolgozás során, a NIST ajánlásai alapján, elkészült egy újfajta annotáló eljárás [4], amelyet a munkacsoport minden tagja használ.

Az eredmények kiértékeléséhez két különböző szoftver került kifejlesztésre. A portugál partner által biztosított kiértékelő szoftvert [7] használtuk a beszéd-nem-beszéd detektáló algoritmusok eredményeinek kiértékelésekor, míg a beszélő csoportosító, a beszélő nemét detektáló szoftverek eredményeinek kiértékeléséhez a belga partner által közreadott szoftvert [8] használtuk.

E tanulmányban a magyar nyelvű híryanag-adatbázis gyűjtéséről, az akusztikai és nyelvi feldolgozásáról, valamint a szegmentálásról és kiértékelésről számolunk be, továbbá bemutatjuk az általunk kifejlesztett új és sikeres beszéd-nem beszéd detektáló eljárást.

## 2. Adatgyűjtés

Laboratóriumunkban korábban több adatbázist is gyűjtöttünk, melyeknek felhasználási célja különböző volt. Készült fonetikai kutatások céljából egy adatbázis (BA-BEL) [12], melynek szöveganyaga süketszobában került rögzítésre, így a felvételek során a jel-zaj viszony magas volt. A bemondott szöveganyag precízen összeszeválogatott mondatokból állt, mivel cél volt, hogy az adatbázis tartalmazza legalább kétszer a magyar nyelvben előforduló félszótagok 98 százalékát. Tartalmaz továbbá számokat és CVC (mássalhangzó-mághangzó-mássalhangzó) kapcsolatokat is. A bemondók száma kicsi (60 beszélő) és nagy hangsúlyt fektetünk arra, hogy a beszélők szépen, artikuláltan beszéljenek.

Egy másik jellegű adatbázis a Magyar Telefonos Beszéd-Adatbázis (MTBA) [13]. Az adatbázis 500 magyar nyelvű beszélő hanganyagát tartalmazza, ezekből 297 vezetékes, 203 pedig mobil telefon-felvétel. Az adatbázis általános fonetikai, nyelvészeti kutatásokhoz szolgál alapul és statisztikai feldolgozási módszereken alapuló személyfüggetlen gépi beszédfelismerők, dialógusrendszerek létrehozását teszi lehetővé.

A Magyar Referencia Beszéd-Adatbázis esetén a cél egy olyan, olvasott folyamatos szöveget tartalmazó beszédadatbázis létrehozása volt, amely alkalmas PC-s beszédfelismerők betanítására, tesztelésére [13]. Az adatbázis szöveganyagát úgy terveztük meg, hogy az adatbázisba bekerülő mondatokban a felismerő rendszerekben tipikus felismerési egységek (beszédhangok, difón, trifón egységek) elegendően sokszor forduljanak elő. A mondatok mellett fonetikailag gazdag szavakat is kiválasztottunk, a nem kellő számban előforduló beszédhangok példányszámának növelése érdekében. Így a 332 adatközlő fejenként 12 különböző mondatot és 12 különböző, a mondatoktól független szót olvas fel. Az adatbázis felvételeit irodai helyiségekben, laborokban, otthonokban rögzítettük.

Látható, hogy a fentebb említett adatbázisok esetén bizonyos szempontok alapján megtervezett szöveget mondott be a beszélő egy (vagy több) számítógéphez csatlakoztatott mikrofonba, vagy egy telefon mikrofonjába. A COST278 munkacsoport által létrehozott adatbázist a fentebb említett adatbázisokhoz képest egy teljesen más szemlélet jellemez, hiszen a híryanag-adatbázis egy többértű, multimédiás adatbázis, melynek feldolgozása során a felhasználónak alkalma nyílik egy akusztikai szempontból is sokkal gazdagabb anyagba betekinteni. A kutatócsoport minden tagintézménye egy legalább három órás adatbázist gyűjtött nemzeti (közszolgálati/kereskedelmi) televíziók hírműsoraiból. Jelenleg a teljes BN adatbázis 30 órányi anyagot tartalmaz, melyet 10 különböző televízióállomástól rögzítettünk 9 európai nyelven: flamand, portugál, gall, cseh, szlovén, szlovák, görög, horvát és magyar. A magyar nyelvű BN adatbázis körülbelül 3 óra 30 percnyi híryanagot tartalmaz, melyet közszolgálati és kereskedelmi adók műsoraiból rögzítettünk az 1. táblázatban bemutatott arányban.

1. táblázat A magyar nyelvű BN adatbázis struktúrája

TV állomás	Időtartam (perc:mp)	Híradások száma	Műsorvezetők száma
MTV1	77:09	3	4
RTL Klub (k)	40:51	5	2
TV2 (k)	84:40	3	4

A táblázatban a (k) jelzéssel ellátott televízióállomások kereskedelmi adók. Szerepel továbbá a felvételek hossza televízió-állomásonként, a híradások száma összesen, illetve az adatbázisban rögzítésre került hírműsorok különböző műsorvezetőinek száma.

Látható, hogy a magyar nyelvű BN adatbázis összetett, hiszen az adatbázis három különböző televíziócsatorna különböző típusú hírműsorainak anyagát tartalmazza. Többnyire kereskedelmi adók hírműsorai kerültek be az adatbázisba (2 ó, 5 p és 31 mp, ami az adatbázis kétharmadát teszi ki), a fennmaradó egyharmad tartalmazza a közszolgálati televízió híradóit. Nyilván az, hogy a magyar nyelvű BN adatbázis igen heterogén, hatással lesz a kutatásaink eredményeire.

A felvételek egy személyi számítógéphez csatlakoztatott televíziós készülék segítségével készültek. A számítógép egy speciális jelfeldolgozó kártyával volt felszerelve, így lehetőség nyílt arra, hogy ne csak a hanganyagot, hanem a képi anyagot is rögzíteni tudjuk. A hanganyag digitalizálásakor 16 kHz-es mintavételi frekvenciát használtunk, tároláshoz pedig a hanghullám (wave) formátumot használtuk a következő paraméterekkel: 16 kHz mintavételi frekvencia 16 biten ábrázolva PCM kódolással, 256 kbit/s-os bitsebességgel. A képi anyag tárolásakor két szempontot vettünk figyelembe: egyrészt, hogy megfeleljen a COST278 BN kutatócsoport ajánlásainak, másrészt, hogy a képi anyag valószínűleg segítséget tudjon nyújtani az átírás során. Ezért két különböző tömörítési eljárást használtunk. Az egyik, a COST278 BN kutatócsoport ajánlásainak megfelelően a következő volt: Indeo® video 5.11 verziójú kodek 930 kbit/s-os bitsebességgel (a kép mérete: 180x144 pixel, 25 képkocka másodpercenként). A kép mérete miatt, az ily módon tárolt felvételek nem voltak igazán használhatóak az átírás, címkézés során, ezért belső használatra elkészítettünk egy 360x288 pixel méretű képi anyagot, melyet DivX 5.0.5 verziójú kodekkel tömörítettünk, 998 kbit/s bitsebességgel.

A magyar BN adatbázis, mely tartalmazza a hanganyagot, a képi anyagot és az átíratokat is, CD lemezen és egy belső használatra létrehozott szerveren tároljuk, melyhez a kutatócsoport minden tagja hozzáfér. Minden adatfájl egyedi névvel rendelkezik, melyből kiderülnek a felvételre jellemző legfontosabb paraméterek, a következőképpen:

```
<tv_csatorna_név>_<év>_<hhnn>_<óópp>
      .<kiterjesztés> ,
```

ahol a <tv\_csatorna\_név> annak televíziós csatornának a neve, ahonnan a híradó rögzítésre került, az <év>\_<hhnn>\_<óópp> paraméterek a rögzítés dátumát és pontos kezdési idejét tartalmazza. A fájlok kiterjesztése pedig lehet wav, amennyiben hangfájlról van szó, avi, amennyiben a képi anyagról van szó és trs, seg vagy stm, amennyiben adatfájlokról van szó (átírat, címkézés).

### 3. Az adatbázis akusztikai és nyelvi feldolgozása

Az adatbázis akusztikai és nyelvi feldolgozása során nagyon fontos a hanganyag átírása, címkézése. A címkézés során a LDC (Linguistic Data Consortium) ide vo-

natkozó ajánlásait követték. Mivel a LDC ajánlásai nem voltak eléggé konkrétak és sok hiba forrásául szolgálhattak, kiegészítettük ezeket a BN kutatócsoport által ajánlott szabályokkal [3]. Ezáltal nagyobb lett az esély arra, hogy a BN kutatócsoport résztvevői megfelelően pontos és hasonló módszerekkel készítsék el a címkefájlokat, elősegítve ezzel a közös munkát.

### 3.1. Címkézési szabályok

A címkézés során jelöltük a beszélőváltások során fellépő akusztikai változásokat, a beszélő által elmondott szöveg határait, a híradások szekcióit, a híradások szignáljainak kezdetét és végét, idegen nyelvű beszédet, háttérzajt és a beszélő által keltett zajokat.

A beszélőváltások során fellépő akusztikai változásokat az átviteli csatorna milyensége és annak minősége határozta meg. Két fajta átviteli csatornát különböztettünk meg (stúdióban elhangzott beszéd vagy telefonon keresztül elhangzott beszéd) és mindegyik csatornát minősítettük azok akusztikai minősége szerint (jel-zaj viszony alacsony, közepes vagy magas). Jellemzően a stúdióban elhangzott beszélgetéseket, a stúdióból kommentált riportokat, illetve a műsorvezető beszéde során elhangzó hanganyagot a „stúdió”, „magas” címkéssel láttuk el. A „stúdió, közepes” (jel-zaj viszony) címkével akkor jelöltük a beszédet, ha a riporter stúdióon kívül beszél, jellemzően ezt a címkét az utcán, vagy nyílt terepen elhangzott beszéd kapta. A „stúdió, alacsony” címkével a különösen zajos környezetben készített felvételeket láttuk el. A telefonos beszéd esetén a tiszta beszédet a „magas” címkével, a zajos, de még érthető beszédet a „közepes” címkével, míg a nehezen érthető beszédet az „alacsony” címkével láttuk el. Ezt a kódolási eljárást a 2. táblázatban foglaltuk össze.

Egyik fontos címkézési szabály az, hogy az egy bementő által bementő beszédet több, kisebb egységre bontottuk, ezért a magyar nyelvű BN adatbázisban minden belélegzés egy ilyen egység kezdete is egyben. Amennyiben a beszélőváltáskor keletkezett beszédszünet kisebb 0,5 másodpercnél, nem jelöltük. Amennyiben ez a fajta szünet 0,5 másodpercnél nagyobb, de 1,5 másodpercnél kisebb, akkor ezt jelölni kellett egy címkével a szünet közepén. Amennyiben 1,5 másodpercnél nagyobb szünet keletkezik, akkor a szünet elejét is és a végét is jelöltük.

1. ábra

A Transcriber program kezelői felülete

		Csatorna	
		Stúdió (Sávszélesség > 4 kHz)	Telefon (4 kHz)
Minőség	Alacsony	Zajos	Érthetetlen
	Közepes	Stúdióon kívül	Zajos
	Magas	Stúdióban	Tiszta

2. táblázat A csatorna fajtája és minőségének jelölése

Adatbázisunkban a szekciók határait is jelöltük. Szekció lehet egy riport (hírtértékű esemény prezentációja), kitöltő szövegek (rövidhírek, címszavak stb.) és nem átírt események (reklámok és szignálok).

Minden szignál külön címkét kapott. Mivel előfordulhat, hogy a híradó elején, végén és közben mindig különböző szignálokat hallunk, az adás eleji és az adás végi szignált egy külön címkével jelöltük. Az idegen nyelvű szöveg kezdetét címke jelzi, de nem lett átírva.

A címkefájl tartalmazza továbbá a háttérzajok kezdetét és végét jelölő címkéket is. A háttérzajokat különböző kategóriákba osztottuk: zene, beszéd, susogás, egyéb. Ugyanakkor jelöltük a beszélő által keltett zajokat is, mint a belélegzés, kilélegzés, papírzörgés stb. Az átiratok minden esetben a Transcriber [4] nevű program segítségével készültek (<http://www.etca.fr/CTA/gip/Projets/Transcriber>) és XML formátumú, ISO-8859-2 karakterkódolású szövegfájlban kerültek elmentésre. Az 1. ábrán látható a Transcriber program kezelői felülete.

A kezelői felület közepén látható a felvétel idő-ampplitúdó függvénye, ez képezi a választóvonalat a kezelőfelület alsó és felső része között. A kezelőfelület felső részéhez fér hozzá a felhasználó, ide írhatja le az elhangzottakat, jelölheti be a szekciók elejét, a beszélő-



váltásokat (természetesen ilyenkor az akusztikai paraméterek változását is jelölni kell, amennyiben ez megtörténik), a beszélő által keltett zajokat (belégzés *[ij]*, kilégzés *[e]*, papírcsörgés *[pap]*, stb.), a háttérzajokat stb.

Amennyiben a felhasználó mindezeket helyesen jelöli, a kezelőfelület alsó részén megjelenik hierarchikus formában a bejelölt információ (fentről lefelé haladva): a háttérzajok sávja szürke, amennyiben létezik bejelölt háttérzaj (zene, susogás stb.), alatta jelenik meg a szekció sávja, amelyben a szekció fajtája van bejelölve (riport – „report”, kitöltő szövegek – „filler”, vagy nem átírt szöveg – „nontrans”). Amennyiben a szekció a „report” vagy „filler” címkét kapta, a felhasználó egy néhány címszavas leírást is írhat a szekció tartalmáról. A szekció sávja alatt található a beszélő sávja. Ebbe a

sávba kerülnek az egy szekción belül előforduló bemondók nevei, vagy ennek hiányában valamilyen egyedi azonosító. Minden beszélő esetén egy adatlapot kell kitölteni a beszélőre jellemző adatokkal (a bemondó neve, annak neme, műsorvezető-e vagy sem, akusztikai környezet stb.)

Mivel nem mindig hangzik el a bemondó neve, nagy segítséget tud nyújtani ezen úrlap kitöltésekor a képi anyag. Amennyiben nincs beszéd, a „no speaker” címke kerül ebbe a sávba. A beszélő sávja alatt található a bemondott szöveg átiratának sávja, alatta pedig az időcímkék.

Említettük, hogy kimenetként ez a program egy XML kódolású, igen nehezen kezelhető, szövegfájlt produkál. Az alábbiakban egy részletet tekinthetünk meg belőle:

```
<?xml version="1.0" encoding="ISO-8859-2"?><!DOCTYPE Trans SYSTEM "trans-13.dtd">
<Trans scribe="(unknown)" audio_filename="MTV1_2~1" version="17" version_date="040509">
<Topics>
<!-- a híradóban előforduló témakörök felsorolása-->
  <Topic id="to1" desc="Takarékossági csomag - egészségügy"/>
  [...]
  <Topic id="to12" desc="elköszönés"/>
</Topics>
<Speakers>
<!-- a híradóban előforduló beszélők felsorolása és azok paramétereit -->
  <Speaker id="spk1" name="Hajdú Andrea" check="yes" type="female" dialect="native"
  accent="" scope="local"/>
  [...]
  <Speaker id="spk6" name="reporternol" check="no" type="female" dialect="native"
  accent="" scope="global"/>
</Speakers>
<Episode>
<Section type="report" startTime="0" endTime="122.063" topic="to1">
<!-- a szekció kezdete, időcímkék -->
  <Turn startTime="0" endTime="16.431">
    <!-- beszélő váltás kezdete, időcímkék, akusztikai paraméterek-->
    <Sync time="0"/>
    <Event desc="jingle" type="noise" extent="instantaneous"/>
  </Turn>
  <Turn speaker="spk1" mode="planned" fidelity="high" channel="studio"
  startTime="16.431" endTime="42.03">
    <Sync time="16.431"/>
    <Event desc="i" type="noise" extent="instantaneous"/>
    <!-- zajesemény (belégzés) leírása -->
    Jó reggelt kívánok <!-- az elhangzott szöveg -->
    <Event desc="e" type="noise" extent="instantaneous"/>
    !
    <Sync time="18.686"/>
  </Turn></Section>
  [...]
</Episode></xml>
```

Látható, hogy a Transcriber program által kimenetként előállított fájl nehézkesen olvasható, nehézkesen dolgozható fel, ezért ezt a fájlformátumot egy könnyebben kezelhető fájlformátumra konvertáltuk. Erről bővebben a következő szakasz ad tájékoztatást.

### 3.2. Adatbázis értékelés (statisztikák)

A magyar BN adatbázisban 2425 mondatot különböztettünk meg, amelyek közül 2382 mondat került átírásra.

Összesen 22.500 szó szerepel az adatbázisban, melyek közül a különböző szavak száma 8147. Összevetve a COST278 BN kutatócsoport tagjainak adatbázisaival, azt tapasztaltuk, hogy a magyar adatbázis a cseh és a szlovák adatbázisokkal hasonlítható össze a fenti számok alapján. A cseh adatbázisban előforduló szavak száma 27.642, míg a különböző szavak száma 8834 (a cseh adatbázis 181 percnyi híryanagot tartalmaz). A szlovák adatbázisban 25.770 szó található, a különböző szavak száma 8887 (a szlovák adatbázis 191 percnyi híryanagot tartalmaz) [3].

Mint azt láthattuk, az átírat során keletkezett fájlformátum nem mondható ideálisnak automatikus gépi feldolgozáshoz. Ezért ezt egy olyan formátumra konvertáltuk, amelyben soronként a következő információkat rögzítettük:

```
[fájlnev] 1 [bemondó neve] [időcimke1]
[időcimke2] <o,[F állapot],
[beszélő neve]> [bemondott szöveg]
```

Például:

```
MTV1_2004_0220_1200 1 Rábai_Balázs 395.151 408.813
<o,F0,male> [i] Bizonytalanná vált a ^szegedi légi-
mentők működése. [i] A szolgálatot fenntartó alapít-
vány kormányzati támogatása [e] több, mint harminc
százalékkal, huszonnolc millió forinttal [pap] csök-
kent a tavalyihoz képest.
```

Látható, hogy a bemondott szövegben már bejelöltük a beszélő által keltett zajokat is. Az időcímkék határozzák meg a bemondás kezdetét és végét ms-ban. Ebben a formátumban a csatorna minőségét és milyenségét is átkódoltuk a könnyebb kezelhetőség és a pontosabb leírás érdekében (*F-állapotok*–*F-conditions*, további információk: [http://www ldc.upenn.edu/Projects/Corpus\\_Cookbook/transcription/broadcast\\_speech/english/conventions.html](http://www ldc.upenn.edu/Projects/Corpus_Cookbook/transcription/broadcast_speech/english/conventions.html)).

A 3. táblázatban bemutatjuk az F-állapotok szerinti statisztikát a magyar nyelvű BN adatbázis esetén. A táblázat világosan mutatja, hogy a telefonon keresztül interjúkészítés igen kedvelt módszer a magyar médiában, hiszen közel az adatbázis közel 18%-a telefonon

<i>F-állapot</i>	<i>Időtartam (óó:pp:ss)</i>	<i>Százalék</i>
Szélessávú tiszta (olvasott) beszéd – F0	00:22:07	11.43 %
Szélessávú tiszta (spontán) beszéd – F1	00:17:24	9.1 %
Telefonos beszéd – F2	00:34:40	17.72 %
Beszéd háttérzene jelenlétében – F3	00:13:15	6.95 %
Rossz akusztikai viszonyok között elhangzott beszéd – F4	01:31:03	45.43 %
Nem anyanyelvű beszélő által mondott szöveg – F5	00:01:16	0.83 %
Egyéb beszéd – FX	00:16:27	8.54 %

3. táblázat

*F-állapotok a magyar nyelvű BN adatbázisban*

keresztül bemondott szöveget tartalmaz. A COST278 BN adatbázisainak statisztikái szerint a telefonos interjúkészítés hungaricum, hiszen partnereink adatbázisaiban elenyésző mértékben volt jelen az ilyen körülmények között rögzített beszéd (kevesebb, mint 4% felelt meg az F2 állapotnak).

A televíziós híradás egy másik jellemzője az, hogy a riportokat a hírértékű esemény megtörténének helyszínén készítik részben, vagy akár teljes egészében. Ez az adatbázisunk statisztikájában a rossz akusztikai viszonyok között elhangzott, azaz F4 állapotú felvételek formájában jelennek meg, amelyek a teljes adatbázis csaknem felét teszik ki. Ugyanakkor egy másik magyarizat erre a tényre az lehet, hogy a magyar BN adatbázis kétharmada kereskedelmi adók hírműsoraikat tartalmazza. Azért lehet ez is egy magyarizat, hiszen közkedvelt a kereskedelmi adók híradóiban a viszonylag hangos háttérzene alkalmazása.

Ez a 45%-os arány átlagosnak mondható, hiszen a többi BN adatbázis statisztikája is az F4 állapotra ezt a százalékos arányt prezentálja többé-kevésbé. Kivételként megemlíthető a két szélsőértéket produkáló BN adatbázis: a portugál nyelvű BN adatbázis, amelynek nagy része F4 állapotú beszédet tartalmaz (76.4%) és a szlovén BN adatbázis, amelynek igen kis részét teszi ki az F4 állapotú beszéd (8.1%) [3]. Messzemenő következtetéseket nyilván nem tudunk az előbb említett számok alapján levonni, de valószínűsíthető, hogy a portugál adatbázisban nagyobb arányban voltak jelen a kereskedelmi adóktól átvett híryanag a közszolgálati adóhoz képest, míg a szlovén adatbázis esetében ez az arány fordítva volt jelen.

A 4. táblázatban látható, hogy a BN adatbázisokban milyen arányban jelentek meg férfi, illetve női beszélők.

4. táblázat

*A beszélők eloszlása a BN adatbázisokban nemek szerint*

	<i>Női beszélők száma</i>	<i>Női beszéd időtartama</i>	<i>Férfi beszélők száma</i>	<i>Férfi beszéd időtartama</i>
<i>Belga BN</i>	27	0:39:48	88	1:59:46
<i>Cseh BN</i>	128	1:05:51	285	1:45:06
<i>Gall BN</i>	34	1:39:47	100	1:20:30
<i>Görög BN</i>	42	0:42:01	111	1:32:04
<i>Horvát BN</i>	69	1:05:58	140	1:47:03
<i>Magyar BN</i>	39	1:15:28	102	1:35:04
<i>Portugál BN</i>	39	0:44:53	133	2:31:38
<i>Szlovén BN</i>	55	1:02:27	139	1:30:58
<i>Szlovén 2 BN</i>	28	0:42:01	69	1:24:14
<i>Szlovák BN</i>	29	1:08:39	95	1:51:22
<b>összesen:</b>	<b>490</b>	<b>10:06:53</b>	<b>1262</b>	<b>17:17:45</b>

A táblázatban szereplő beszélők száma az összes olyan beszélőt takarja, akinek a hangja elhangzott a híradás során. Látható, hogy jóval nagyobb számban szerepelnek a híradásokban a férfi beszélők a női beszélőkhöz képest. Valószínűsíthető, hogy a televíziós társaságok a nagyobb hitelesség reményében inkább férfiakat bíznak meg a műsorvezetéssel, riportkészítéssel stb. Ugyanakkor látható az is, hogy annak ellenére, hogy általában jóval kevesebb a női beszélő a híradásokban, mégis az egy főre jutó beszélt percek száma a nőknél nagyobb, mint a férfiaknál. Egy női beszélő átlagosan 1 percet és 13 másodpercet beszél, míg egy férfi beszélőre jutó beszéidő 49 másodperc. Természetesen ez a szám adatbázisonként változik, például a gall BN adatbázis esetén az egy női beszélőre eső percek száma majdnem 3, míg a férfi beszélők csupán 1 percet és 24 másodpercet beszéltek, azaz feleannyit. A legkiegyenlítettebb arány talán a portugál adatbázisban fedezhető fel, ahol egy női beszélő 69 másodperc beszéidővel, míg egy férfi 68,57 másodperc beszéidővel rendelkezik.

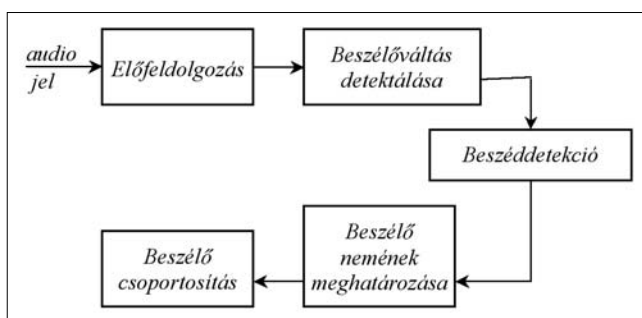
Mindezek a különbségek, eltérések az adatbázisok között, sőt még a magyar adatbázison belüli különbségek is az adatbázis többretűségét, újszerűségét emelik ki.

#### 4. Szegmentálás, kiértékelés

A COST278 BN kutatócsoportjának célkitűzései között szerepelt az is, hogy a már rögzített és szabályosan átírt, címkézett adatbázist alapul véve olyan intelligens automata rendszereket fejlesszen ki, amelyek képesek néhány nyelvfüggetlen paraméter alapján feldolgozni a teljes BN adatbázist, majd ezeket a rendszereket egy egységes kiértékelő elv alapján osztályozni. A kutatócsoport a következő algoritmusok megvalósítását tűzte ki célul: beszélőváltás detektálása, beszéd-detekció, beszélő nemének meghatározása, beszélők csoportosítása [3]. A kutatócsoport mindegyik tagja kifejlesztett egy vagy több olyan módszert, amely valamely, a 2. ábrában is jelölt feladat elvégzésére alkalmas volt.

2. ábra

Az audio jel feldolgozásának egyszerűsített blokkvázlata



\* A BME Beszédakusztikai Laboratóriumban kifejlesztett folyamatos beszédfelismerő (MKBF 1.0) optimális működését az akusztikai, fonetikai [10] és nyelvi modellek változtatásával állítottuk be. A felvételek mindegyike – mind a betanításnál, mind a tesztelésnél – 16 kHz-en mintavételezett, 16 biten lineárisan kvantált jel, amely a megfelelő előfeldolgozás után kerül felismerésre. A fonémaszintű felismerőnk 16 kHz mintavételezésű, 17 Bark frekvenciatérbeli derivált, + 17 időbeni derivált, + 17 időbeni második derivált, + energia bemeneti jelvektor mellett, 4-5 állapotú kvázi-folytonos, 24 lépcsős, rejtett Markov-modellekkel (QCHMM), fonéma alappal dolgozik. Az akusztikai, fonetikai szint optimalizálásáról további információk [11]-ben találhatóak.

A Beszédakusztikai Kutatólaboratórium egyik fő célkitűzése a magyar nyelvű hírszóanyag-adatbázis létrehozása és a fentebb említett algoritmusok közül a beszéd-detektáló algoritmus implementálása volt.

A különböző algoritmusok különböző jellegű eredményeket produkálnak, ezért szükség volt egy közös kiértékelő szoftverre, mellyel az eredmények összehasonlíthatóságát biztosították. Ezt a szoftveres eszközt a BN kutatócsoport minden tagja használta és a portugál partner bocsátotta közre.

#### 5. Beszéddetekció

A beszéd-detektáló algoritmus (*speech-non-speech classification – SNC*) feladata az, hogy detektálja a legalább 1,5 másodperces beszéd-szünetet [3], tehát a rögzített anyag olyan részeit, ahol nincs beszéd, de előfordulhat háttérzaj vagy egyéb zaj, zene vagy egyéb hanghatás.

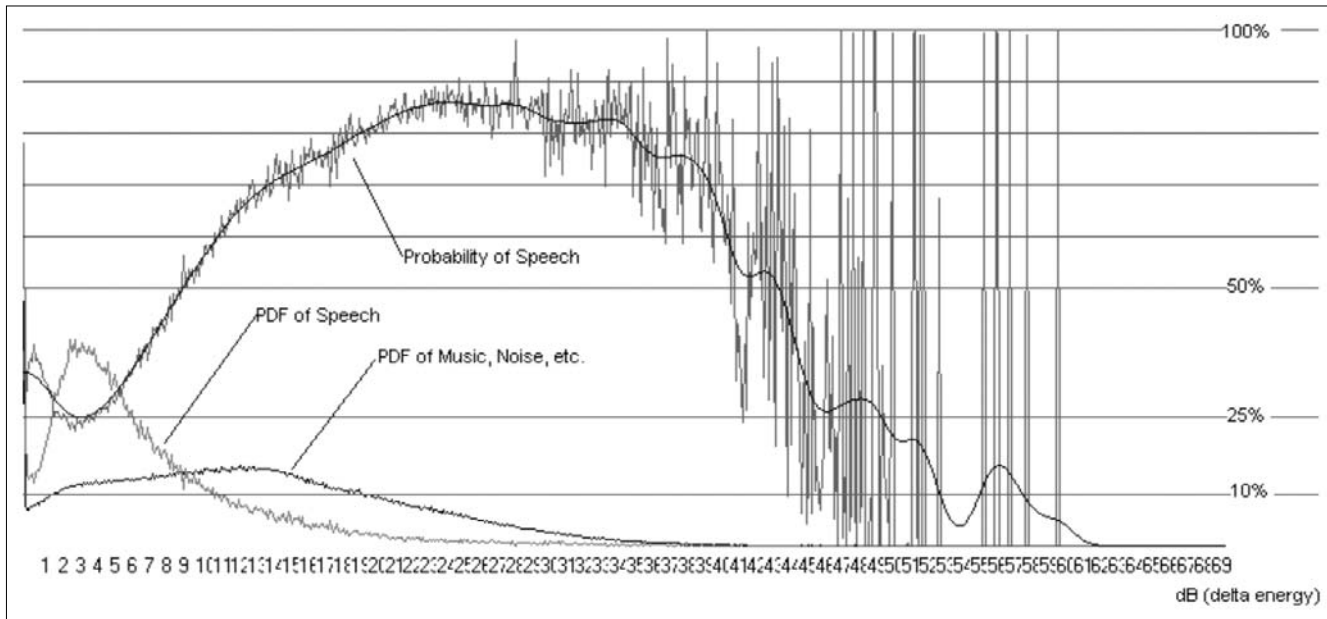
A laboratóriumunkban kifejlesztett algoritmus két különböző, ideiglenes döntésre alapozva hozza meg a végső döntést arról az akusztikai szegmensről, amit a bemeneten talál. Az első ideiglenes döntést egy statisztikai alapú (HMM) fonémafelismerő\* kimenete alapján hozza. Kimenetnek a fonéma-bigram valószínűségeket tekintjük. Ezek a valószínűségek jellemzően más-más értékeket vesznek fel attól függően, hogy a bemeneten beszéd vagy egyéb akusztikai jelenség (zene, zaj stb.) található. A felismerő betanításához bármely, a BN kutatócsoportjában szereplő, nyelven elhangzott hanganyagot használhatunk. A betanítás során két különböző fonéma-bigram mátrix készül, egy a beszédre, egy pedig a „nem-beszédre”, azaz zene, zaj stb. A döntés a felismerés során születik meg a modell tranziensei és a beszéd vagy „nem-beszéd” fonéma-bigram mátrixok közötti távolság függvényében.

A második ideiglenes döntés a beszéd, illetve a „nem-beszéd” szegmens energiaváltozásának statisztikai analízisének eredménye alapján jön létre. A mért energiaváltozást valószínűségi változóként kezelve, egy valószínűség-sűrűség függvényt (*probability density function – PDF*) számolunk a beszédre és a „nem-beszédre” is. A PDF által adott eredmények alapján a beszéd valószínűségét határozzuk meg (3. ábra). Csak abban az esetben fog a rendszer beszédet detektálni, amennyiben mindkét algoritmus beszédet detektál. Minden egyéb esetben „nem-beszédet” fog detektálni.

A módszer kiértékelésekor azon szegmensek százalékos aránya dönt, amelyekre helyes döntést hozott a rendszer. Ez a százalékos arány a rendszer pontosságát fejezi ki („*accuracy*”) [3].

Ezt a feladatot a COST278 BN kutatócsoportjának hét tagintézménye végezte el: a Ghent-i Egyetem –





3. ábra

Valószínűség-sűrűség függvények beszéd, „nem-beszéd” (zene, zaj, egyéb) esetre és annak valószínűsége, hogy beszéd hangzott el

Belgium (ELIS), a Ljubljana-i Egyetem – Szlovénia (ULJ), a Maribor-i Egyetem – Szlovénia (UMB), a Liberéc-i Műszaki Egyetem – Cseh Köztársaság (TUL), INESC ID – Portugália (INESC), a Vigo-i Egyetem – Spanyolország (UVIGO), és a Budapesti Műszaki és Gazdaságtudományi Egyetem (BUTE).

A kiértékelés eredményei (4. ábra) világosan mutatják az algoritmus létjogosultságát, hiszen a feladat megoldása során módszerünk, a többi módszer eredményeivel összevetve, kiemelkedő eredményeket prezentál. A probléma megoldása nem triviális, hiszen előfordulhat, hogy az adatbázisba reklám is belekerül, ami tartalmazhat beszédet is (megjegyzés: a reklámok a BN adatbázisokban nem kerültek átírásra). Az általunk bemutatott módszer sikeresen alkalmazható „nem-beszéd” (zene, zaj stb.) szegmensek detektálására, címkésítésére. Az eredményeket a grafikonon prezentáljuk.

Az ábrán látható, hogy a fentebb bemutatott módszer a COST278 munkacsoport többi tagja által kifejlesztett módszerhez viszonyítva a beszédet majdnem 95% arányban osztályozza beszédnek, ami egy közepes eredmény ebben a kontextusban. Ugyanakkor látható az is, hogy a módszer igazi erőssége abban rejlik, hogy a nem-beszéd eseményt csaknem 85% arányban sorolja a nem-beszéd kategóriába, ami ebben a kontextusban egy kiváló eredmény, hiszen csak a portugál partner tudott olyan algoritmust kifejleszteni, amely 75% feletti arányban teszi ugyanezt. Tehát kimondható az, hogy erre a feladatra a legalkalmasabb módszer az általunk bemutatott módszer.

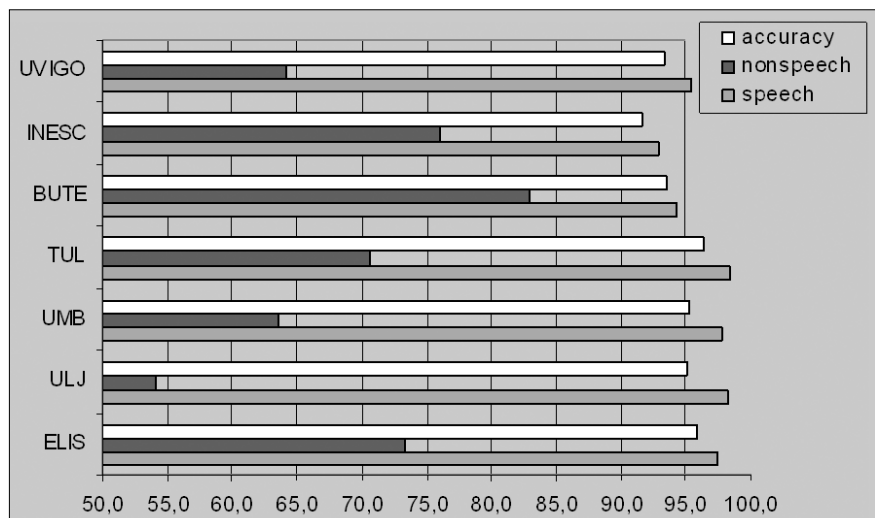
## 6 . Összefoglalás

E tanulmányban a szerzők bemutattak egy újszerű, multimédiás adatbázist, annak struktúráját, az adatbázison végzett statisztikai vizsgálatok eredményeit és egy a „nem-beszéd” detektálásához és annak címkésítéséhez alkalmazható algoritmust, melyet a BME Beszédakusztikai Kutatólaboratóriumában fejlesztettünk ki. Az eredmények alapján kimondható az, hogy ez a módszer megbízhatóan és megfelelően nagy pontossággal ismeri fel a „nem-beszéd szegmenseket”.

Ugyanakkor Laboratóriumunk nem tekinti lezártnak a kutatást ezen a területen, hiszen még nagyon sok kiaknázatlan területe van még. Például, az adatbázis kiválóan alkalmazható automata reklámszámláló szoftverek betanításához, teszteléséhez.

4. ábra

A különböző beszéddetektálási módszerek eredményei



## Irodalom

- [1] Becchetti C., Ricotti L.P.,  
'Speech Recognition, Theory and C++ implementation'  
Fondazione Ugo Bordoni, Rome, (1999)  
ISBN 0-471-97730-6.
- [2] D. Graff, Z. Wu, R. MacIntyre, M. Liberman,  
'The 1996 broadcast news speech and language-model corpus'.  
In: Proceedings of the 1997 DARPA  
Speech Recognition Workshop, February 1997.  
Chantilly, Virginia.
- [3] A. Vandecatseye, J. Martens, J. Neto,  
H. Meinedo, C. Mateo, J. Dieguez, F. Mihelic,  
J. Zibert, J. Nouza, P. David, M. Pleva, A. Cizmar,  
H. Papageorgiou, C. Alexandris,  
'The COST278 –  
pan-European Broadcast News Database',  
In: Proceedings of LREC 04,  
Lisboa, Portugal (2004)
- [4] C. Barras, E. Geoffrois, Z. Wu, M. Libermann,  
'Transcriber : Development and use of a tool  
for assisting speech corpora production',  
Speech Communication,  
Volume 33, Issues 1-2., pp.5–22. (2001)
- [5] J. Zibert, F. Mihelic, J. Martens, J. Neto,  
H. Meinedo, J. Neto, L. Docio, C. Mateo, P. David,  
J. Nouza, M. Pleva, A. Cizmar, A. Zgank, Z. Kacic,  
Cs. Teleki, K. Vicsi,  
'The COST278 Broadcast News segmentation and  
speaker clustering evaluation – overview,  
methodology, systems, results',  
INTERSPEECH 2005,  
Lisboa, Portugal (2005)
- [6] Siegler, M. A., Jain, U., Raj, B., Stern, R. M.,  
'Automatic segmentation, classification and  
clustering of broadcast news',  
In: Proceedings of DARPA Speech Recognition  
Workshop, Chantilly VA, pp.97–99. (1999)
- [7] K. Vicsi, Cs. Teleki, Sz. Velkei,  
'Development and evaluation of  
a Hungarian Broadcast News database',  
In: Proceedings of Forum Acousticum 2005,  
Budapest, Magyarország (2005)
- [8] Perez-Freire, L., Garcia-Mateo C.,  
'A multimedia approach for audio segmentation  
in TV Broadcast News',  
In: Proceedings ICASSP (2004)
- [9] Vandecatseye A., Martens, J.P.,  
'A fast, accurate and stream-based speaker  
segmentation and clustering algorithm'.  
In: Proceedings Eurospeech (2003)
- [10] Deng Y., Mahajan M., Acero A.,  
'Estimating Speech Recognition Error Rate  
without Acoustic Test Data'  
Elérhető: [http://research.microsoft.com/srg/papers/  
2003-milindm-eurospeech.pdf](http://research.microsoft.com/srg/papers/2003-milindm-eurospeech.pdf)
- [11] Sz. Velkei, K. Vicsi,  
'Beszédfelismerő modellépítési kísérletek akusztikai,  
fonetikai szinten, kórházi leletező beszédfelismerő  
kifejlesztése céljából',  
MSZNY 2004, Szeged, Magyarország (2004)
- [12] Roach, P., S. Arnfield, W., Barry, J.,  
Baltova, M., Boldea, A., Fourcin, W., Gonet, R.,  
Gubrynowicz, E., Hallum, L., Lamel, K., Marasek, A.,  
Marchal, E., Meister, E., Vicsi, K.,  
'BABEL:  
An Eastern European Multi-language database'.  
International Conference on Speech and  
Language Processing 1996, Philadelphia.
- [13] Vicsi, K., Valyon, Z., Gordos, G., Csirik, J.,  
Kocsor, A., Tóth, L.,  
'MTBA – Magyar nyelvű telefonbeszéd adatbázis'.  
Technical report. IKTA 3 project,  
a.sz.: 11025888, (2000)  
<http://alpha.ttt.bme.hu/speech/hdbMTBA.php>  
„György Békésy” Acoustics Research Laboratory  
of the Budapest University of Technology and  
Economics (2002).
- [14] Vicsi Klára, Kocsor András,  
Teleki Csaba, Tóth László,  
'Beszédatbázis irodai számítógép-felhasználói  
környezetben',  
II. Magyar Számítógépes Nyelvészet Konferencia,  
(2004)

# Algoritmus formánsok követésére, módosítására és szintézisére

BÓHM TAMÁS, NÉMETH GÉZA

BME Távközlési és Médiainformatikai Tanszék  
{bohm, nemeth}@tmit.bme.hu

Lektorált

**Kulcsszavak:** formánskövetés, formáns szintézis, lineáris predikció, hangkarakter módosítás

*A pontos formánskövetés problémája régóta kihívást jelent a beszédtechnológia kutatóinak. Ebben a cikkben a szerzők egy olyan eljárást mutatnak be, amely alkalmas a formánsmenetek követésére és módosítására beszédjelben. Ez egy hatékony eszköz a beszéd spektrális tartalmának elemzésére és megváltoztatására, lehetőséget ad a hangminőség és a hangszínezet módosítására. A módszer a lineáris predikció modelljén alapul.*

## 1. Bevezetés

Zöngés hangok képzésekor a hangszalagok kváziperiodikus rezgést végeznek. Az így keletkezett gerjesztőjelet a toldalékcső (garat, orrüreg és szájüreg) rezonátor-rendszerként viselkedve módosítja: a felharmonikusok egyes csoportjait felerősíti (a rezonanciahelyekenél), másokat gyengít. Ezeket a rezonanciahelyeket és környéküket formánsoknak nevezzük.

A rezonanciahely frekvenciája a formánsfrekvencia, ami a toldalékcső átviteli függvényében lokális maximumként jelenik meg. A formánst jellemzi még sávszélessége és amplitúdója. Az előbbi alatt a lokális maximum körüli, annál legfeljebb 3 dB-lel alacsonyabb értékeket tartalmazó leghosszabb frekvenciatartomány szélességét értjük. Az utóbbi a függvény értéke a csúcsonál ([1] 52. o.).

Bár a formánsok és azok változása az idő függvényében (az úgynevezett formánstrajektóriák vagy formánsmenetek) a spektrum és a spektrogram alapján az ember számára könnyen felismerhetőek, a formánsmérés és -követés automatizálása egyáltalán nem triviális feladat.

Jelentős az igény pontos formánskövetésre és a formánsmenetek módosítására, mert ez elősegíti a nyelvészeti kutatásokat és számos alkalmazás megvalósítását is lehetővé teszi. Ilyen alkalmazás például a hangkarakterének megváltoztatása (mint dialektus módosítás, beszédkorrekció vagy a hangszínezet átalakítása) és konkatenatív szövegfelolvasó rendszerek által létrehozott hullámformák formáns-trajektóriáinak simítása. Egy ilyen algoritmus arra is alkalmas lehet, hogy a beszéd személyfüggő jellemzőit megváltoztassa, így a hallgatóban egy másik beszélő érzetét keltse a nyelvi tartalom módosítása nélkül.

A fenti alkalmazások szempontjából fontos, hogy egy olyan módszert alkalmazzunk, amely képes visszaszintetizálni a beszédet a formánsstruktúra megváltoztatása után. Ezt csak egy nagy pontosságú formánsmeghatározó algoritmussal lehet elérni. A számunkra érdekes területeken alkalmazható formánsmódosítás-

sal kapcsolatban eddig nem végeztek kiterjedt kutatásokat, az irodalomban nem találtunk megfelelő megoldást.

A formánskövetést ellenben széles körben vizsgálták az elmúlt évtizedekben. A tradicionális módszerek valamilyen nem-lineáris simítással nyert spektrumon alkalmazznak egy csúcskereső algoritmust ([2] 154. o., [3]). Ide sorolhatjuk a kepsztrális spektrumszűrést is, amikor a kepsztrumból eltávolítjuk a periodicitás miatt megjelenő maximumokat (melyek a zöngének felelnek meg) és Fourier-transzformáljuk azt ([1] 241. o.). Rabiner és Schafer ezt használja fel csúcskeresővel és az „analízis szintézissel” módszerrel kombinálva [4]. Egy másik megközelítés különböző szűrőkészletek alkalmazása [5].

Néhány, elsősorban beszéd felismerésre használt módszer is alkalmazható formánskövetésre, például elterjedtek a rejtett Markov modellen (HMM) [6] és LSP-n (Line Spectrum Pair) alapuló eljárások. Ez utóbbi az LP analízis egyik megvalósítása, ami nem az idő-, hanem a frekvenciatartományban dolgozik és az előállított együtthatók a spektrum nagy amplitúdójú szakaszainak a mozgását követik (ezek a szakaszok nagyjából megfelelnek a formánsoknak).

Ebben a cikkben egy olyan algoritmusról számolunk be, amely képes precíz formáns-meghatározásra, módosításra és visszaszintetizálásra. A kidolgozott algoritmus szemléltetésére egy grafikus alkalmazást hoztunk létre, amely a formánsmenetek sokrétű megjelenítésére képes. A második részben a módszer alapelvét írjuk le, míg a harmadikban az algoritmust ismertetjük. A negyedik fejezet a grafikus alkalmazást ismerteti, míg az ötödikben összefoglaljuk eredményeinket.

## 2. Működési elv

### 2.1. Spektrumszámítás lineáris predikcióval

Mint már korábban tárgyaltuk, a formáns a beszéd-spektrum helyi maximuma. A spektrum számítására alkalmazhatunk gyors Fourier-transzformációt (FFT), de

ez egy számos maximumot és minimumot tartalmazó függvényt állít elő. Egy ilyen spektrum csúcsait nehéz megbízhatóan meghatározni. Formáns-meghatározáshoz a lineáris predikción alapuló spektrumszámítás [2] az elterjedtebb, mert számos előnnyel rendelkezik:

- A predikció fókuszának (azaz az átviteli függvényben szereplő pólusok számának) változtatásával határozhatjuk meg a spektrum „felbontását”.
- A lineáris predikció a (formáns-detekció szempontjából érdekes) csúcsoknál közelíti legpontosabban a spektrumot.
- Még rövid szakaszok elemzésekor is elfogadható eredményeket ad.

A lineáris predikció együttthatóiból becsült átviteli függvény:

$$H(z) = \frac{1}{1 - \sum_{k=1}^p \alpha_k z^{-k}} \quad z = r \cdot e^{j2\pi \frac{f}{f_s}}$$

Az eredmény egy olyan spektrumközelítés, amely jóval simább, mint az FFT, de a formánsokat nem torzíja el.

## 2.2. Formánsmeghatározás

A formánsmeghatározás számos publikált módszere két nagyobb csoportba sorolható: a spektrum-alapú és a pólus-alapú eljárások. Míg az előbbiek az amplitúdó- és fázisspektrumot használják fel, az utóbbiak a z-tartománybeli pólusokkal számolnak.

Spektrumalapú módszer a McCandless-algoritmus, amely a spektrum abszolútértékének logaritmusában keres csúcsokat [7]. Christensen, Strong és Palmer hasonló eljárást dolgozott ki, de ők a log spektrum negatív második deriváltján alkalmazzák a csúcskeresést [8]. Yegnanarayana bebizonyította, hogy a komplex spektrum fázisának első deriváltja jelentős hasonlóságot mutat az amplitúdóspektrum alakjához [9]. Ha ezt deriváljuk kétszer a log spektrum helyett, akkor pontosabb formánsfrekvencia-bebecslést kapunk. Reddy és Swamy módszere egyszerre dolgozik a z- és az f-tartományban [10], így közeli formánsokat is képes elkülöníteni. Bár ezeket a módszereket már többen implementálták és vizsgálták, egyik sem jelent általánosan alkalmazható megoldást a problémára.

Formánskövető rendszerünk pólus-alapú formáns-meghatározást alkalmaz – hasonló ahhoz, amit Slifka és Anderson dolgozott ki beszélőmódosítás céljára [11]. A  $H(z)$  átviteli függvény egy all-pole, tehát zérusokat nem tartalmazó modellt ad a toldalékcsőre. Ennek pólusai a rendszer rezonanciapontjainak, azaz az amplitúdó-spektrum helyi maximumainak felelnek meg. A pólusok az átviteli függvény nevezőjének gyökei:

$$H(z) = \frac{1}{1 - \sum_{k=1}^p \alpha_k z^{-k}} = \frac{1}{\prod_{i=1}^p (1 - p_i z^{-1})}$$

ahol az  $\alpha_k$ -k a lineáris predikciós együttthatók.

A formánsfrekvenciákat és -sávzélességeket a pólusok  $p_i = r_i \cdot e^{j\theta_i}$  alakjából határozhatjuk meg:

$$F_i = \frac{f_s}{2\pi} \varphi_i \quad B_i = \frac{f_s}{\pi} \ln\left(\frac{1}{r_i}\right)$$

Az eredmények rendkívül pontosak, de nem feltétlenül tartozik minden pólushoz formáns (például valós pólusok nem lehetnek formánsok). Ilyen pólusok például a sugárzási ellenállás vagy háttérzaj miatt jelenhetnek meg.

## 3. Algoritmus

Bemeneti adat a hullámforma periódushatár és hanghatár címkékkel. Pontosabb eredmények érdekében megadhatjuk a bemondás fonetikus átíratát is. A jelfeldolgozást két szakaszra bonthatjuk: analízis és szintézis. Az előbbi a formánsmenetek meghatározását jelenti és kimenetei a formánsokat leíró adatok valamint néhány segédinformáció (a lineáris predikció maradékjele és az energiaszintek). Szintézis alatt a formánsok módosítását és az eredmény visszaszintetizálását értjük.

### 3.1. Analízis

#### 3.1.1. LP analízis és pólus-meghatározás

A lineáris predikciós analízist a zöngeperiódusokkal szinkron módon végezzük. Az analízishez Hamming-ablakot használunk, amit mindig két egymást követő periódus hosszára állítunk és periódusonként léptetjük, ezzel biztosítva a megfelelő időfelbontást. Zöngétlen hangok esetén állandó időközönként virtuális alapperiódusokat jelölünk ki.

Előbb Burg módszerrel meghatározzuk a PARCOR együttthatókat, majd azokat lineáris predikciós együttthatókká konvertáljuk, hogy kiszámíthassuk az átviteli függvényt [2].

A lineáris predikciós analízis és szintézis nem garantálja, hogy a bemenő jel energiája megegyezik a kimenő jelével. Annak érdekében, hogy az ebből eredő torzítást elkerüljük, érdemes tárolni minden egyes időszelvény energiáját. A visszaszintetizálás során ezt felhasználhatjuk az eredeti szint visszaállítására. Ugyanezt az adatot csenddetekcióra is használjuk, miután normalizáltuk az időszelvény hosszára. A veszteségmentes kódolás érdekében a lineáris predikció maradékjelét is tárolni kell.

Ezután ki kell számítanunk a rendszer pólusait, azaz az átviteli függvény nevezőjének gyökeit. Bár a nevezőben levő polinom valós együttthatós, a gyökök komplexek is lehetnek, ezért a Newton-Raphson és Brent módszerek nem alkalmasak a feladatra. Helyettük a Laguerre eljárást [12] használja az algoritmus.

Laguerre algoritmusát iteratíván alkalmazva kaphatjuk meg az összes gyököt: egyszer lefuttatva az eredmény egy valós gyök vagy egy komplex konjugált gyökképár. Miután ezekkel leosztottuk a polinomot, újra lefuttatjuk a gyökkeresést.

A komplex konjugált póluspárokhoz a 2. részben leírt formulákkal, valamint a spektrumból meghatározha-

tóak a formánsok adatai (rendre frekvencia, sávszélesség és amplitúdó). Ezeket az adatokat azonban a szintézis során nem használhatjuk fel: pontosabb a számítás, ha a formánsfrekvenciák módosítása során közvetlenül a megfelelő pólus argumentumát változtatjuk.

Az újraszintetizáláshoz tárolni kell azokat a pólusokat is, amelyekhez nem tartozik formáns.

### 3.1.2. Formánsokhoz tartozó pólusok kiválasztása

Nem feltétlenül tartozik minden pólushoz formáns. Annak érdekében, hogy az ilyen pólusokat ki lehessen szűrni, néhány feltételt érdemes megvizsgálni:

- A formánsfrekvencia (amit a pólus argumentumából számíthatunk) nagyobb kell, hogy legyen, mint az alapfrekvencia.
- A pólus abszolútértéke el kell, hogy érjen egy minimumértéket, azaz limitáljuk a formáns sávszélességét.
- Az időszelvény energiája meg kell, hogy haladjon egy küszöbértéket (beszéddetekció).
- Azokat a pólusokat, amelyek valós része nulla vagy nullához közeli, figyelmen kívül hagyhatjuk (ezek például alacsony frekvenciájú, keskenysávú zaj miatt jelenhettek meg).

Kizárólag azokat a pólusokat tekintjük formánsnak, amelyek eleget tesznek mind a négy feltételnek. A kiválasztott komplex konjugált póluspárokból meghatározhatóak a formánsok adatai (frekvencia, sávszélesség és amplitúdó).

A fenti feltételekkel mindössze a nyilvánvalóan téves eredményeket tudjuk kiszűrni és adhatunk egy első közelítést a formáns-pólus összerendelésre. A végleges leképezést a formánstrajektóriák folytonosságát korlátok szerint történő meghatározásával végezzük.

### 3.1.3. Formánstrajektóriák

A feladat tehát a formánsok trajektóriáinak követése a hanghatárok ismeretében. Rendelkezésre áll még a formáns-pólus összerendelés első közelítése. Minden egyes időszelvény formánsait le kell képeznünk a következő időszelvény formánsaira úgy, hogy azok folytonos formánsmeneteket képezzenek.

Egy formánst ahhoz a formánsmenethez rendeljük hozzá, amelyik trajektória utolsó frekvenciaértéke hozzá legközelebb van. Azokat a formánsmeneteket, amelyek nem fedik át egymást, összevonjuk. Szintén összevonjuk a párhuzamosan egymáshoz rendkívül közel futó formánstrajektóriákat. A valószínűtlenül rövid formánsmeneteket töröljük.

Zöngétlen obstruensek (zárhangok, réshangok és zár-rés hangok) előtt és után nem szükséges folytonossági korlátokat alkalmazni, mert ezen hangok képzése olyan artikulációs mozgásokkal jár, amelyek hirtelen változást idéznek elő a toldalékcső rezonanciahelyeiben. Ha a felvétel fonetikus átírata rendelkezésre áll, az algoritmus nem próbálja a formánsmeneteket ilyen hangokon keresztül összekötni. Ha nem, akkor minden hanghatáron töréspontokat iktat be – így néhány hangátmenetnél pontatlan lesz az eredmény, de

a leképezés pontossága az egész felvételre tekintve jobb lesz.

## 3.2. Szintézis

Az első lépésben az analízis kimenetéből előállítjuk a visszaszintetizáló algoritmus bemenetét, azaz módosíthatjuk a formánsmeneteket. A transzformáció módját az adott alkalmazás határozza meg (például trajektóriák simításakor valamilyen interpolációs eljárás). A trajektóriák módosításával közvetve a pólusokat mozgathatjuk a z-síkon.

A második lépés a visszaszintetizálás. A módosított formánsfrekvencia értékek megadják a módosított pólusokat (az analízis során alkalmazott összefüggésekkel). Ezekből összeállítjuk az átviteli függvény nevezőjét alkotó polinomot, majd ezt kifejtjük. A polinom együtthatóiként megkapjuk a lineáris predikció együtthatóit. Ezek és a külön eltárolt maradékkal felhasználásával elvégezzük a lineáris predikciós szintézist. Utolsó lépésként visszaállítjuk az egyes időszelvény energiáját (amit az analízis során eltároltunk). Az eredmény egy új hullámforma fájl, amelyben a formánsmenetek módosultak az eredetihez képest.

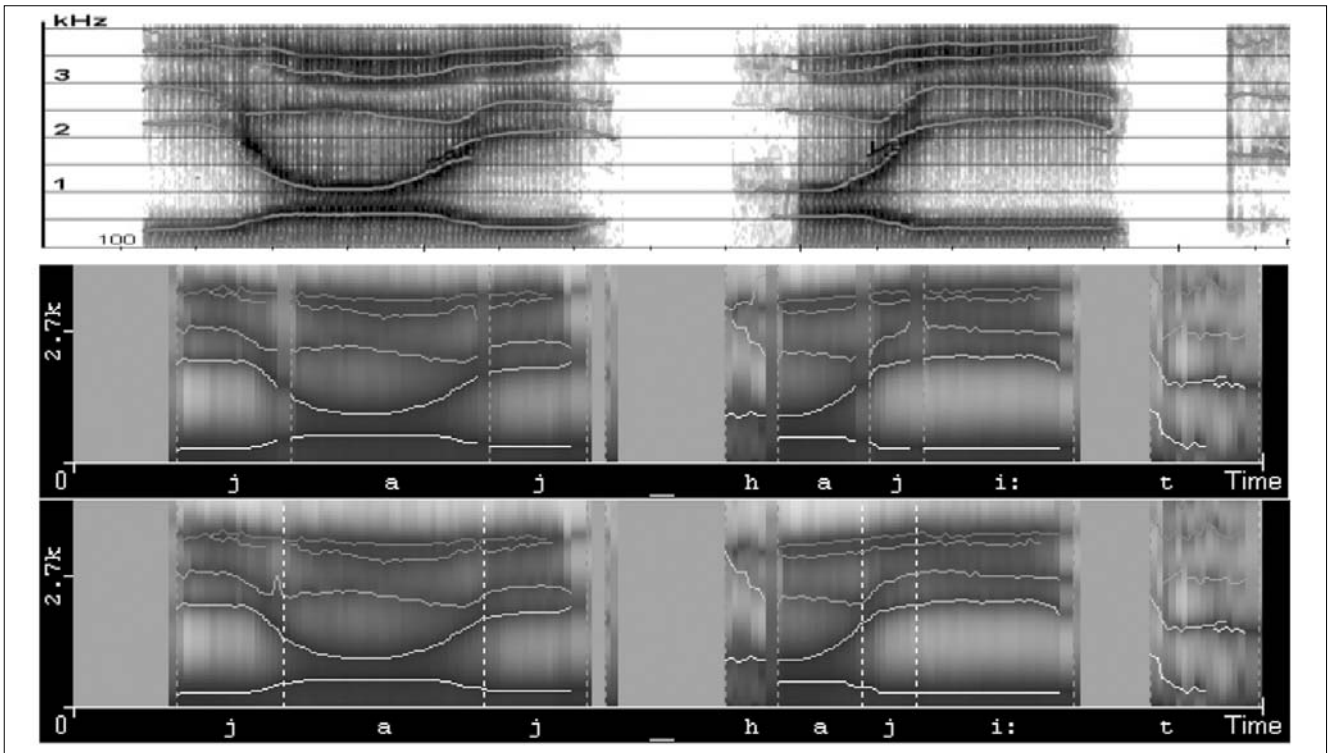
## 4. Eredmények

Az algoritmust implementáltuk és külön ellenőriztük az analízis és a szintézis funkciókat magyar nyelvű felvételekkel.

### 4.1. Analízis

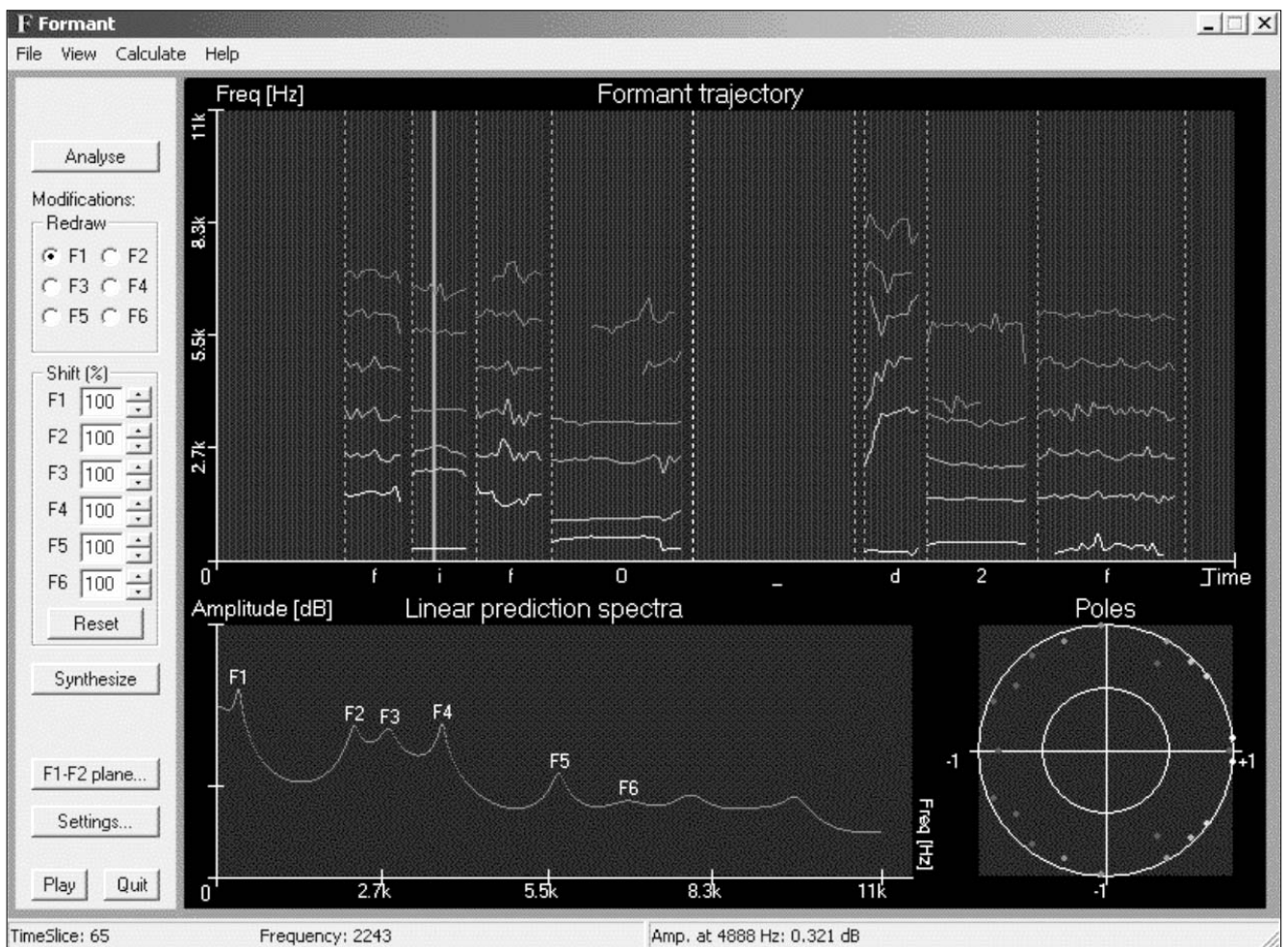
A formáns-meghatározás és -követés pontosságát három módon ellenőriztük. Először az MTA Nyelvtudományi Intézet által a rendelkezésünkre bocsátott spektrogramokkal vetettük össze az eredményeket. Ezek a referenciaként használt spektrogramok a Kay Elementrics CSL 4300B típusú digitális jelfeldolgozóval készültek. Második lépésként az egyes beszédhangok formánsfrekvenciáit összehasonlítottuk a magyar nyelvre jellemző értékekkel ([13] 40. o.).

Végül lemértük a leképezési hibaarányt a [14] definíciója szerint. Ehhez a később ismertetendő grafikus formánskövető program segítségével összehasonlítottuk a felvételek spektrogramjait a kinyert formánsmenetekkel. A tesztalmaz 29 darab, kétszavas, magyar nyelvű felvétel volt férfi bemondóval. Mindössze két felvétel esetén (6,90%) találtunk leképezési hibát az első három formánstrajektóriában – az egyik esetben a harmadik formáns, a másik esetben mind a három formáns leképezése téves volt. [14] szerint egy névleges formánsfrekvencia értékeket használó algoritmus leképezési hibaaránya 3,62-3,99%. Ez alacsonyabb az általunk elért aránynál, de az itt bemutatott algoritmus nem használ előre definiált, tipikus formánsértékeket, így a bemondó személyétől, nemétől és nyelvétől függetlenül működik. Az említett cikk egy hasonló tulajdonságokkal rendelkező módszer hibaarányát 13,04%-ban állapítja meg.



1. ábra Egy rövid felvétel elemzése a Kay Elemetrics CSL 4300B-vel (fent), a cikkben ismertetett algoritmussal – fonetikus átírat nélkül (középen) és fonetikus átíráttal (lent)

2. ábra Grafikus formánskövető és módosító program



A „jaj hajít” bemondásra adott eredmény az 1. ábrán látható. Az ábra alapján látszik, hogy zöngés hangokra pontos volt a formánskövetés, még a második szó /h/-magánhangzó átmenete után is. Ezt az esetet az irodalomban a névleges formánsfrekvenciákat nem használó algoritmusok számára problematikusként említik [14].

#### 4.2. Szintézis

A szintézis lehetőségeit a lehetséges alkalmazások szempontjából vizsgáltuk, mert ezek alapján terveztük a módosító eljárást. A szintézis értékelése és a módszer finomítása még folyamatban van, jelenleg két egyszerű kísérletről tudunk beszámolni.

A formánsszerkezet megváltoztatásával egy magánhangzót egy másik magánhangzóvá alakíthatunk át. Például a „fésű” szóból „fásű”-t csinálhatunk, ha az első formánsmenetet magasabbra, a második formánsmenetet alacsonyabbra toljuk a frekvencia-tengelyen. Egy ilyen formánsmódosított felvételt négy magyar anyanyelvű kísérleti alannyal meghallgattunk és mindannyian egyértelműen a „fásű” értelmetlen szót hallották. Ez a technika hatékonynak bizonyult konkatenatív beszéd szintetizátor akusztikai adatbázisának kiegészítésére olyan hangokkal, amelyeket nem rögzítettünk

a bemondótól (például olyan beszédhangok, amelyek a magyarban nem fordulnak elő).

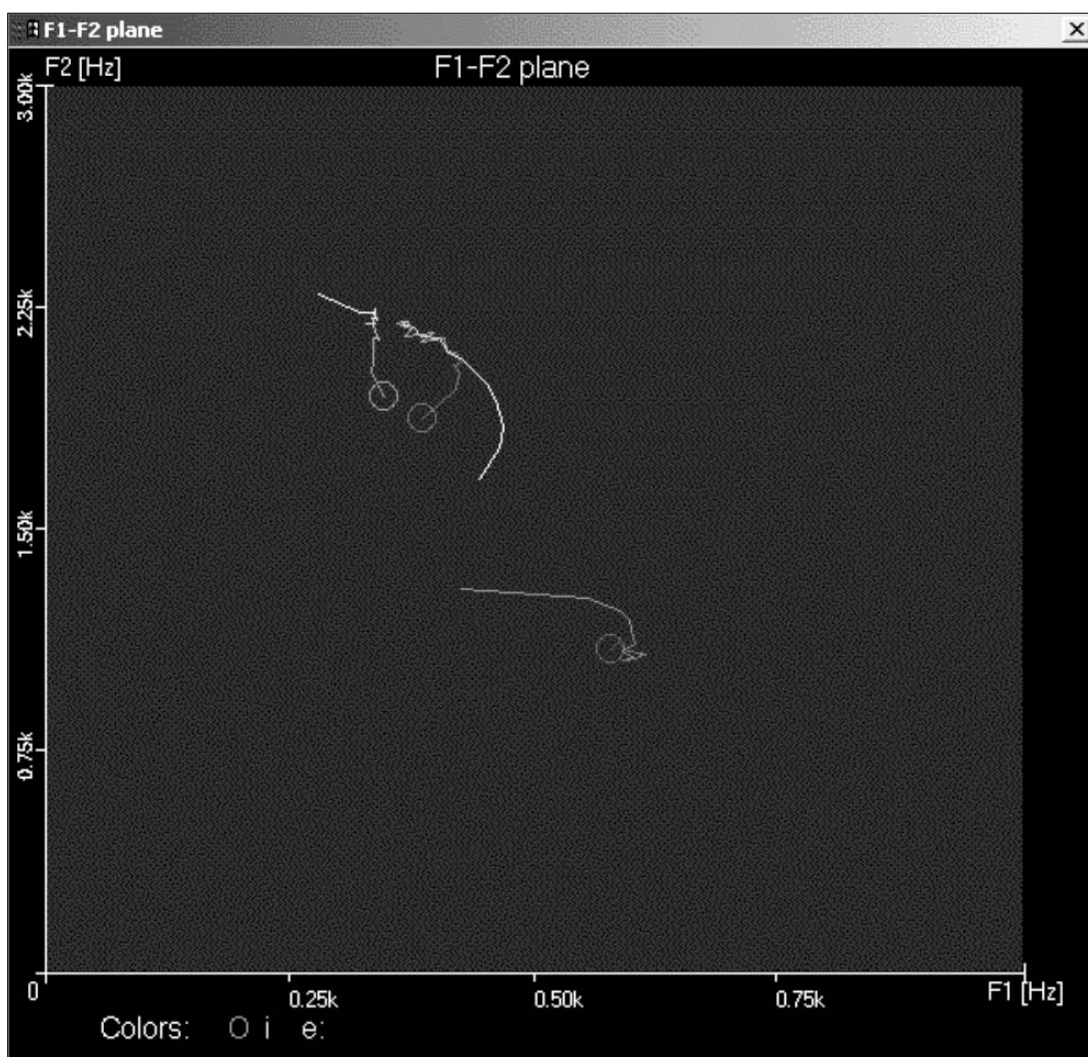
A beszédkutatók körében általánosan elfogadott nézet, hogy a magasabb formánsok a beszélőre jellemzőek. Kezdeti kísérleteket végeztünk beszélőmódosításra, amivel célunk a beszélőspecifikus jegyek megváltoztatása volt annak érdekében, hogy a beszélő személye ne legyen felismerhető. Meghallgatásos tesztek azt mutatták, hogy néhány formánsmenet megfelelő módosítása után a beszélő személyek azonosítása jóval nehezebb a hallgatók számára.

A formánsmenetek módosítása során azt tapasztaltuk, hogy nagy mértékű változtatás esetén jelentősen torzulhat a beszéd. Kisebb (legfeljebb kb. 20%) módosítások során általában jó minőségű, természetes hangzású az újrasszintetizált beszéd.

#### 4.3. Grafikus formánskövető és módosító program

Egy grafikus alkalmazást készítettünk a módszer lehetőségeinek bemutatására és a modell ellenőrzésére (2. ábra). Ez a program oktatási és kutatási célra nyilvánosan hozzáférhető [15].

A program az analízis után az egyes formánstrajektóriákat (legfeljebb hatot) külön színekkel rajzolja ki a frekvencia-idő síkra. Az ellenőrzés érdekében a prog-



3. ábra  
Magánhangzók  
formánsmenetei  
az  $F_1$ - $F_2$  síkon

ram képes spektrogram megjelenítésre, valamint a kétféle ábrázolás egyszerre is látható. Egy időszelvet kijelölve megtekinthetjük a szelethez tartozó rövid idejű lineáris predikciós spektrumot (a program a formánsokat ezen is feltünteti) és a lineáris predikciós szűrő pólusait a komplex z-síkon (az ábra jobb alsó sarkában). Ez utóbbi egy újfajta vizualizációs megoldás, amely ugyanazt az információt hordozza, mint a lineáris predikciós spektrum.

Van lehetőség a formánsmenetek ábrázolására az F1-F2 síkon is (3. ábra). A hagyományos szórásképek helyett az első két formáns együttes mozgását folytonos görbével lehet megjeleníteni.

Az ábrán a felvétel magánhangzóinak F1-F2 menete látható. Az egyes pontok vízszintes koordinátája az első, a függőleges koordinátája a második formáns értékét fejezi ki. A görbék egyes szakaszainak fényereje az idő múlásának felel meg: a magánhangzó első alapperiódusából származó adat jelenik meg a legsötétbben, majd periódusonként világosodik a görbe és az időben utolsó szakasz a legvilágosabb. A jobb láthatóság érdekében egy kör is megjelenik a kezdőpontok körül.

A formánsmeneteket módosíthatjuk kézi átrajzolással vagy trajektóriánként egy szorzó beállításával. Az eredményt újraszintetizálhatjuk és elmenthetjük egy hullámforma fájlba.

A program széles felhasználási lehetőségekkel rendelkezik a fonetika oktatása területén. Szemléltető eszközként használható például a beszédhangok formánsstruktúráinak bemutatására vagy a magánhangzók megkülönböztető tulajdonságainak képi megjelenítésére. Alkalmas továbbá egy fonéma különböző realizációi közötti hasonlóságok és különbségek vizsgálatára és koartikulációs hatások bemutatására. Fonetikai vizsgálatok eszközeként is szolgálhat, többek között perceptiós tesztek hanganyagának előállításához, nyelvjárássok vizsgálatához „analízis szintézissel” módszerrel és szemléltető ábrák készítéséhez. Jelen cikk ábráit is ezzel a programmal állítottuk elő.

## 5. Összefoglalás

Egy általános formánskövető és módosító algoritmus számos alkalmazás megvalósításához szükséges. Cikkünkben egy ilyen algoritmust mutattunk be és értékeltünk. Eljárásunk elfogadható eredményeket ad kevés információ alapján is és pontos eredmények érhetőek el további bemeneti adatokkal (fonetikus átirat). Az algoritmust implementáltuk és beépítettük egy grafikus alkalmazásba.

Ez a nyilvánosan elérhető program hasznos oktatási és kutatási eszköz lehet – ezt igazolja, hogy a szerzők oktatási tevékenységén kívül az ELTE Bölcsészettudományi Karán is használják. Továbbá laboratóriumunkban sikeresen alkalmaztuk a magyar magánhangzók formánsstruktúrájának kiterjedt vizsgálatára (melynek eredményei feldolgozás alatt állnak).

## Irodalom

- [1] Gordos G., Takács Gy.:  
Digitális beszédfeldolgozás,  
Műszaki Könyvkiadó, Budapest, 1983.
- [2] Markel, J. D., Gray, A. H.:  
Linear Prediction of Speech,  
Springer-Verlag, Berlin, 1976.
- [3] Lobanov, B., Levkovskaya, T., Kheidorov, I.:  
„Speaker and channel – Normalized set of formant parameters for telephone speech recognition”,  
Proc. of Eurospeech 1999, Vol. 1., pp.331–334.
- [4] Rabiner, L. R., Schafer, R. W.:  
Digital Processing of Speech Signals, Prentice-Hall,  
Englewood Cliffs, 1978.
- [5] Ouni, K., Lachiri, Z., Ellouze, N.:  
„Formant estimation using Gammachirp filterbank”,  
Proc. of Eurospeech 2001, Vol. 4., pp.2471–2474.
- [6] Weber, K., Bengio, S., Bourlard, H.:  
„HMM2 – Extraction of formant structures and their use for robust ASR”,  
Proc. of Eurospeech 2001, Vol. 1, pp.607–610.
- [7] McCandless, S. S.:  
„An algorithm for automatic formant extraction using linear prediction spectra”,  
IEEE Trans. on Acoustics,  
Speech and Signal Processing, Vol. 22, no.2, 1974.
- [8] Christensen, R. L., Strong, W. J., Palmer, E. P.:  
„A comparison of three methods of extracting resonance information from predictor coefficient coded speech”,  
IEEE Trans. on Acoustics,  
Speech and Signal Processing, Vol. 24, no.1., 1976.
- [9] Yegnanarayana, B.:  
Formant extraction from linear prediction phase spectra,  
Journal of the Acoustical Society of America,  
Vol. 63., p.1638., 1978.
- [10] Reddy, N. S., Swamy, M. N. S.:  
„High-resolution formant extraction from linear prediction phase spectra”,  
IEEE Trans. on Acoustics,  
Speech and Signal Processing, Vol. 32, no.6., 1984.
- [11] Slifka, J., Anderson, T. R.:  
Speaker modification with LPC pole analysis,  
Proc. of ICASSP 1995, pp.644–647.
- [12] Orchard, M. T.:  
„The Laguerre method for finding the zeros of polynomials”,  
IEEE Transactions on Circuits and Systems, 1989.  
Vol. 36, no.11, pp.1377–1381.
- [13] Olasz G.:  
Elektronikus beszédelőállítás  
(A magyar beszéd akusztikája és formánszintézise),  
Műszaki Könyvkiadó, Budapest, 1989.
- [14] Lee, M., van Santen, J., Möbius, B., Olive, J.:  
„Formant tracking using segmental phonemic information”,  
Proc. of Eurospeech 1999, Vol. 6., pp.2789–2792.
- [15] <http://fonetika.nytud.hu>



# Az akusztikai arculat szerepe az infokommunikációs szolgáltatók megítélésében

NÉMETH GÉZA

BME Távközlési és Médiainformaticai Tanszék  
nemeth@tmit.bme.hu

Lektorált

**Kulcsszavak:** vállalati akusztikai arculat, infokommunikációs szolgáltatásminőség, IVR, számfelolvasás, dátumfelolvasás

A jelen cikkben a jól ismert „vállalati arculat” kifejezésnek az akusztikai jelenségekkel kapcsolatos kiterjesztésére teszünk kísérletet. Az „akusztikai arculat” (acoustic company image) fogalmát a szerző 1998-ban alkotta meg. Ennek tudatos kialakítása a szervezetek részéről a külvilág felé hangjelenségekkel történő megjelenés jellemzőinek és minőségének rendszerezett tervezését és megvalósítását igényli. Az általános definíciókon és témakör ismeretetésén túlmenően infokommunikációs szolgáltatások konkrét példáival illusztráljuk az akusztikai arculat jelentőségét.

## 1. Bevezetés

Az „akusztikai arculat” fogalmát a szerző egy, a beszédtechnológiák bevezetésére vonatkozó 1998-as vizsgálata kapcsán [1] alkotta meg. Az akkori (és sajnos nagyrészt a mai) helyzetet jól jellemzi egy bank informatikai vezetőjének válasza az egyik kérdőív kérdésre: „Tudjuk, hogy a bemondásaink és a generált válaszaink minősége rossz, de ez nem igazán probléma. A fontos az, hogy rövid időn belül ki tudtunk fejleszteni egy rendszert, ami információt tud adni az adatbázisainkból.”

Annak ellenére, hogy a vállalatok jelentős erőfeszítéseket fordítanak a minőség szempontjainak érvényesítésére (lásd total quality management), az akusztikai minőség fogalma a legtöbb ember fejében koncerttermekre, színházakra, audiofil berendezésekre és hasonlókra korlátozódik. Az angol nyelvű szakirodalomban újabban jelentős figyelmet kap a hangminőség (sound quality, [2]) témaköre olyan jelenségek tervezése kapcsán, mint a gépkocsik kipufogó hangja, az ajtó becsapódásának zöreje, a háztartási gépek működésével járó hangjelenségek stb.

Az akusztikai arculat tudatos kialakítása a szervezetek részéről a külvilág felé hangjelenségekkel történő megjelenés jellemzőinek és minőségének rendszerezett tervezését és megvalósítását igényli. A következő szakaszban röviden áttekintjük a vállalati arculat általános definícióját, majd a vállalati akusztikai arculat témakörét tekintjük át röviden. A negyedik szakaszban infokommunikációs szolgáltatások konkrét példáival illusztráljuk az akusztikai arculat jelentőségét.

## 2. A vállalati arculat fogalma

A „vállalati arculat” hosszabb ideje ismert és használt fogalom. Ezt jól mutatja, hogy a „company image” kulcsszóra a Google a cikk írásának időpontjában (2006. február) mintegy 25,6 millió, a magyar változatra pedig 551 találatot adott. Érdemes megkülönböztetni a vállalati azonosságtudat (company identity) és a vállalati ar-

culat (company image) fogalmát [3]. A megnevezések általánosított formája a szervezetekre általában vonatkozik (corporate identity/image). Előbbi a vállalat „mindazon intézkedéseinek összessége, amely meghatározó a szervezet egészére nézve” [4]. Másszóval azon „tulajdonságok, intézkedések összessége, melyek azonosítják a szervezetet önmagával, illetve megkülönböztetik más szervezetektől” [5]. Az utóbbi pedig arra vonatkozik, hogy a külső személyek és szervezetek milyenek látják az adott vállalatot (gyakran beleértik azt is, hogy a vállalat milyenek akarja láttatni magát). Sajnos, a hazai szóhasználat ezekben a témakörökben sem egységes. Még a magyar PR Szövetség ajánlásával kiadott meghatározás-lista is számos angol szót tartalmaz [5].

Fontos megjegyezni, hogy a vállalati arculat szubjektív tényezőktől függ, ezért a cégek hangsúlyt fektetnek az egyszerű, könnyen értelmezhető jellemzőkre. Érthető, hogy elsősorban a vizuális jegyekre, különösen a logóra összpontosítanak, hiszen gyakran annak alapján azonosítják a vállalatot. Jellemző, hogy egy – a vállalati arculat tervezéséhez szempontokat adó – huszonekét jellemzőt tartalmazó paraméterlista [3] csak egyetlen akusztikához kapcsolható tanácsot tartalmaz: „Egységes, professzionális módon válaszoljuk meg a telefonhívásokat”.

Mindez megfelelő lehet fizikai formában megjelenő, megnézhető és megtapintható termékek esetén, bár az éles versenyben döntő lehet egy autóajtó záródásának mikéntje is. Az infokommunikációs szolgáltatók piacán azonban sokkal szélesebb látókörű megközelítés szükséges.

## 3. Az akusztikai arculat áttekintése

Az akusztikai arculat definícióját az általános definícióból származtathatjuk:

**Az akusztikai arculat azt fejezi ki, hogy külső személyek és szervezetek milyenek látják az adott vállalatot, a vállalatot magát, valamint a termékeit**

## és szolgáltatásait jellemző akusztikai jegyek, paraméterek és események alapján.

Az akusztikai arculat néhány eleme hosszabb ideje reflektorfényben van. Ilyenek elsősorban a televíziós reklámok. Ebben az esetben a vizuális és az akusztikai információ együtt jelenik meg. Az akusztikai paraméterek kevésbé lényegesek, mert a vizuális élmény döntő módon befolyásolja a nézőt [6]. A rádiós műsorok és vállalati szignálók valamint a reklámok azon akusztikai arculati elemek közé tartoznak, melyeket többnyire tudatosan terveznek meg a vállalatok általános arculatáért felelős szervezetek.

A beszédminőség („voice quality” – 1,59 millió, magyarul 144 Google-találat) a távközlő hálózatok alapvető jellemzője és olyan paraméterekkel írjuk le, mint a jel/zaj viszony (SNR), torzítás, bithibaarány (BER) vagy szubjektív minőségi érték (Mean Opinion Score, MOS).

A hangminőség („sound quality” – 7,56 millió, magyarul 34.800 találat) két lényegesen eltérő megközelítést takar. A legtöbb esetben hagyományos hangtechnikai-elektroakusztikai értékelést fejez ki, pl. hangszórók, fejhallgatók, erősítők, CD-játszók, termek, stb. minősítésére szolgál. Az elmúlt években azonban előtérbe került a használata más, hangjelenségeket produkáló eszközök, például gépkocsik, mosó- ill. fűrógépek esetében is. Kiderült, hogy az olyan objektív paraméterek, mint pl. a hangosság szint, nem feltétlenül korrelálnak a felhasználók értékelésével. Előfordul, hogy egy hangosabb mosógép hangját kellemesebbnek tartják, mint egy csendesebb, de „furcsa” hangú versenytársát [2].

Noha az ilyen jellemzők is befolyásolják a vállalati arculatot, de mégis erősen kötődnek egy bizonyos termékhez. Nehéz olyan fűrógépet elképzelni, ami működés közben az egységes vállalati dallamot játssza. Bizonyos termékcsoportok esetében, – például sportkocsik motor és kipufogó hangja (kapcsolható reklámszöveg: „az erő hangja”), vagy asztali számítógépek ventilátorzaja (reklámszöveg: „csendes segítő társ”) – vagyis az adott termékek hangminősége a terméket gyártó egész vállalat arculatára is jelentős hatást gyakorolhat.

Az akusztikai arculat szerepe egyre növekszik és egyre kifinomultabb megközelítést igényel az alábbi okokból:

- Erőtéljes az a tendencia, hogy egyre több szolgáltatást és információt beszéddel nyújtsanak (főleg telefonkapcsolaton alapuló hívásközpontokban – call center – de az internet-alapú megoldások száma is növekszik).
- Hangsúlyt kap a célcsoportokra vagy akár személyekre szabott megközelítés (például léteznek már olyan vállalkozások az USA-ban, melyek arra szakosodtak, hogy egy adott vállalatra jellemző hangposta üdvözlő üzenetet, vagy várakozás alatti zenét terveznek és valósítanak meg).

Az akusztikai arculat tudatos formálásához fel kell mérni és kezelni kell azokat az objektív és szubjektív tényezőket, melyek befolyásolhatják azt. Ennek két megközelítése lehetséges.

A *felülről lefelé haladó* (top-down) megközelítés során egy központi szervezeti egység felméri azokat a szolgáltatásokat és jellemzőket, melyek befolyásolhatják az akusztikai arculatot, majd megfelelő szolgáltatásokat, előírásokat és eljárásrendet alakít ki, melyeket a vállalat valamennyi egysége alkalmaz. Ez a megközelítés kis-közepes vállalatok esetében lehet eredményes, ahol egyetlen szervezeti egység képes lehet minden kapcsolódó feladatot megoldani.

Az *alulról felfelé haladó* (bottom-up) megoldás esetén a vállalati akusztikai arculat tervezői az általános koncepció megalkotásáért, annak az üzleti szolgáltatási és kommunikációs folyamatokba történő beillesztésért és alkalmanként a megvalósításban vagy annak ellenőrzésében történő részvételért felelnek. Az arculati jellemzőket az adott szolgáltatást vagy paramétert létrehozó szervezeti egység többé-kevésbé önállóan valósítja meg. Az eredményeket és az esetleges problémákat a szervezeti egységek visszacsatolják a vállalati akusztikai arculat tervezőihez. Ez a megközelítés lehet a legmegfelelőbb nagyvállalatok számára, különösen, ha több nyelv és/vagy kultúra is érintett működési területükön. Szerencsés, ha az akusztikai arculat tudatos kialakítását a meglévő szervezetbe és eljárásrendbe (pl. TQM) építik be. Fontos szempont, hogy a helyes megoldáshoz a marketing szakembereknek és az adott akusztikai esemény szakértőinek (akusztikus, fonetikus, gépészmérnök, beszédtechnológus stb.) szoros együtt kell működniük [7].

Az akusztikai arculat színvonalas kialakításához elengedhetetlen, hogy a vállalat minden vezetési szintje tudatosan törekedjen arra, hogy a vállalattal kapcsolatba kerülő minden ügyfél a lehető legjobb benyomást szerezhessen. Az a megközelítés, melyben egy külső szakértőt úgy foglalkoztatnak, hogy *„fizetek neked az akusztikai arculat kialakításáért, de belső vállalati munkatársakat és információt nem tudok rendelkezésedre bocsátani”* –, kis eséllyel hoz jó eredményt.

Jellegzetes gond az, hogy a vállalati beszerzési eljárásrendben gyakran nem jelennek meg az akusztikai arculat szempontjai. Helyette elsősorban az ár és néhány műszaki jellemző dominál. Jó példa erre az, hogy a magyar távközlési szolgáltatók többsége olyan hangposta rendszert használ, ami a telefonszámokat úgy mondja be (az angolból átvett szoftver magyarítása következtében számjegyenként), ahogyan egyetlen magyar ember sem beszél. A rendszerek teljes bekerülési értéke több milliárd forint, működésük több millió ügyfelet érint és a problémát legfeljebb egy emberhónapnyi programozói és beszédtechnológia szakértői munkával meg lehetne oldani. A jó hangminőségű üzenetet csak egyszer kell gondosan, igényesen előkészíteni – beszédtechnológiai szakember bevonásával – és utána milliók fogják hallgatni nap mint nap. Érdemes lenne tehát az egyszeri befektetés.

Gyakran felmerül a felelősségi körök átfedése. Például hagyományosan többnyire a marketing osztályok a felelősek azért, hogy kiválasszák azt a színészt vagy színésznőt, akinek a hangját TV és rádió reklámokban

a vállalat azonosítására használni fogják. Mikor azonban egy részben automatizált hívásközpont kialakítása merül fel, akkor elsősorban az ügyfélszolgálati osztály határozza meg, hogy kinek a hangjával veszik fel (a sokszor gyakran változó) rendszerüzeneteket. A két hang többnyire különbözik. A magasabb vezetési szintek bevonása nélkül a két megközelítés harmonizálására kis esély van.

#### 4. Infokommunikációs szolgáltatások és az akusztikai arculat

A vállalatok akusztikai arculata talán az infokommunikációs szolgáltatások területén a legfontosabb, mert itt a felhasználói interakciók és élmények többsége akusztikai jelenségekhez kapcsolódik és a bevételek többsége is ebből származik. A következőkben néhány – kimondottan erre a területre jellemző –, akusztikai arculathoz kapcsolódó tényezőt veszünk számba, valamint kapcsolódó felhasználói tesztek is bemutatunk.

##### 4.1. Az akusztikai arculat összetevői infokommunikációs szolgáltatásokban

Beszédszolgáltatások esetén a legkézenfekvőbb paraméter az átviteli minőség. Az ezekre vonatkozó jellemzőket szabványosítási folyamatok során részletesen szabályozzák. Az észlelt minőség azonban állandó műszaki minőség mellett is változhat. Jó példa erre az, hogy a mobil és az internet-telefonos széleskörű használata során az ügyfelek gyakran találkoznak erősen torzított beszéddel és hozzászoknak ahhoz. Emiatt a változatlan műszaki tartalom mellett is a hagyományos vezetékes szolgáltatás sok mobil felhasználás után jobbnak tűnhet.

Egy másik meghatározó tényező az ügyfélszolgálatok alapvető eszközévé vált hívásközpontok szolgáltatásminősége. Két alaptípusukat különböztetjük meg. A *kimenő* (outbound) *hívásközpontból* felhívják az ügyfelet. Ilyen rendszereket elsősorban a marketing vállalkozások, közvéleménykutatók, biztosítók, utazásközvetítők, stb. alkalmaznak. A legtöbb esetben jól képzett kezelők veszik át a szót az általában automatizált hívásfelépítési folyamat után.

Egy jellemző kivétel az úgynevezett SMS-felolvasás, amikor rövid szöveges üzenetet küldenek egy olyan vezetékes telefonszámra, melyhez nem kapcsolódik annak szöveges fogadására alkalmas telefon- vagy faxkészülék. Ekkor az üzenetet egy gépi szövegfelolvasó (Text-To-Speech, TTS) rendszer gépi beszéddel olvasa fel. Attól függően, hogy a TTS megoldást milyen környezetbe ágyazzuk, az észlelt szolgáltatásminőség és ezzel az akusztikai arculat jelentősen eltérővé válhat. A mai TTS rendszerek általában jól érthető hangot szolgáltatnak, azonban a beszédük kissé robotos. Azt is figyelembe kell venni, hogy a TTS szöveget (karakter-sorozatot) olvas fel, tehát ha a szöveg hiányos (például hiányoznak ékezetek), akkor a felhangzó beszéd sem lesz helyes.

Ha először magát az üzenetet a tetszőleges szöveg felolvasására felkészített TTS rendszerrel olvastatjuk fel, majd a feladó telefonszámát és a feladási időpontot is szintén a TTS mondja be, a felhasználónak nem lesz lehetősége a gépi hanghoz történő alkalmazkodásra, a feladó száma alapján a várható tematikára való felkészülésre, stb. A lényegi üzenet után korlátozott minőségben felhangzó kiegészítő információk feldolgozása nehezíti a fő mondanivaló megértését. Ha viszont előbb a telefonszámot és az időpontot mondatjuk be egy erre a célra fejlesztett, az emberi bemondó hangminőségét közelítő rendszerrel (például [8]) és csak ezután érkezik a fő üzenet a TTS segítségével, akkor az ügyfél jobb eséllyel értheti meg a felhangzó információkat és ily módon a vállalat akusztikai arculata is kedvezőbb lesz. Természetesen nagyon fontosak a további kiegészítő tulajdonságok is. Az éjjel egy órakeres csörgő telefon, majd a felhangzó SMS üzenet a legjobb felolvasási technológia mellett is leronthatja a vállalati arculatot.

A *fogadó* (inbound) *hívásközpontokat* az ügyfél hívja fel és előfordulhat, hogy azonnal jól képzett kezelőkhöz kapcsolják. Az ellenérzések zöme az automatizált interaktív hangválasz (Interactive Voice Response, IVR) megoldásokhoz kapcsolódik. Ha az ügyfelek költségmentesen vehetik igénybe jól képzett ügyfélszolgálati munkatársak idejét, kevesen kezdenek el összetett menürendszerekben barangolni és a szükséges kezelők száma gyorsan nőni kezd. Ennek ellensúlyozására a vállalatok gyakran arra kényszerítik ügyfeleiket, hogy olyan, viszonylag egyszerű feladatokat, mint előre fizetett kártya feltöltése, egyenleglekérdezés stb. automatizált rendszerekkel oldjanak meg.

Ezeket a rendszereket gyakran olyan nagyvállalatok szállítják, melyek elsődleges piaca(i) angol (vagy legfeljebb 6-8 másik) nyelvet használ(nak). A kisebb piacokra készített nyelvi változatok gyakran rossz minőségűek (az elsődleges nyelv szerkezetét követik) még olyan alapvető témakörök esetében is, mint a telefonszámok, dátumok és pénzüsszegek felolvasása (lásd 4.2 szakasz). A minőséget tovább rontják a határidős elvárások, mert az ilyen vizsgálatokat, illetve honosításokat többnyire a projektek végére hagyják. Gyakran nehezen megoldható, vagy teljesen elmarad a menürendszerek szerkezeti és hangminőségének, hatékonyságának mérése és optimalizálása.

Ha egyszer egy rossz minőségű, de az elemi használhatóságot kielégítő rendszert üzembe állítanak, nehéz javítani rajta. A nagyvállalatok gyakran vizsgálják az ügyfélpanaszok számát és okát, de – még akkor is, ha elégedetlen a rendszerrel – nagyon ritka az, hogy az ügyfél panaszokkal töltse az idejét olyan esetekben, amikor a közvélekedés szerint a változás esélye csekély, nem remélhető, hogy az egyéni megjegyzésekre figyelni fognak.

Az illetékes menedzserek esetleg elismerik, hogy a rendszer rossz minőségű, de azzal érvelnek, hogy a javítás profitnövelő hatását nem (vagy nehezen) lehet kimutatni és a változtatást körülményes és költséges len-

ne végrehajtani. A helyzetet jól jellemzi egy ügyfélszolgálati vezető véleménye: „Tudjuk, hogy az ügyfelek utálják az IVR-t. Mi is utáljuk. De nincs pénzünk több kezelőre...”

Nemrég kezdett el jelentős teret nyerni – elsősorban az USA-ban – a beszédfelismerésen alapuló adatbevitel, illetve lekérdezés (menetrend, név, cím, időpont stb.) Bizonyos ilyen alkalmazások (pl. automatizált alközponti kezelő) egy ideje hazánkban is elérhetőek [9]. Ezekben az esetekben az akusztikai arcukat javítására vonatkozó hatás vagy a tranzakció gyorsulásán (pl. banki egyenleg lekérdezése), vagy pedig emberi kezelővel nem vagy nehezen megvalósítható szolgáltatások kialakításán (például a legközelebbi bolthoz vagy bankhoz való irányítás) alapul.

Gyakran előforduló hiányosság annak figyelmen kívül hagyása, hogy a stúdiókörülmények között kiválóan hangzó felvétel (vállalati „akusztikus logó”, zene, csengőhang, előre felvett üzenetek) nem garantálja azt, hogy a telefonos alkalmazás is sikeres lesz. Egy akusztikus jel 20-ról 3,1 kHz-re történő sávkorlátozása többnyire jelentős torzulásokat okoz (különösen a széles spektrumú – pl. sziszegő – beszédhangoknál és a tranziens jelenségeknél, például az ütős hangszereknél). Ahhoz, hogy reális legyen a megítélés, minden minősítést a végfelhasználói csatornán kell elvégezni.

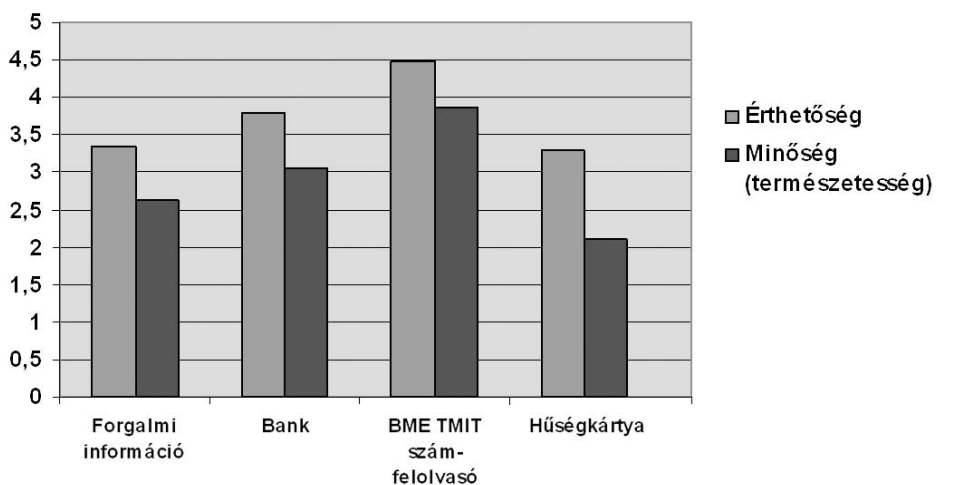
#### 4.2. Az akusztikai arcukat meghatározó néhány szolgáltatás vizsgálata

Annak érdekében, hogy bemutathassuk az akusztikai arcukat változását a különböző szolgáltatások tükrében, empirikus vizsgálatokat végeztünk beszédinformációs rendszerekben alkalmazott alapvető bemondástípusokra [10]. A hangfelvételeket a BME TMIT-en valódi szolgáltatásokról készítettük. A felvételeket ötven ép hallású egyetemi hallgató értékelte ötfokozatú skálán (1=legrosszabb, 5=legjobb). A tesztalanyoknak külön-külön kellett értékelniük a minták funkcionalitását (érthetőségét) és minőségét (természetességét). Minden mintát kétszer játszottunk le. Először valamennyi mintát végighallgatták, majd az újbóli lejátszás során a minták között rövid szünetet tartottunk. Ezalatt kellett az előzőleg meghallgatott minta osztályozását elvégezni.

Az 1. ábrán számfelolvasási technológiák értékelésének átlagát láthatjuk. Noha az érthetőségi értékek változása is jelentős és csak egy megoldás haladta meg a 4-es átlagot, a minőségi paraméterek eltérései a legfontosabbak a jelen cikk szempontjából. Még a legjobb rendszer is 0,6 ponttal kisebb értéket kapott a minőségi skálán, mint az alapvető funkcionalitást jelző érthetőség. A legrosszabb esetben pedig a különbség az 1-et is meghaladja. Egyetlen rendszer közelíti meg a 4-es értéket, ami a hasonló vizsgálatok esetén a széleskörű

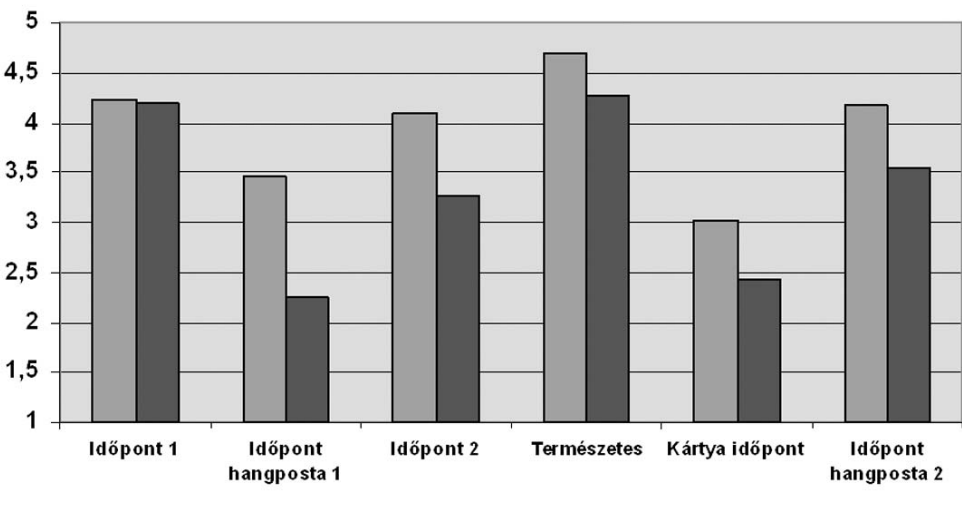
1. ábra  
Különböző számfelolvasók érthetősége és minősége [10]

Forgalmi információ = egy előfizető beszélgetéseinek havi díja,  
Bank = számlaegyenleg értéke,  
BME TMIT számfelolvasó = a [6] szerinti mintarendszer,  
Hűségkártya = egy pontgyűjtő alkalmazás



2. ábra  
Különböző dátum- és időpont felolvasók érthetősége és minősége [10]

Időpont 1 és 2 = távközlési vállalatok pontos idő bemondása,  
Időpont hangposta 1 és 2 = távközlési vállalatok hangpostájának az üzenet időpontját tartalmazó bemondása,  
Kártya időpont = előre fizetendő (pre-paid) előfizetéshez kapcsolódó időpont bemondás,  
Természetes = professzionális bemondó



elfogadottság szintjének felel meg. A Hűségkártya jelle-  
gű megoldások egyik legfontosabb célja az, hogy ked-  
vező képet alakítsanak ki az adott vállalatról. A jelen  
esetben az alacsony érték valószínűsíti azt, hogy a  
számfelolvasó rossz minősége ezt a hatást jelentősen  
csökkenti vagy esetleg negatív irányba fordítja.

A legjobb és a legrosszabb természetességű rend-  
szer között a különbség csaknem két jegynyi, ami meg-  
alapozhat egy „hagyományos mosópor <> korszerű mo-  
sópor”-jellegű marketing kampányt.

A 2. ábrán különböző dátum és időpont bemondó  
technológiák értékelését láthatjuk. Noha az időpont be-  
mondások és az egyik hangposta érthetőség szem-  
pontjából egyaránt megközelítik a természetest, a mi-  
nőségi oszlopok jelentős eltéréseket mutatnak. Itt már  
csak az időpont 1 rendszer áll közel a természeteshez.  
Az egyik hangposta és a pre-paid rendszer (Kártya idő-  
pont) minőségi értékelése olyan alacsony (2,5 alatt),  
hogy az valószínűleg az általános vállalati arculat le-  
romlását is eredményezheti.

Az ebben a szakaszban bemutatott bemondások  
alapvetőek és az ügyfelek által gyakran használt rend-  
szerelemek nagy forgalmú infokommunikációs szolgá-  
latásokban. Sajnálatos, hogy noha viszonylag kis be-  
fektetéssel minőségük jelentősen javítható lenne (eh-  
hez hazánkban is rendelkezésre állnak a szükséges  
műszaki-tudományos ismeretek), egyelőre a legtöbb  
vállalat nem fordít erre figyelmet.

## 5. Összefoglalás

A vállalatok akusztikai arculata kialakul, függetlenül at-  
tól, hogy azt tudatosan formálják vagy sem. A cikkben  
első lépésként javaslatot tettünk ennek a témakörnek  
a rendszerezett áttekintésére. Kísérletekkel bizonyítottuk,  
hogy az infokommunikációs szolgáltatások értéke-  
lésében a minőségi/természetességi szempontok be-  
vezetése finomabb megkülönböztetést tesz lehetővé.  
A rendszerek alaposabb tervezése és értékelése lény-  
egesen hozzájárulhat életminőségünk növeléséhez.  
Ennek elhanyagolása az ügyfelek számára a termé-  
szetestől távolálló interakciókat eredményez, ami le-  
ronthatja a vállalati arculat javítására jelentős költsé-  
gekkel más médiumokban kifejtett kampányok hatását.

A jelen dolgozat a 2005-ös Forum Acousticum konfe-  
rencia meghívott angol nyelvű előadásának [11] átdol-  
gozott és kibővített változata.

### Köszönetnyilvánítás

A szerzőnek a témával kapcsolatos elgondolásait sok be-  
szélgetésben formálták a BME TMIT Beszédtechnológiai  
Laboratóriumának munkatársai és ipari partnerei. Külön  
köszönet illeti Olasz Gábort és Zainkó Csabát. Hálával  
tartozom Szőke Veronikának, hogy közgazdász diploma-  
tervének területétől ezt a témakört választotta, új gondola-  
tokkal gazdagította eredeti elképzeléseimet és hozzájárult  
eredményeinek felhasználásához.

### Irodalom

- [1] Németh, G.,  
„From Near-Nil to Everyday Life: Speech Technology  
Based Telecommunications Services in Hungary”,  
Proc. of IEEE-ESCA 4th Workshop on Interactive  
Voice Technology for Telecom. Applications,  
IVTTA98, Sept. 1998, Torino, Italy,  
pp.191–196.
- [2] R.H. Lyon,  
„Designing for Product Sound Quality”,  
Marcel Dekker, 2000.
- [3] S. MaGee,  
„How to Create Your Company's Identity”,  
www.EdwardLowe.org,  
©1992-2005 Edward Lowe Foundation.
- [4] Barát Tamás,  
„Tolmács a hídon”, egyetemi jegyzet, Gödöllő, 2001.  
részletek: <https://phigy.hu/node/849>
- [5] Magyar Public Relations Szövetség,  
„PR meghatározások”, 2000.  
<http://www.mprsz.hu/szolgáltatatas/szotar/>
- [6] Illényi A., Csányi K.,  
„Mérnöki pszichoakusztika”,  
BME egyetemi jegyzet, Budapest, 2001.
- [7] G. Olasz, G.  
„Phonetics and Speech Technology”,  
Proc. of Forum Acousticum 2005, Budapest,  
pp.2663–2667.
- [8] G. Olasz, G. Németh,  
„IVR for Banking and Residential Telephone  
Subscribers Using Stored Messages Combined with  
a New Number-to-Speech Synthesis Method”,  
In: D. Gardner-Bonneau ed.,  
Human Factors and Interactive Voice Response  
Systems, Kluwer, 1999.,  
pp.237–255.
- [9] Fegyő, T., Mihajlik, P., Szarvas, M., Tatai, P., Tatai, G.:  
„Voxenter – Intelligent Voice Enabled Call Center  
for Hungarian”, Proc. of EUROSPEECH'03,  
Geneve, Switzerland, September 1-4.
- [10] Szőke Veronika,  
„Távközlési vállalatok arculatának új dimenziója”,  
diplomaterv, BKE, 2003.
- [11] G. Németh,  
„Acoustic Company Image and  
Telecommunications Services” (meghívott előadás),  
Proc. of Forum Acousticum 2005, Budapest,  
pp.2633–2637.

# MPEG-4 modell alkalmazása szájmozgás megjelenítésére

TAKÁCS GYÖRGY, TIHANYI ATTILA, BÁRDI TAMÁS,  
FELDHOFFER GERGELY, SRANCSIK BÁLINT

Pázmány Péter Katolikus Egyetem, Információs Technológia Kar  
{takacs.gyorgy, tihanyia, bardit, flugi, sraba}@itk.ppke.hu

Lektorált

**Kulcsszavak:** *audiovizuális beszédfeldolgozás, fej animáció, multimodális kommunikáció*

A cikk áttekinti az MPEG-4 szabványnak a fej és az emberi test alakjának és mozgásával foglalkozó részének kódolási alapelveit. Bemutatja a nyílt forráskódú LUCIA dekódoló modellt jellemzőit és egy speciális alkalmazását. Ebben az alkalmazásban beszédjelből előállított jellemzők vezérlik a fejmodellt, amely siketek számára szolgál kommunikációs segédeszközként. A cikk kitér az alkalmazási kísérletek mérési eredményeire is.

## 1. Bevezetés

Egyre több szempontból felmerül, hogy az emberi beszédátvitel nem csak hangtani jellemzőkre épülő folyamat. A multimodális kommunikáció megközelítése foglalkozik azzal, hogy a beszédinformáció közlése és felfogása nem csak hallható, hanem látható folyamatok együtteséből áll. A szabványosítás elérte ezt a területet is. Az MPEG-4 szabvány része a fej és az emberi test alakjának és mozgásának kódolása. Kidolgoztunk egy speciális alkalmazást, amely végletesen használja a beszéd folyamat kettős természetét, azaz magából a hallható beszédjelből származtat látható beszéd-folyamatot és ezzel siketek számára ad egy kommunikációs segédeszközt.

Ebben a cikkben nem a segédeszköz felépítését és működésének részleteit ismertetjük, hanem az MPEG-4 szabványon alapuló dekódoló rész technikai részleteit taglaljuk. Ehhez a 3. szakaszban a szabvány érdekesebb részeit ismertetjük és értékeljük, ami előkészíti a következő részt, melyben a nyílt forráskódú MPEG-4 kompatibilis LUCIA modell részletes ismertetése következik. Azok a finom részletek kaptak nagyobb figyelmet, amelyek a szájról olvasás szempontjából kritikusak. Az alkalmazás hatásosságát mérési eredmények támasztják alá az 5. szakaszban kifejtettek szerint.

Folyamatos beszédjelből mozgókép folyamatot hozunk létre. Ez egy olyan transzformáció, amelynek lényegi részét egy neurális hálózat hajtja végre. A neurális hálózat komplexitását korlátok között kellett tartani, ezért elengedhetetlen volt az emberi beszéd folyamat lényegét jól megragadó, tömör és hatékony leírása a vizuális beszédnek.

A neurális hálót előfeldolgozott hangadatokkal tanítottunk és képi koordinátákon vártunk a kimeneteken. Főkomponens analízist alkalmaztunk a képi koordináták tömör reprezentálására. Így mindössze 6 kimeneti jellemző kisebb, mint 2% hibával leírta a szükséges képi koordinátákat. A rendszer fejlesztésében külön kezelt probléma volt a mozgókép megjelenítés modellje.

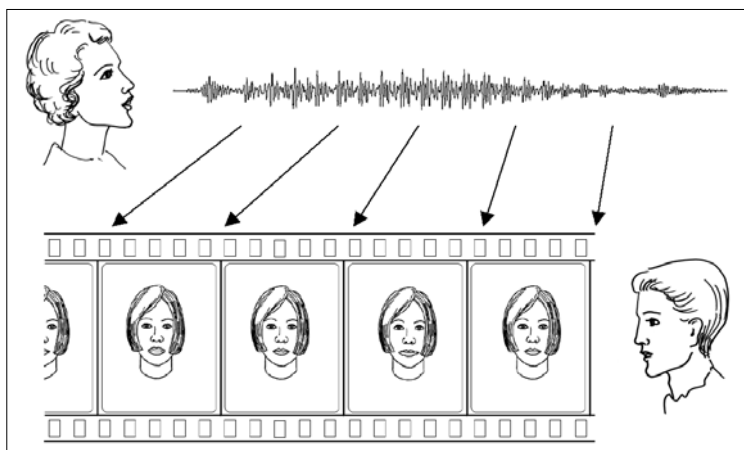
## 3. Az MPEG-4 szabvány fejmozgások tömörített kódolására

Az MPEG (Moving Picture Expert Group) szabványok fő célja a hang és videó jelek tömörítése. A tömörítés alapvető követelményei a hatékonyság és élethűség. A multimédia-alkalmazásokban elterjedt az MPEG-2 kódolás. Az ezt meghaladó MPEG-4 kódolás is ígéretes jövő előtt áll, ugyanakkor céljainkat közvetlenül támogatja. Az MPEG-4-et nem csak nagy tömörítésre alakí-

## 2. Előzmények

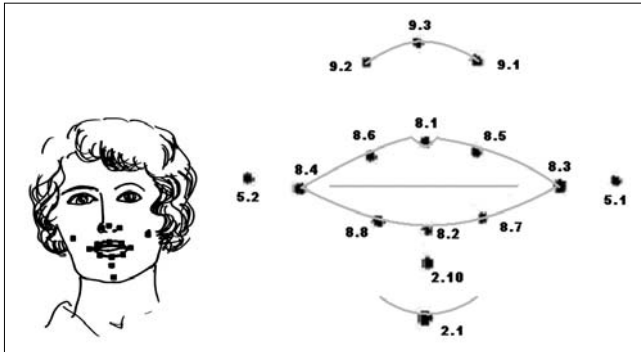
Egy teljes rendszert dolgoztunk ki, amely alkalmas arra, hogy beszédjelből mozgó száj képét állítsa elő. A mozgó szájról a siketek képesek a beszédet leolvasni. A rendszer ismertetése ugyanezen folyóirat számban megtalálható [3]. Itt azokat a részleteket és általános megfontolásokat taglaljuk, amelyek kifejezetten a megjelenítő egységre vonatkoznak.

1. ábra  
Mozgó száj előállítási vázlat



tották ki, hanem figyelembe vettek olyan multimédia alkalmazásokat is, mint a 3D-s jelenetek, animációk, szintetizált hangok, képek, szövegek, grafikák külön vagy akár együttes kezelése és élethű megjelenítése.

Az MPEG-4 szabvány egyik legösszetettebb része a fej és az emberi test megjelenítése és mozgatása (Face and Body Animation, FBA). Az FBA-ra vonatkozó szabványrész leírja az arc és a test alakjának és mozgásának kódolási alapelveit. Az FBA egyik legfontosabb tulajdonsága tehát, hogy nem adja meg pontosan a kódolási és a dekódolási eljárást, csak a küldött adat formáját és értelmezését.



2. ábra Felhasznált tartópontok

Az MPEG-4 szabvány az arc modelljét az arc normál állapotával írja le, megad több tartópontot (Feature Point, FP) és az arc mozgását leíró paramétereket (Facial Animation Parameter, FAP), melyek lényegében a normál archoz képesti elmozdulást jellemzik (2. ábra). Az elmozdulások méretét és arányát a szabvány szerint mindig az emberi arcra jellemző alapvető méretek alapján fejezi ki. A szakirodalomban ennek elterjedt rövidítése FAPU (Face Animation Parameter Unit – lásd a 6. ábrát). A FAPU-t az arc olyan jellegzetes távolságaiból kell számolni, mint például a szegélyök távolsága vagy a száj szélessége.

A szabványban 84 tartóponttal írják le az arcot. (Az adatbázisunk összeállításánál mi 15 FP-t használtunk a száj és környékének leírására).

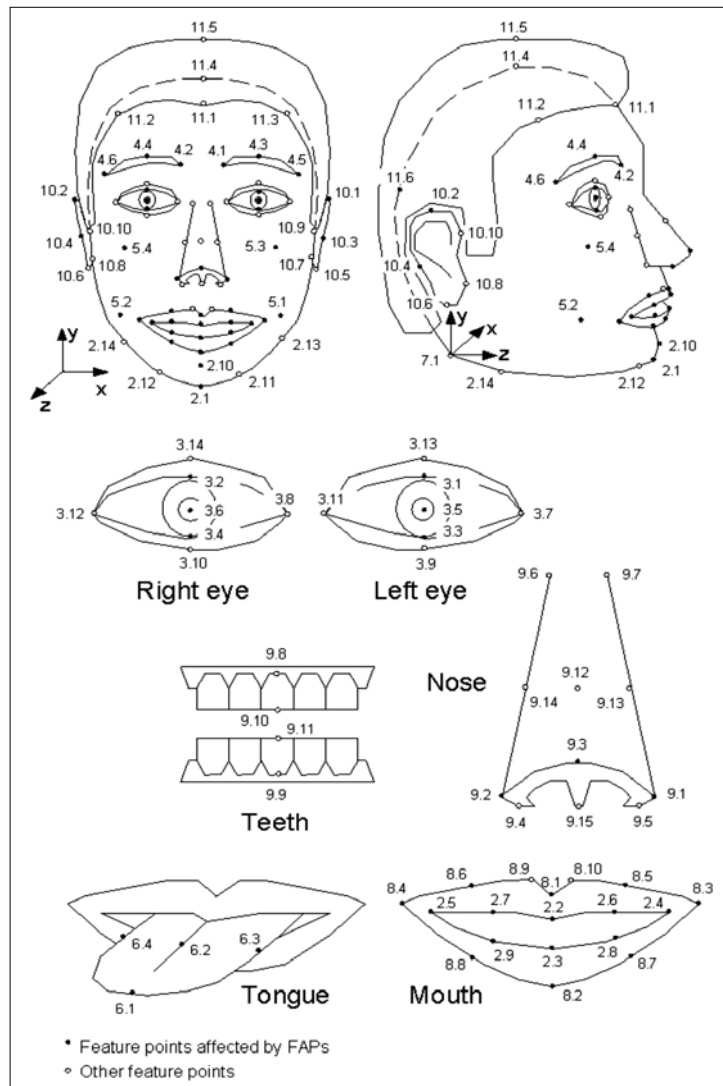
A tartópontok fő feladata, hogy referencia-ként szolgáljanak a FAP-ok számára. A FAP-ok által leírt összetett mozgások mindig a normál tartópontok által leírt fejre vonatkoznak. A normál fej csukott száját és semleges arc kifejezést jelent. Vannak olyan FP-k is, melyekre egy FAP sincs közvetlen hatással (például az orr szélei). Ezeket mindössze az arc alakjának meghatározására használják. Az FP-eket minden MPEG-4 kompatibilis modellen a 3. ábra alapján kell elhelyezni.

FAP-ból a szabvány 68-at különböztet meg, melyet 10 csoportba sorol az alapján, hogy az arc mely részét mozgatja.

3. ábra  
A tartópontok szabványos elhelyezkedése a fején

Az első két FAP magas szintű paraméter. Ez azt jelenti, hogy ezekkel előre beállított komplexebb mozgást lehet kivitelezni. Az első FAP egy megadott vizéma szerinti megjelenést határoz meg. A vizéma a fonéma képi megfelelője. A második FAP a hat alap érzelm megjelenítésére szolgál, úgy mint öröm, bánat, harag, félelem, undor és meglepetés. Tovább érzelmkifejezéseket az alap érzelmek keveréséből lehet megjeleníteni.

A többi FAP alacsony szintű. Ezek abban különböznek a magas szintű FAP-októl, hogy itt a mozgás irányát és amplitúdóját kell megadni, nem pedig egy összetett feladatra előre összeszerkesztett mozgásvezérlést kell kezdeményezni. Az alacsony szintű FAP-ok általában egy-két tartópontot mozgatnak. Előfordul olyan FAP is, amely az összes FP-t mozgatja, ilyen például a fej forgatása. Az alacsony szintű FAP-oknál a szabvány meghatározza, hogy a mi a hozzá illő FAPU, amiből a mozgás a mérték alapja. A FAP előjele a tartópont mozgásirányára vonatkozó információt hordoz, például a száj nyitására vonatkozó paraméterek pozitív, a zárásra vonatkozóak negatív előjelűek. Ez független attól is, hogy a tartópont a száj alsó vagy felső részéhez tartozik. A mozgatás lehet eltolás, forgatás vagy skálázás.

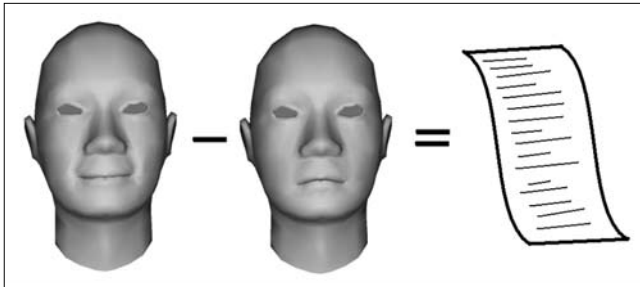


#### 4. A LUCIA modell

A legtöbb modell, legyen az két- vagy háromdimenziós, hálóból áll. A háló (mesh) több egymáshoz illeszkedő nem feltétlenül egy síkban levő sokszöget tartalmazó felület. A hálóban a csúcspontok koordinátáin kívül a lapok, az élek és a csúcsok illeszkedési viszonyait is nyilván kell tartani [12].

A modell felületi jellemzői, textúrája erre a rácsra van ráhúzva. Ahogy mozgatjuk a háló csúcspontjait, úgy mozog vele a textúra is. Ám az MPEG-4 szabványban csak az FP-k mozgatására van mód, az egyes hálókérra közvetlenül nincs. Egy modell tetszőleges számú és finomságú hálóból állhat, a szabvány erre nem terjed ki. Minden MPEG-4 kompatibilis fejmodell azonban azonos tartópont rendszerre épül. A hálók mozgása a tartópontok mozgatásával történik.

A LUCIA modellt Cosi vezetésével olasz kutatók fejlesztették ki [1]. Ez egy nyílt forráskódú mozgó fejmodell. A LUCIA egy MPEG-4 megvalósítás, ami alkalmas vizémák és érzelmi állapotok FAP paraméter alapú közvetlen megjelenítésére. Az MPEG-4 modell tömörítést kifejtő (decompress) része egy grafikus modell mozgási feladat, alapvetően az 5. ábra szemléltetése szerinti információk felhasználásával képes átvinni a mozgás jellegzetességeit. A szabványosított eljárás során az alaphelyzetű fej teljes képének meghatározása és vevő oldalra történő átvitele valósul meg, és a továbbiakban csak az alaphelyzettől történő eltérések átvitelére van szükség a tömörített adatközlés során.

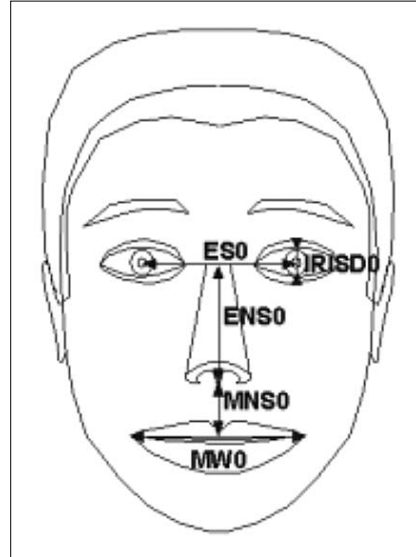


4. ábra  
Az MPEG-4 rendszerű tömörítés koncepciója

Az MPEG-4 tömörítési folyamat (4. ábra) azon az elven működik, hogy a tömörítendő mosolygós fej lényeges paramétereinek valamint az alaphelyzetű fej paramétereinek különbségéből meghatározza a tömörített jellemzőket. Az MPEG-4 koncepció szerint ez a jellemzősor a fej alakjától és környezetétől független adatokat tartalmaz.

A visszaállítási folyamat (5. ábra) során a tömörített jellemzőkhöz, amely jelenleg a mosolygós adatait tartalmazza „hozzáadva” egy tetszőleges alaphelyzetű fej paramétereit, egy mosolygós fej képét kapjuk. Az alaphelyzetű fej meghatározó adatai között kell elhelyezni a felületi jellemzőket valamint az esetleges további adatokat, mint például a modell haja, szeme stb. A visszaállítás során kell létrehozni a felületeket azok megvilágítástól függő színezésével együtt [2].

Az MPEG-4-ben a tömörítés során meghatározott és felhasznált távolság mértékrendszer (6. ábra) lehetőséget biztosít arra, hogy a tömörített információ felhasználásával tetszőleges más alaphelyzetű fejre lehessen alkalmazni a visszaállítást, és így lehessen változtatni a visszaállítás folyamatát.



6. ábra  
Az emberi arcra jellemző méretek

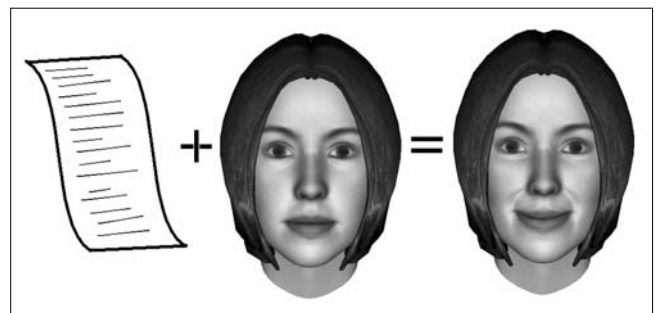
ESO= szemgolyók távolsága;  
IRISD0= az írisz átmérője;  
ENS0= az orr hossza;  
MNS0= orr és a száj távolsága;  
MW0= a száj szélessége

Az ESO; IRISD0; ENS0; MNS0; MW0; távolságok határozzák meg az adott arcberendezésen alkalmazandó távolságegységek halmazát. A távolságmérésnek ez a módszere biztosítja azt a lehetőséget, hogy a visszaállítás során az eredetitől jelentősen eltérő felépítésű alaphelyzetben álló fejre is visszaállíthatók legyenek a tömörített információk.

Az MPEG-4 szabványnak köszönhetően az arc mozgásához nem kell képkockáról képkockára megadni a videó minden egyes pixelét, mindössze a mozgatott FP-khez tartozó FAP-okat kell továbbítani. Ennek köszönhetően igen alacsony sáv szélességen keresztül is elérhető a real-time arcanimáció.

Az MPEG-4 szabvány előnyeit leginkább internetes alkalmazásokban használják. Találkozhatunk olyan rendszerrel, mely az e-mail-eket alakítja át olyan videóvá, ahol az általunk kiválasztott személy mondja el az üzenetet. Léteznek olyan alkalmazások, melyek internetes áruházakban „eladókat” alkalmaznak, vagyis egy MPEG-4 szabványú modell ad segítséget az árákról, a minőségről vagy éppen a készletről.

5. ábra  
Az MPEG-4 rendszerű visszaállítás koncepciója



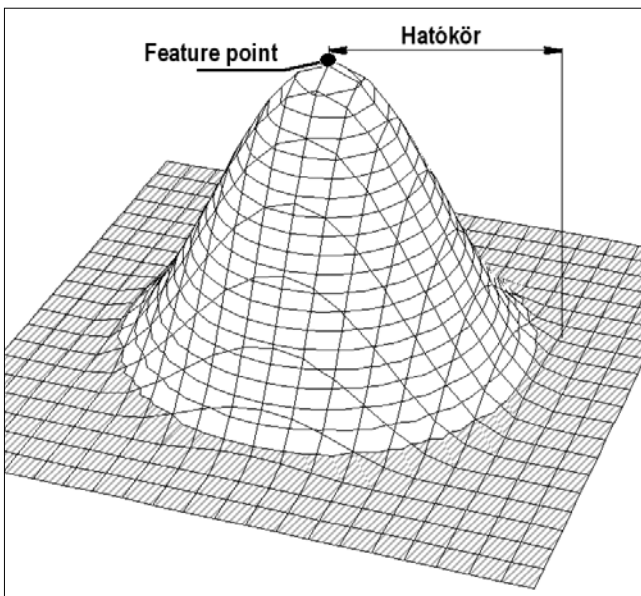


A szintetikusan létrehozott szájmogzás megjelenítésére felhasznált LUCIA modell egy szokásos 3D grafikus modell, amely animálható és így a céljaink megvalósítására alkalmas.

Az adatbázis felvételénél az arcra felfestett pontok kis mértékben eltérhetnek a szabványban előírt tartópontok helyétől. Ezt a hibát úgy korrigáltuk, hogy a tartópontokat ráillesztettük a felfestett pontokra úgy, hogy szintetizáláskor egybeessenek.

Az 7. ábra bemutatja egy eredetileg vízszintes elhelyezkedő négyzögháló felhasználásával készített animáló eljárás hatását abban az egyszerű esetben, ha az eredeti helyzetből függőleges irányban felfelé kívánjuk elmozdítani a síknak egyetlen pontját.

7. ábra  
Az FP függőleges elmozdításának hatása a vízszintes felületre



A hatókörökben összeérő, egymás mellé eső pontok egymásra hatását megfelelő súlyozással kell kiküszöbölni. Elképzelhető, hogy egy hálórészt több tartópont is mozgatni akar. Ilyenkor természetesen súlyozottan összegződnek az elmozdulások. A súlyozás meghatározásánál az elmozdítást eredményező pont hatását annak távolságával fordított arányban határozzuk meg, ez a módszer azt eredményezi, hogy a modell rácspontjainak elmozdulását a FP-hez közeli rácspontok esetén nagymértékben az FP helyzete határozza meg. A vázolt eljárással lehetséges kijelölt pontok és hozzájuk tartozó területek rögzítése. Ilyen technikával oldottuk meg a 3D-s LUCIA fej állának mozgását.

Annak érdekében, hogy az állcsont a megfelelő forgáspont körül elforduljon, az állcsúcsot (2.1-es FP) mozgtattuk. Az állcsont miatt nagy hatókörrel kell a 2.1-es FP-t mozgatni, aminek az a hatása, hogy szemből nézve úgy tűnik, mintha a fej egész álla leesne. A jelenséget meg lehet szüntetni oly módon, hogy az arc körvonalához tartozó 2.13 és 2.14-es FP-t minden irányban 0-val mozdítjuk el, ennek hatására a 2.13 és

2.14-ös FP-k nagy súllyal helyben tartják az arc körvonalát és csak elenyésző mértékben mozdul a környezetük a 2.1 és 2.10 pontok mozgatásának hatására. Az alkalmazott technika teljesen kiküszöböli az áll leesésének a jelenségét.

A LUCIA modell tartalmazza az alsó és felső fogsort valamint a nyelvet is. Az alsó fogsor mozgatását kizárólagosan az állcsúcs mozgása határozza meg, a felső fogsor mozgatását az orr megfelelő pontjaihoz kötöttük, így annak elmozdulása minimális, hiszen az orr középpontját tekintettük a munka során referenciának. A nyelv mozgatásával a projekt nem foglalkozott.

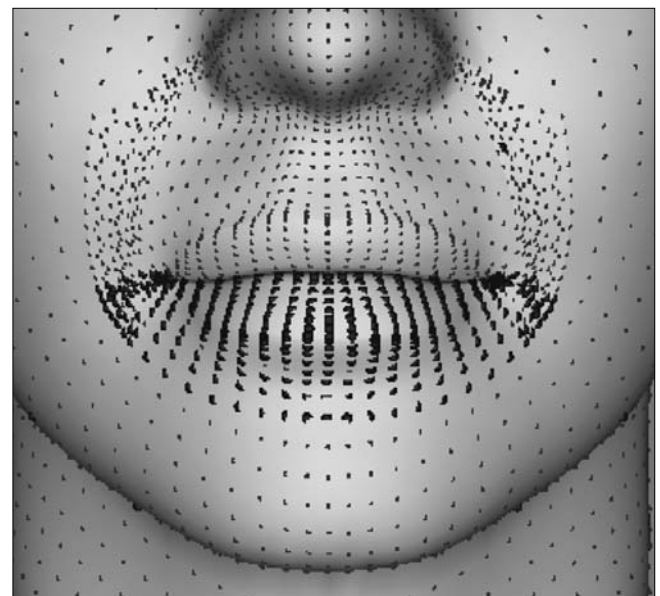
A mozgatandó felületen a háló törése, szakadása (például szem, száj) azt a problémát jelenti, hogy a szakadási vonalnál tovább azon átnyúlva nem alkalmazhatjuk az előzőekben vázolt módszert. Például az alsó ajak mozgatása nem húzza magával a felső ajkak hálórészét, pedig azok a hatókörön belül esnek. Ezzel a módszerrel kezelhető a száj, a szemek természetes nyitása.

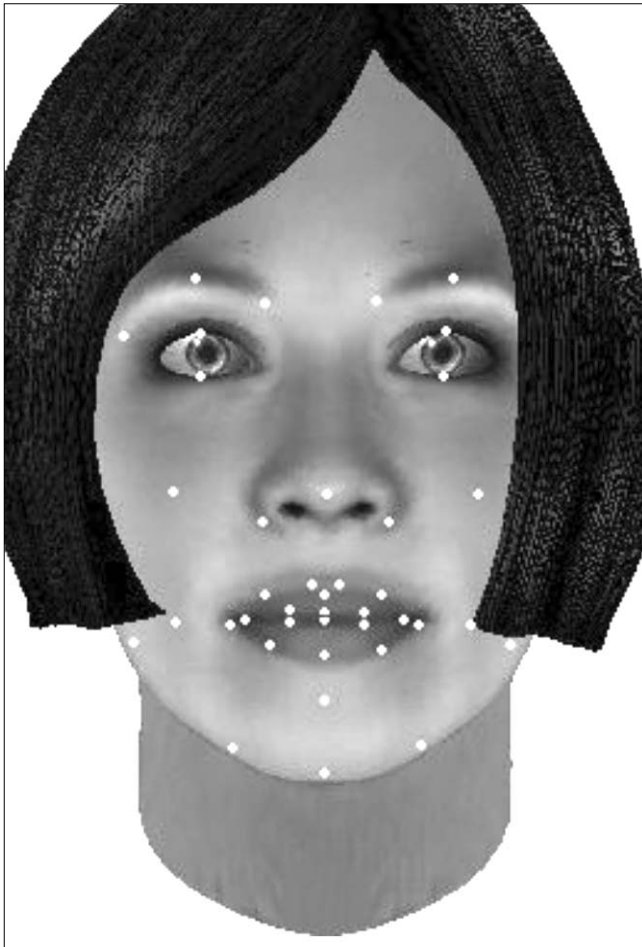
Azt a módszert választottuk, hogy minden mozgatott FP-hez meghatároztuk a modellünk egy-egy háló csúcspontokkal leírt egybefüggő részét. Ez jelentősen gyorsítja a mozgó algoritmusokat, mivel nem kell a teljes fej összes rácspontjának távolságát meghatározni minden egyes FP helyzetétől, hanem elegendő a kijelölt rész halmaz pontjainak a figyelembe vétele a számítások folyamán. Az alsó és felső ajakrész szétválasztását szemlélteti a 8. ábra.

Az ábrán sötétebb pontok jelölik a száj alsó szélét. Ezekre a pontokra hatnak, ezeket mozgtatják a 8.2; 8.7; 8.8 tartópontok (lásd 3. ábra).

Minden FP-hez tartozik egy mozgatási hatókör – egy gömb alakú térrész – és azon a hatókörön belül levő rácspontok elmozdulását határozza meg az adott FP elmozdulása az MPEG-4 rendszerben meghatározott skálázás szerint.

8. ábra  
Az alsó szájszélet meghatározó hálópontok





9. ábra LUCIA modellen alkalmazott FP-k

A 3D grafikus modellt az MPEG-4 rendszernek megfelelően ki kell egészíteni a 3 dimenzióban értelmezett FP-vel, és azok hatókörének meghatározásával, valamint az egyes FP-k által mozgatható rácpontok halmozásával (9. ábra).

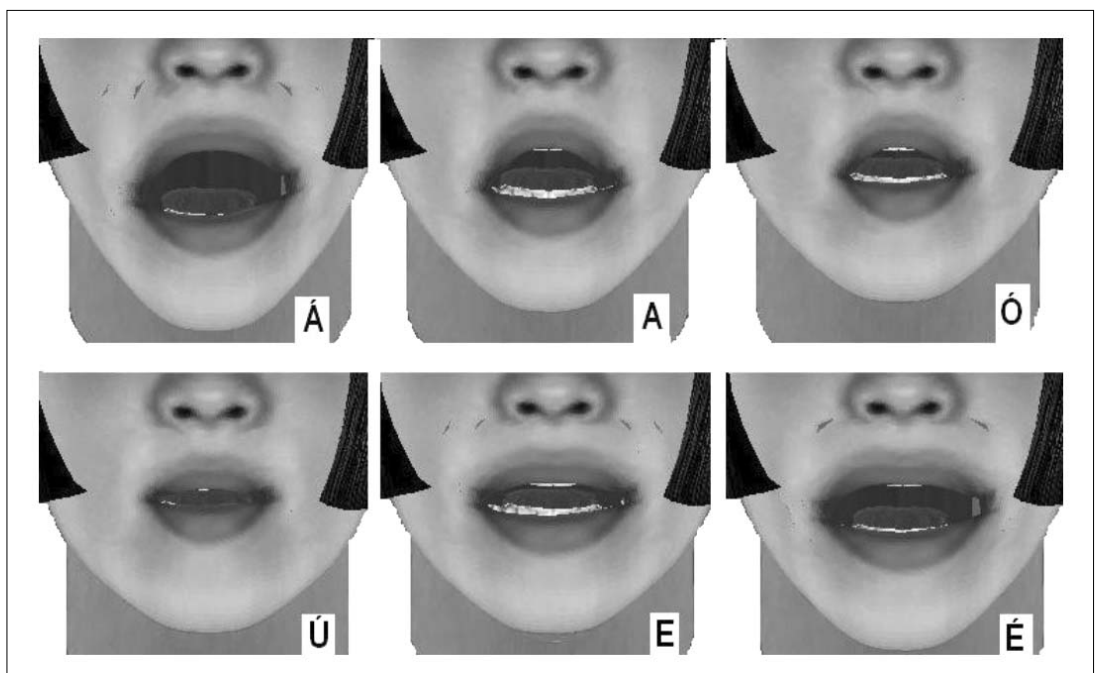
A beszédjelből származó mozgást előállító projekt során az előzőekben részletezett módon kialakított LUCIA modellt alkalmaztuk. A projekt eredeti elképzelései szerint a megvalósításkor a beszédjelből közvetlenül az FP mozgató paramétereket állítottunk elő, tehát nem volt szükség arra, hogy az egyes vizémákat külön-külön meghatározzuk és előállítsuk, de a hosszan kitarított magánhangzók tiszta fázisainál jól megkülönböztethető szájállásokat hozott létre a fejmodell (10. ábra).

## 5. MÉRÉSI EREDMÉNYEK, KÖVETKEZTETÉSEK

Az animációs rendszerünk komponenseinek ellenőrzésére szájról olvasási kísérleteket végeztünk sikertest-talányokkal. A szájról olvasási feladatok nehézségét úgy állítottuk be, hogy körülbelül 95% és 100% közötti felismerési arányt kapjunk a vetített eredeti videó felvételekre, hogy referenciaként szolgálhasson.

Ilyen jó arányt az előzetes kísérletek leírásánál [3] már ismertetett módon a felismerendő szövegben használt szókinccs és nyelvtan erős szűkítésével, valamint egy jól artikuláló jeltolmács szerepeltetésével értünk el. Ezután mértük a felismerési arányt, úgy, hogy a videófelvétel helyett az animált beszélőfej-modell volt látható, ugyanakkor minden más kísérleti körülményt változtatlanul hagytunk.

A fejmodellre való áttérés két lépcsőben történt. Az elsőben a felvételeken festékpöttyel megjelölt MPEG-4 pontok koordinátáit igyekeztünk átvinni a modellre: vagyis a fejmodell vázát képező háló megfelelő csomópontjait minden képkockán a felvételen mért koordinátájú pozíciókba mozgattuk. Ezzel azt kívántuk elérni, hogy a modell közvetlenül utánozza a jeltolmács artikulációját, ebben a lépésben a hang még nem játszott szerepet.



10. ábra  
Példák  
magyar nyelvű  
jellegzetes  
magánhangzó  
szájállásokra  
(viziémákra)

A második lépcsőben a beszédhang alapján számított koordináták szerint vezéreltük a fejmodellt. Itt már csak a hangbemenetre volt szükség az animáció előállításához [3].

A kísérlet során a felismerési arányok a következők szerint alakultak:

- eredeti felvételek (referencia):  
97,1%;
- animáció a jeltolmácsra festett tartópontok alapján vezérelt modellnél (1. lépcső):  
54,9%;
- animáció a hang alapján (2. lépcső):  
47,9%.

Jelen cikk szempontjából a felvételekről a LUCIA modellre való áttérés, vagyis az első lépcső érdekes. Itt elég jelentős romlás tapasztalható a felismerési arányban, ennek lehetséges (valószínű) okaira térünk ki röviden.

Megállapítható, hogy az általunk kiválasztott és a felvételeken megjelölt MPEG-4 pontok helyzete hiányosan (információ veszteséggel) reprezentálja azokat a látható beszédképzési jellemzőket, melyek a szájról olvasásban szerepet játszanak. A kísérletek után minden alkalommal kikértük a résztvevő siketek véleményét, hogy mely tényezők gátolták őket leginkább a szájról olvasásban.

A felvételek és az animációk között talán a legfontosabb különbség, hogy a fejmodellnek nincs nyelve. De ha a LUCIA modell lehetővé tenné a nyelv animálását, akkor is problémát jelentene, hogy nincsenek referencia adataink a nyelv pillanatnyi helyzetéről, nem tudjuk, hogyan is kéne mozgatni. A nyelvre a felvételeken nem festhettünk pontot. A nyelv hiányában például a *kilenc* vagy a *nulla* szavak felismerése gondot okozott az animáció esetében, míg a felvételeken jól látható volt a nyelv főntről lecsapódása az /hang után, így valamennyi tesztalanyunk könnyedén felismerte azokat.

A másik problémánk volt, hogy a felvételekhez csak az ajkak külső kontúrján tudunk pontokat megjelölni, beljebb nem. Ezek viszont az ajakkerekítésről kevés információt tartalmaznak. Az animációkon elsősorban az ajakkerekítéses magánhangzók (pl. *u*, *ü*) voltak kifogásolhatók. Szintén a pontok elhelyezésére vezethető vissza, hogy nincs elegendő információnk a fogak láthatóságáról. Pedig elsősorban ettől függ az ajkakon belüli terület világossága, ami egy igen karakteres és könnyen észlelhető vizuális jellemző [4].

Az MPEG-4 szabvány eredeti célja egy olyan modell megalkotása, aminek segítségével tömöríteni, majd rekonstruálni lehet mozgó fej adatokat. Munkánk során megoldottuk, hogy a szabványra építve olyan minőségben mozgatható a száj és környezete, hogy ennek alapján a siketek a beszédet képesek szájról leolvasni.

Fontos eredménynek tartjuk azt is, hogy az animáció működik akkor is, ha nem képpontok mintavételezése alapján származtattuk a tartópont paramétereiket, hanem beszédjelből számoltuk. Az eredményeink azt mutatják, hogy igen kis különbség van a mintavétele-

zéssel vezérelt arc, és a beszédjel alapján vezérelt arcmodell felismerhetősége között.

További fejlesztést igényel a fejmodell finomítása. A száj külső körvonalán túl a belső kontúr, fogak vagy nyelv láthatósága tűnik a továbblépés első lehetőségeinek.

### Köszönetnyilvánítás

A szerzők ezúton is kifejezik köszönetüket a Nemzeti Kutatási és Technológiai Hivatalnak a 472/04 szerződés keretében nyújtott támogatásáért.

### Irodalom

- [1] Cosi P., Fusaro A., Tisato G.,  
"LUCIA: a New Italian Talking-Head Based on a Modified Cohen-Massaro's Labial Coarticulation Model",  
Proc. of Eurospeech 2003,  
Geneva, Switzerland, September 1, 2003, Vol. III,  
pp.2269–2272.
- [2] Szirmai-Kalos László, Antal György, Csonka Ferenc,  
„Háromdimenziós grafika animáció és játékfejlesztés”,  
ComputerBook Kiadó Kft., Budapest 2003.
- [3] Takács György, Tihanyi Attila, Bárdi Tamás,  
Feldhoffer Gergely, Srancsik Bálint:  
„Beszédjel átalakítása mozgó száj képévé  
siketek kommunikációjának segítésére”  
Híradástechnika 2006/3,  
pp.31–37.
- [4] László Czap, János Mátyás,  
„Virtual Speaker”  
Híradástechnika – Selected Papers 2005/6,  
pp.2–5.
- [4] I. Pandzic, R. Forchheimer,  
„MPEG-4 Facial Animation:  
The Standard, Implementation and Applications”,  
Wiley, 2002.

# Beszédjelek pillanatnyi jellemzőinek becslése a Teager-operátorral és a Hilbert-Huang-transzformációval

PINTÉR ISTVÁN

Kecskeméti Főiskola GAMF Kar, Automatizálási és Alkalmazott Informatikai Tanszék  
pinter.istvan@gamf.kefo.hu

Lektorált

**Kulcsszavak:** Teager-operátor, HHT, pillanatnyi amplitúdó és frekvencia, visszaállítás pillanatnyi jellemzőkből

*A beszédjelek finomszerkezetének vizsgálatához a nemlineáris és nemstacionárius jellemzők meghatározására szolgáló módszerek szükségesek. Jelen dolgozatban a Teager-operátort és a Hilbert-Huang-transzformációt (HHT) ismertetjük, mint a pillanatnyi amplitúdó és a pillanatnyi frekvencia becslésére alkalmazható jelfeldolgozási eljárást. A HHT-vel előállítható pillanatnyi amplitúdó és pillanatnyi frekvencia paramétereket összehasonlítjuk a Teager-operátorra alapozott becslések eredményeivel mind vizsgálójel, mind beszédjel esetén.*

## 1. Bevezetés

A gépi beszédfeldolgozásban számos feladat megoldásának alapja az úgynevezett kvázi-stacionárius jelmodell. Eszerint a beszédjel feldolgozható úgy, hogy elegendően rövid időtartamú szakaszok egymást időben átfedő sorozatainak végezzük az adott feladat megoldása érdekében számításainkat. Feltételezzük, hogy a beszédszakasz időtartama alatt a beszédjel-modell paraméterei nem változnak. Az elegendően rövid időtartamot a hangszalagok nyitási-zárási ütemének megfelelő alapperiódus-idő 2...5-szöröseként határozza meg a szakirodalom, az átfedési idő 1...3 ugyanebben az időegységben [1].

A gépi beszédfeldolgozás fejlődése során felmerült az igény olyan elemző módszerek iránt, amelyekkel az alapperiódus időtartamánál rövidebb idő alatt lejátszódo változások is vizsgálhatók. Az ilyen változások alkotják a beszédjel finomszerkezetét. A nemlineáris módusú hangszalag-rezgés okozta kismértékű alapperiódus-idő ingadozás jelensége – sok egyéb mellett – olyan jelenség, aminek vizsgálatához a finomszerkezet leírására alkalmas módszerek szükségesek. A módszerekkel szembeni elvárás az, hogy néhány beszédmintányi adathoz tudjanak fizikailag is értelmezhető jellemzőket rendelni. Következésképpen erre a célra nem használható a kvázi-stacionárius jelmodell alapján kidolgozott gépi beszédfeldolgozási eszköztár [2].

A probléma lényegét tömören összefoglalva azt mondhatjuk, hogy az időfelbontás növelése a részletes frekvenciakép megtartása mellett nem lehetséges, mert fennáll az időpont és a frekvenciaérték együttes meghatározásának bizonytalanságát összekapcsoló Gábor Dénes-féle határozatlansági reláció, ezáltal a gördülő Fourier-transzformációra (STFT, Short-Time Fourier Transform) alapozott – vagy azzal kapcsolatba hozható – módszerek a beszédjel finomszerkezetének leírására nem alkalmasak.

Ma már elterjedt a megnövelt időfelbontást igénylő alkalmazásokban a wavelet-transzformáció használata,

de a wavelet-es beszédelemzés időfelbontását is korlátozza az, hogy a fentebb említett idő-frekvencia bizonytalanság helyére az idő-skála bizonytalanság lép.

Van olyan beszédábrázolás is, amelynél nincs jelen a határozatlansági reláció okozta korlát – ilyen például a Wigner-Ville-eloszlás vagy a Choi-Williams-eloszlás, ám itt más problémák jelentkeznek a finomszerkezet feltárásakor (például a transzformáltban megjelenő kereszt-tag elnyomása jelent megoldandó feladatot). Ezzel az izgalmas témakörrel jelen dolgozatban nem foglalkozunk, a részleteket [2,3] tartalmazza.

A beszédjel finomszerkezetének elemzésére szolgáló – az előző bekezdésben említettektől lényegesen eltérő – módszer a Teager-operátorra alapozott ES-algoritmus (Energy Separation algorithm) [2], amivel becslés adható a beszédjel pillanatnyi amplitúdójára és pillanatnyi frekvenciájára. A cikk hátralévő részében ezeket együtt pillanatnyi jellemzőknek nevezzük. Az előző bekezdésben foglaltakat is figyelembe véve talán nem meglepő, hogy a wavelet-es elemzés és a Teager-operátor összekapcsolása mára sikeres alkalmazásokhoz vezetett [4].

További lehetőség a pillanatnyi jellemzők meghatározására a Hilbert-Huang transzformáció [5] alkalmazása. Mivel a Teager-operátorra illetve a HHT-re alapozott módszerek összehasonlításáról a számunkra hozzáférhető – nyomtatásban, illetve elektronikusan megjelenített – beszédfeldolgozási szakirodalomban nem találunk közölt eredményeket, jelen dolgozatunk témájának ezt választottuk.

## 2. A Teager-operátor és az ES-algoritmus

### 2.1. A folytonos idejű Teager-operátor és a pillanatnyi jellemzők becslése

A címben szereplő operátor fogalmának megalkotása és a vele elvégzendő művelet meghatározása az emberi beszédkeltés közben fellépő nemlineáris fizikai jelenségek gondos vizsgálata után vált lehetségessé.

H. M. Teager és S. M. Teager először 1980-ban közöltek ilyen mérési eredményeket, majd 1985-ös publikációjukban a modellalkotásról számoltak be. Kiderült, hogy az alapperiódus-időn belüli gyors jelenergia-változás jelenségének leírásához célszerű meghatározni a jelet előállító rendszer összenergiáját. Ezen összenergia becslését kapjuk meg, ha a jelre egy alkalmasan megválasztott operátor hat – ma ezt az operátort Teager-operátornak nevezzük. A részleteket és a bőséges szakirodalmi forrást [2]-ben találhatjuk meg.

Azt, hogy miként lehet egy rendszer által előállított jelből a rendszer összenergiájára következtetni, a rugóra függesztett test harmonikus rezgőmozgásának példáján szokás bemutatni. Ezt a mozgást másodrendű differenciál-egyenlet írja le, ami ideális esetben a következő alakú:

$$\frac{d^2x(t)}{dt^2} + \frac{k}{m} \cdot x(t) = 0,$$

ahol  $x(t)$  a kitérés-idő függvény,  $k$  a rugóállandó,  $m$  a harmonikus rezgőmozgást végző test tömege. A differenciál-egyenlet megoldása  $x(t) = a \cdot \cos(\omega \cdot t + \varphi)$  alakú – a fentebbi szóhasználat szerint ez a rendszer által előállított jel. A rugóból és a harmonikus rezgőmozgást végző testből álló rendszer összenergiája a rugóban tárolt energia és a mozgási energia összege:

$$E = \frac{1}{2} \cdot k \cdot x^2(t) + \frac{1}{2} \cdot m \cdot \left( \frac{dx(t)}{dt} \right)^2.$$

Behelyettesítés után adódik, hogy  $E = m \cdot a^2 \cdot \omega^2$ , ezáltal ha a kitérés-idő függvényből mérésrel meghatározuk az amplitúdót és a körfrekvenciát, akkor ezek szorzatának négyzete arányos a jelet előállító rendszer összenergiájával. A Kaiser által javasolt általánosítás alapja az, hogy – egy állandó szorzótényezőtől eltekintve – ugyanezt az eredményt kapjuk, ha a következő operátort alkalmazzuk a kitérés-idő függvényre, mint jelre [2]:

$$\Psi\{x(t)\} = \left( \frac{dx(t)}{dt} \right)^2 - x(t) \cdot \frac{d^2x(t)}{dt^2}, \quad (1)$$

ahol  $\Psi\{\cdot\}$  a Teager-operátor. A fenti kitérés-idő függvényre alkalmazva a következőképpen számolhatunk:

$$\frac{dx(t)}{dt} = -a \cdot \omega \cdot \sin(\omega \cdot t + \varphi), \quad (2)$$

$$\frac{d^2x(t)}{dt^2} = -a \cdot \omega^2 \cdot \cos(\omega \cdot t + \varphi),$$

$$\Psi\{x(t)\} = a^2 \cdot \omega^2 \quad \text{amivel} \quad (3)$$

adódik. Ellenőrizhető, hogy ugyanezt az eredményt kapjuk, ha az operátort az  $x(t) = a \cdot \sin(\omega \cdot t + \varphi)$  jelre alkalmazzuk – amint az várható is. Érdekességképpen megemlítjük még, hogy fennáll:

$$\Psi\{a \cdot e^{j(\omega t + \varphi)}\} = 0. \quad (4)$$

Az  $x(t) = a \cdot \cos(\omega \cdot t + \varphi)$  jel egy lehetséges általánosítása az, amikor mind az amplitúdó, mind a fázis időfüggő, az így keletkező AM-FM jel alakja:

$$x(t) = a(t) \cdot \cos(\varphi(t)). \quad (5)$$

Közvetlen számolással hamar belátható, hogy tetszőleges amplitúdó- és fázis időfüggvény esetén az (1)-ben megadott operátor nehezen kezelhető kifejezéshez vezet. Ám abban az esetben, ha mind az amplitúdó, mind a fázis lassan változik az időben, vagyis amikor fennállnak az alábbi közelítések:

$$\frac{da(t)}{dt} \approx 0, \quad \frac{d\varphi(t)}{dt} \approx \text{állandó}, \quad \frac{d^2\varphi(t)}{dt^2} \approx 0, \quad (6)$$

akkor az (5)-beli AM-FM jelre alkalmazva a Teager-operátort, a következőket kapjuk:

$$\begin{aligned} \frac{dx(t)}{dt} &\approx -a \cdot \frac{d\varphi(t)}{dt} \cdot \sin \varphi(t), \\ \frac{d^2x(t)}{dt^2} &\approx -a \cdot \left( \frac{d\varphi(t)}{dt} \right)^2 \cdot \cos \varphi(t), \\ \Psi\{x(t)\} &\approx a^2(t) \cdot \left( \frac{d\varphi(t)}{dt} \right)^2. \end{aligned} \quad (7)$$

Alkalmazhatjuk az operátort a jel deriváltjára is, ekkor:

$$\Psi\left\{ \frac{dx(t)}{dt} \right\} = \left( \frac{d^2x(t)}{dt^2} \right)^2 - \frac{dx(t)}{dt} \cdot \frac{d^3x(t)}{dt^3}. \quad (8)$$

A (6)-beli közelítéseket figyelembe véve az (5)-ben szereplő AM-FM jelre a részletes számítás után adódik, hogy:

$$\Psi\left\{ \frac{dx(t)}{dt} \right\} \approx a^2(t) \cdot \left( \frac{d\varphi(t)}{dt} \right)^4. \quad (9)$$

A kapott közelítések segítségével becslést adhatunk az amplitúdó abszolút értékére, mivel fennáll:

$$\frac{\Psi\{x(t)\}}{\sqrt{\Psi\left\{ \frac{dx(t)}{dt} \right\}}} = |a(t)|, \quad (10)$$

valamint a fázis deriváltjának (a pillanatnyi frekvenciának) abszolút értékére:

$$\sqrt{\frac{\Psi\left\{ \frac{dx(t)}{dt} \right\}}{\Psi\{x(t)\}}} = \left| \frac{d\varphi(t)}{dt} \right|. \quad (11)$$

Az (1), (10) és (11) egyenletekkel tehát a jelből becsülhető az időben lassan változó  $a(t)$  burkoló, és a lassan változó pillanatnyi frekvencia. Ellenőrizhető, hogy az  $x(t) = a \cdot \cos(\omega \cdot t + \varphi)$  jelre ezek a becslések megadják az (állandó) amplitúdó és az (állandó) körfrekvencia értékét.

## 2.2. A diszkrét idejű Teager-operátor és az ES-algoritmus

A gépi számítás alapjául is szolgálhat (1), (10) és (11) megfelelő mintavételezés valamint a differenciálás alkalmas diszkrét közelítése után. Mint numerikus eredményeink mutatják, ez utóbbi célra a Savitzky-Golay-féle 5 pontos simító deriválási algoritmus [6] megfelelő. Ezt a továbbiakban közvetlen számításnak nevezzük.

A diszkrét idejű Teager-operátort a folytonos idejű Teager-operátor (1)-ben megadott alakjából úgy tudjuk származtatni, hogy a differenciálást a  $d(n) = x(n) - x(n-1)$  differenciával közelítjük.

Ezzel a diszkrét idejű Teager-operátor alakja a következő lesz:

$$\Psi_D \{x(n)\} = x^2(n) - x(n-1) \cdot x(n+1). \quad (12)$$

Némi számolás után adódik, hogy  $x(n) = a \cdot \cos(\omega \cdot n + \varphi)$  mintasorozatra alkalmazva a diszkrét idejű Teager-operátort, az eredmény

$$\Psi_D \{x(n)\} = a^2 \cdot \sin^2 \omega, \quad (13)$$

ahol  $\omega$  a digitális körfrekvencia.

A diszkrét idejű Teager-operátor esetén megmutatható, hogy az  $x(n) = a(n) \cdot \cos(\varphi(n))$  mintasorozatból kiindulva a lassan változó pillanatnyi jellemzők becslésére a következő összefüggések érvényesek [2]:

$$a(n) \approx \frac{2 \cdot \Psi_D \{x(n)\}}{\sqrt{\Psi_D \{x(n+1) - x(n-1)\}}}. \quad (14)$$

$$\omega(n) \approx \arcsin \left( \sqrt{\frac{\Psi_D \{x(n+1) - x(n-1)\}}{4 \cdot \Psi_D \{x(n)\}}} \right). \quad (15)$$

A (12), (14) és (15) kifejezésekkel adott számítási eljárást nevezi a szakirodalom ES (Energy Separation) algoritmusnak.

Az ES-algoritmusnak megvan az az előnye, hogy csak három mintát igényel a becslés meghozatalához, míg a közvetlen számítás a simító deriválás miatt öt mintát használ a becsléshez, ám ez utóbbi esetben nem szükséges az arcsin(.) függvény a digitális körfrekvencia értékének meghatározásához.

### 3. A Hilbert-Huang-transzformáció és a pillanatnyi jellemzők számítása

Az előző pontban láttuk, hogy a Teager-operátor alkalmazásával történő pillanatnyi jellemző-számítás meghatározott feltételek mellett lehetséges, amit például alkalmas sávszűrővel biztosíthatunk.

Felmerülhet a kérdés, hogy nincs-e ennél általánosabb módszer a fizikailag is értelmezett pillanatnyi paraméterek – a pillanatnyi frekvencia és pillanatnyi amplitúdó – becslésére? Az igenlő választ Norden E. Huang és munkatársai adták meg 1998-ban közölt dolgozatukban [5]. A cikkükben felvetett egyik első kérdés az, hogy mi jellemzi a fizikailag értelmezhető pillanatnyi frekvenciát? A természetes válasz az, hogy a pillanatnyi frekvencia legyen pozitív valós szám. Ezt követően felmerül, hogy olyan jel esetében, aminek nincs egyenáramú komponense, milyen jelbéli szerkezet az, ami negatív pillanatnyi frekvenciát ad? Ennek ismeretében ugyanis törekedni lehet az ilyen jelszerkezet elkerülésére a pozitív pillanatnyi frekvencia biztosítása érdekében. A szerzők érveléséből kiderül, hogy abban az esetben, ha két egymást követő pozitív helyi maximum között található pozitív helyi minimum, avagy két negatív helyi minimum között található negatív helyi maximum, a pillanatnyi frekvencia negatív lesz.

Tehát a feladat az, hogy a pillanatnyi jellemzők számítása előtt a meglévő mintasorozatból olyan összetevőket kell kinyerni, amelyekre az előző tulajdonság nem teljesül. Ezt követően már sor kerülhet a pillanatnyi jellemzők számítására is. A természetes módusfelbontás (EMD, Empirical Mode Decomposition) nevű algoritmust adták meg ezen összetevők előállítására, amelyeket benső módusfüggvényeknek (IMF, Intrinsic Mode Functions) neveztek el. A felbontást követően már a jellemzők számítására ismert módszerekkel, nevezetesen az egyes benső módusfüggvények kanonikus reprezentációjának segítségével lehet meghatározni a pillanatnyi amplitúdót és a pillanatnyi frekvenciát.

vöket kell kinyerni, amelyekre az előző tulajdonság nem teljesül. Ezt követően már sor kerülhet a pillanatnyi jellemzők számítására is. A természetes módusfelbontás (EMD, Empirical Mode Decomposition) nevű algoritmust adták meg ezen összetevők előállítására, amelyeket benső módusfüggvényeknek (IMF, Intrinsic Mode Functions) neveztek el. A felbontást követően már a jellemzők számítására ismert módszerekkel, nevezetesen az egyes benső módusfüggvények kanonikus reprezentációjának segítségével lehet meghatározni a pillanatnyi amplitúdót és a pillanatnyi frekvenciát.

#### 3.1. A természetes módusfelbontási eljárás és a benső módusfüggvények

A benső módusfüggvények tehát eleget tesznek az előző bekezdésben leírt feltételeknek, aminek következménye, hogy két alapvető tulajdonsággal kell rendelkezzenek [5]:

- a szélsőértékek és a nullaátmenetek száma vagy azonos, vagy eltérésük 1,
- rendre a helyi maximumok és minimumok által kijelölt burkolók középértéke zérus.

A benső módusfüggvények előállítása az [5]-ben közölt algoritmussal történik.

Az algoritmusban főszerepet játszik a leválasztási eljárás (sifting process), mert – szemléletesen szólva – ezzel fejtünk le a jelről rendre egy-egy benső módusfüggvényt. Mindeközben az eredeti jel (adatsor) helyi jellemzőivel kell számolni, így a benső módusfüggvények a jelhez igazítottak lesznek, vagyis az eljárás ebben az értelemben adaptív. A leválasztási eljárás ezen felül olyan, hogy az eredeti jel – egy maradékjeltől eltekintve – a benső módusfüggvények összegzésével állítható elő. A benső módusfüggvények számára [5] nem tartalmaz előírást, így azt többnyire tapasztalati úton kell meghatározni. A leválasztási eljárás után az eredeti valós mintasorozat tehát a következőképpen írható fel:

$$x(n) = r(n) + \sum_{k=0}^{K-1} m_k(n), \quad (16)$$

ahol  $r(n)$  a maradékjel,  $m_k(n)$  a  $k$ -adik benső módusfüggvény.

#### 3.2. A jel kanonikus reprezentációja és a pillanatnyi jellemzők

Gábor Dénes részletes vizsgálatainak [7] eredménye, hogy az  $x(t) = a(t) \cdot \cos(\varphi(t))$  alakú jelmodell, amit az előző pontban használtunk, nem minden esetben egyértelmű. Ha azonban a jelből és

$$\hat{x}(t) = H\{x(t)\} = \frac{1}{\pi} \cdot P \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (17)$$

Hilbert-transzformáltjából előállítjuk a

$$Z(t) = x(t) + j \cdot \hat{x}(t) = A(t) \cdot e^{j\varphi(t)} \quad (18)$$

komplex analitikus jelet, akkor az ebből származtatható

$$x(t) = A(t) \cdot \cos(\varphi(t)) \quad (19)$$

kanonikus reprezentáció már egyértelmű,

továbbá a pillanatnyi paraméterek is definiálhatók:

$$A(t) = \sqrt{x^2(t) + \hat{x}^2(t)} \quad (20)$$

$$\omega(t) = \frac{d\Phi(t)}{dt} = \frac{d}{dt} \left( \arctan \left( \frac{\hat{x}(t)}{x(t)} \right) \right). \quad (21)$$

Megjegyezzük, hogy a (17) egyenletben az improprius integrál főértéke, a létező,  $\lim_{\Lambda \rightarrow \infty} \int_{-\Lambda}^{+\Lambda} f(x) dx$  alakú határérték szerepel – erre utal a P betű.

A (21) egyenletben a pillanatnyi körfrekvencia az analitikus jel fázisának deriváltjaként áll elő, de számítható az

$$\omega(t) = \text{Im} \left\{ \frac{\partial}{\partial t} \ln(Z(t)) \right\} \quad (22)$$

összefüggés alapján is, amivel

$$\omega(t) = \frac{x(t) \cdot \frac{d\hat{x}(t)}{dt} - \frac{dx(t)}{dt} \cdot \hat{x}(t)}{x^2(t) + \hat{x}^2(t)}, \quad (23)$$

akárcsak a (21)-ben kijelölt deriválás tényleges elvégzésével. Mind (21), mind (23) alapján származtathatunk algoritmust a pillanatnyi frekvencia becslésére. Fontos tulajdonság, hogy a jel és Hilbert-transzformáltjának Fourier-transzformáltja között fennáll az

$$\hat{X}(j\omega) = -j \cdot \text{sgn}(\omega) \cdot X(j\omega), \quad (24)$$

összefüggés, továbbá teljesül, hogy

$$F\{x(t) + j \cdot \hat{x}(t)\} = \begin{cases} 2 \cdot X(j\omega) & \omega > 0 \\ 0 & 0 \leq \omega \end{cases}, \quad (25)$$

ahol  $F\{\cdot\}$  a Fourier-transzformáció műveletét jelöli.

### 3.3. A diszkrét idejű Hilbert-transzformált számítása és a pillanatnyi jellemzők becslése

A diszkrét idejű Hilbert-transzformáltat előállíthatjuk (24)-ből kiindulva megfelelő digitális szűrővel [8], vagy (25) alapján FFT-re alapozott számítási eljárással, amit jelen munka során is használtunk.

A mintasorozat és a Hilbert-transzformált sorozat ismeretében következhet a pillanatnyi amplitúdó és a pillanatnyi frekvencia becslése. A pillanatnyi amplitúdó mintákat (20) alapján a következőképpen határozhatjuk meg:

$$A(n) = \sqrt{x^2(n) + \hat{x}^2(n)} \quad (26)$$

A pillanatnyi frekvencia minták számítására egyrészt (21), másrészt (23) alapján származtathatunk eljárást. A (21) alapján a fázis mintasorozat

$$\Phi(n) = \arctg \left( \frac{\hat{x}(n)}{x(n)} \right), \quad (27)$$

ám a jel időbeni fejlődése során a fázis úgy változik, hogy  $\Phi_u(n) = \Phi(n) + r(n) \cdot 2\pi$ ,  $\Phi(n) \in [-\pi; \pi]$ , (28)

ahol  $r(n)$  pozitív egész szám. A számítások során azonban közvetlenül a fázis főértékének  $\Phi(n)$  mintái adódnak, ebből kell a tényleges fázis mintáit előállítani. A feladat az, hogy minden minta esetében ismert legyen a  $2\pi$  ide tartozó egész szám-szorosa, vagyis elő kell áll-

lítani az  $r(n)$  sorozatot. Erre a célra például a  $\text{mod}(2\pi)$  fázis-visszahajtogatási (phase-unwrapper) eljárás használható [2].

Ha rendelkezésre áll a pillanatnyi fázis, a pillanatnyi digitális körfrekvencia meghatározásához szükséges deriválást az alábbi differencia kiszámításával közelíthetjük:

$$\omega(n) = \Phi_u(n) - \Phi_u(n-1). \quad (29)$$

Más eljárás adódik (23) alapján, ahol a deriválás alkalmos közelítése szükséges. Ahogy az előző pontban, itt is alkalmazható a Savitzky-Golay-féle 5 pontos simító deriválás.

## 4. A Teager-operátor alapján és a HHT-vel számított pillanatnyi jellemzők összehasonlítása

### 4.1. A jel visszaállítása a pillanatnyi jellemzőkből

A 2. pontban ismertettük, hogy a lassan változó jel pillanatnyi amplitúdójának és frekvenciájának abszolút értéke két algoritmus-párral is becsülhető, míg a 3. pontban a benső módusfüggvényekhez rendelt analitikus jel alapján becsültük a pillanatnyi amplitúdót, továbbá vagy közvetlenül, vagy a pillanatnyi fázis előállítását követően a pillanatnyi frekvenciát. Ezekre a becslésekre is megadtunk két algoritmus-párt. Mivel az előző két pontban tárgyalt algoritmusok megközelítési módja, az alkalmazott jelmodell lényegesen különbözik egymástól, felmerül a kérdés, hogy ugyanazon a jelen számolt pillanatnyi jellemzőik hogyan viszonyulnak egymáshoz? Esetleg valamilyen szempontból hasonlóak-e?

Ebben a pontban ezt a kérdést vizsgáljuk meg az alábbi négy összetartozó algoritmus-pár összehasonlításával (zárójelben az ezt követő táblázatokban szereplő elnevezések):

- közvetlen számítással becsült pillanatnyi amplitúdó és frekvencia (közvetlen számítás),
- a diszkrét idejű Teager-operátorral becsült pillanatnyi amplitúdó és frekvencia (ES-algoritmus),
- a kanonikus reprezentáció alapján számolt pillanatnyi amplitúdó és a fázis-visszahajtogatással kapott pillanatnyi frekvencia (HHT (fázis-differencia)),
- a kanonikus reprezentáció alapján számolt pillanatnyi amplitúdó és a simító deriválással kapott pillanatnyi frekvencia (HHT (simító deriválás)).

Az egyes algoritmus-párok összehasonlításának egy lehetséges módja az, hogy az adott jel esetén meghatározzuk velük a pillanatnyi jellemzőket, majd ugyanazon visszaállítási eljárással e pillanatnyi jellemzőkből becsüljük az eredeti jelet. Az  $x(n)$  eredeti jel, és az  $\tilde{x}(n)$  becslés ismeretében az adott algoritmus-pár jóságát az

$$\text{NSR} = 10 \cdot \log \left( \frac{\sum_{n=4}^{N-5} [x(n) - \tilde{x}(n)]^2}{\sum_{n=4}^{N-5} x^2(n)} \right) \quad (30)$$

zaj/jel viszonytal jellemezzük.

Az indexek magyarázata az, hogy a közvetlen számítás során nemcsak a jelre, hanem deriváltjára is alkalmazzuk az 5 pontos simító deriválást, így a jel mindkét széléről elhagyunk 4-4 mintát. Emiatt mindegyik algoritmus-párnál az így adódó jelerészletet vettük figyelembe. A visszaállítási algoritmus alapja maga az adott algoritmus-párhoz tartozó jelmodell. Ehhez a pillanatnyi amplitúdó mindegyik esetben közvetlenül adódik. A saját jelmodelljének megfelelő pillanatnyi fázist azonban csak egy algoritmus állítja elő közvetlenül, a többi három eljárás a pillanatnyi frekvenciára ad becslést, ezért – az egységesség érdekében – mindegyik esetben a pillanatnyi frekvenciából indultunk ki, és ebből határoztuk meg a pillanatnyi fázist az alábbiak szerint:

$$\tilde{\Phi}(k) = \tilde{\Phi}(-1) + \sum_{n=0}^k \tilde{\omega}(n) \quad k = 0, 1, \dots, N-1 \quad (31)$$

Numerikus kísérleteink tanúsága szerint az egyes esetekben a visszaállított jel és az eredeti jel között fázis-ingadozás mutatkozik. Ezért mindegyik algoritmus-párnál kereséssel határoztuk meg a legjobb NSR-t adó  $\Phi(-1)$  kezdőfázist  $\pi/180$  ( $1^\circ$ ) fázisléptetés mellett.

Az összehasonlítást vizsgálójelen és egy szó bemondásából származó beszédjelen is elvégeztük.

#### 4.2. A módszerek összehasonlítása vizsgálójel esetén

Vizsgálójelnek a szakirodalomban található AM-FM-jelet használtuk [2]:

$$s(n) = (0,998)^n \cdot \left[ 1 + 0,8 \cdot \cos(2 \cdot \pi \cdot f_3 \cdot n) \right] \cdot \cos \left[ 2 \cdot \pi \cdot \left( f_1 \cdot n + \frac{1}{2 \cdot \pi} \cdot \sin(2 \cdot \pi \cdot f_2 \cdot n) \right) \right] \quad (32)$$

$$f_s = 10000\text{Hz} \quad f_1 = \frac{1000\text{Hz}}{f_s} \quad f_2 = \frac{100\text{Hz}}{f_s} \quad f_3 = \frac{50\text{Hz}}{f_s}$$

Időbeli alakja alapján ez a jel egyben benső módusfüggvény is, ezért azt várjuk, hogy az EMD-algoritmus egyetlen lényeges IMF-et ad vissza.

Ez így is van, amint az a túlololdali 1. ábrán is látható. A visszaállított jel eltérését mind az eredetitől, mind az IMF-től számszerűen jellemezve az 1. táblázatban látható adatokat kapjuk.

Az 1. ábra a számított eredményeket szemlélteti vizsgálójel-részleten. Az ábra b) részén kivehető, hogy az  $1^\circ$ -os fázisléptetés ellenére egyik-másik módszernél még marad kis fázishiba, ami nyilván rontja a zaj/jel viszonyt.

1. táblázat  
Az algoritmus-párok jellemzése a vizsgálójel esetében

Módszer	Eredeti jel NSR (dB)	IMF1 NSR (dB)
Közvetlen számítás	-8	-8
ES-algoritmus	-18	-19
HHT (fázis-differencia)	-24	-27
HHT (simító deriválás)	-7	-7

2. táblázat  
Az algoritmus-párok jellemzése a sávszűrt beszédjel esetén

Módszer	Eredeti jel NSR (dB)	IMF1 NSR (dB)
Közvetlen számítás	-5	-5
ES-algoritmus	-2	-2
HHT (fázis-differencia)	-29	-30
HHT (simító deriválás)	-14	-14

Az elméleti pillanatnyi frekvenciát az egyes módszerek kis hibával közelítik, az elméleti pillanatnyi amplitúdó közelítése is közel azonosan jó.

#### 4.3. A módszerek összehasonlítása sávszűrt beszédjel esetén

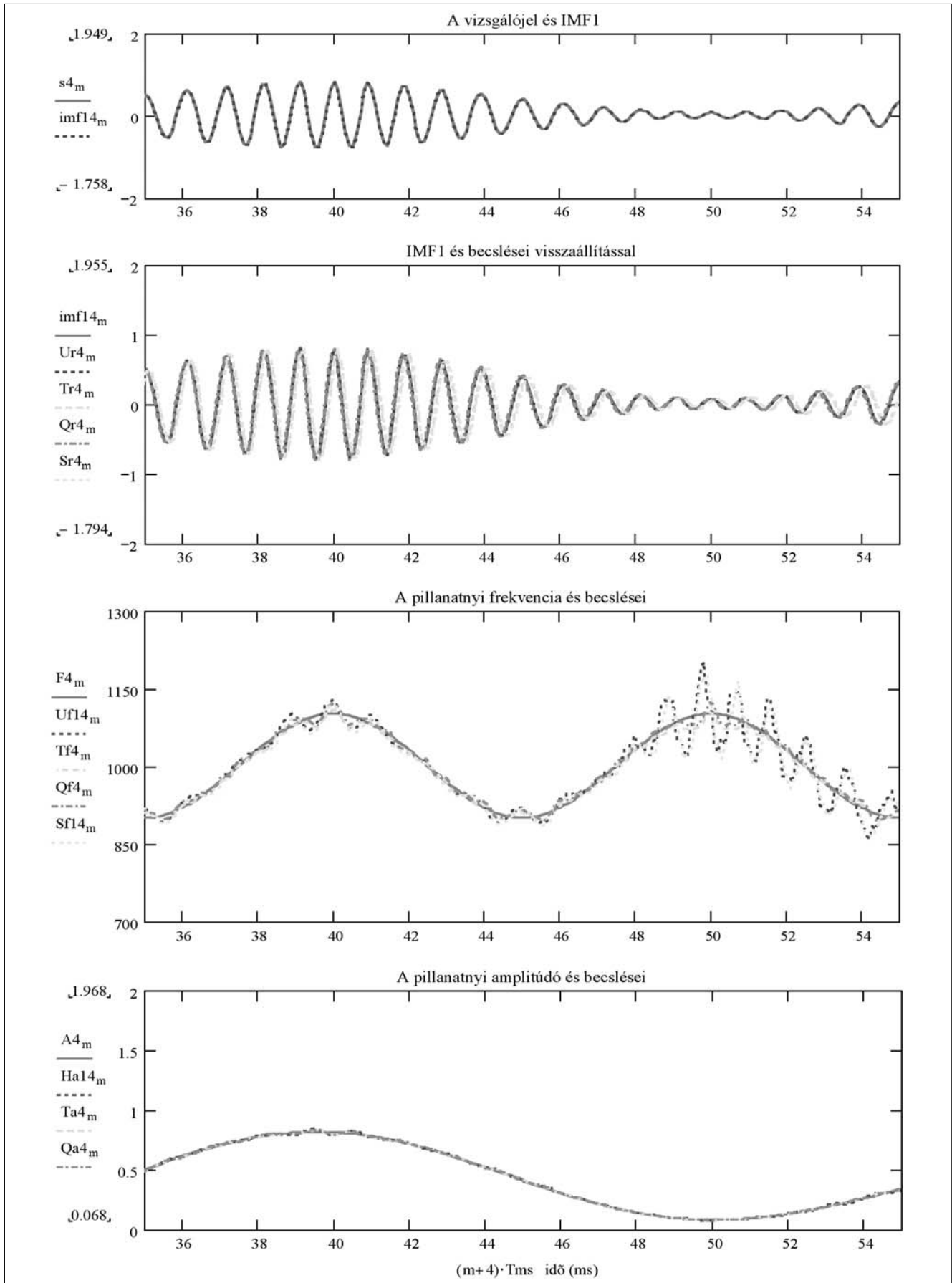
Az előző pontban a vizsgálójel – konstrukciójánál fogva – olyan volt, hogy pillanatnyi jellemzői lassan változtak, így a (6) feltétel teljesült, ami a pillanatnyi paraméterek becsléséhez szükséges mind a közvetlen számítás, mind az ES-algoritmus esetében. Ennek megfelelően a beszédjel esetében is gondoskodni kell arról, hogy a becsléni kívánt pillanatnyi jellemzők lassan változzanak.

Ez megfelelő sávszűréssel biztosítható. A megfelelő sávszűrő tervezésére beszédfeldolgozási feladatokhoz – ismereteink szerint – nincs általánosan elfogadott módszer, de a szakirodalom szerint az egységnyi kritikus sávzélességű szűrősor (valamely tagja) megfelelő a Teager-operátor alkalmazhatóságához [4], ami az említett két eljárás alapja. A sávszűrő alkalmazásának praktikus oka is van, hiszen a tapasztalat szerint megfelelő sávszűrés után a diszkrét idejű Teager-operátor sokkal ritkábban ad negatív, tehát fizikailag nem értelmezett értéket, mint a nélkül.

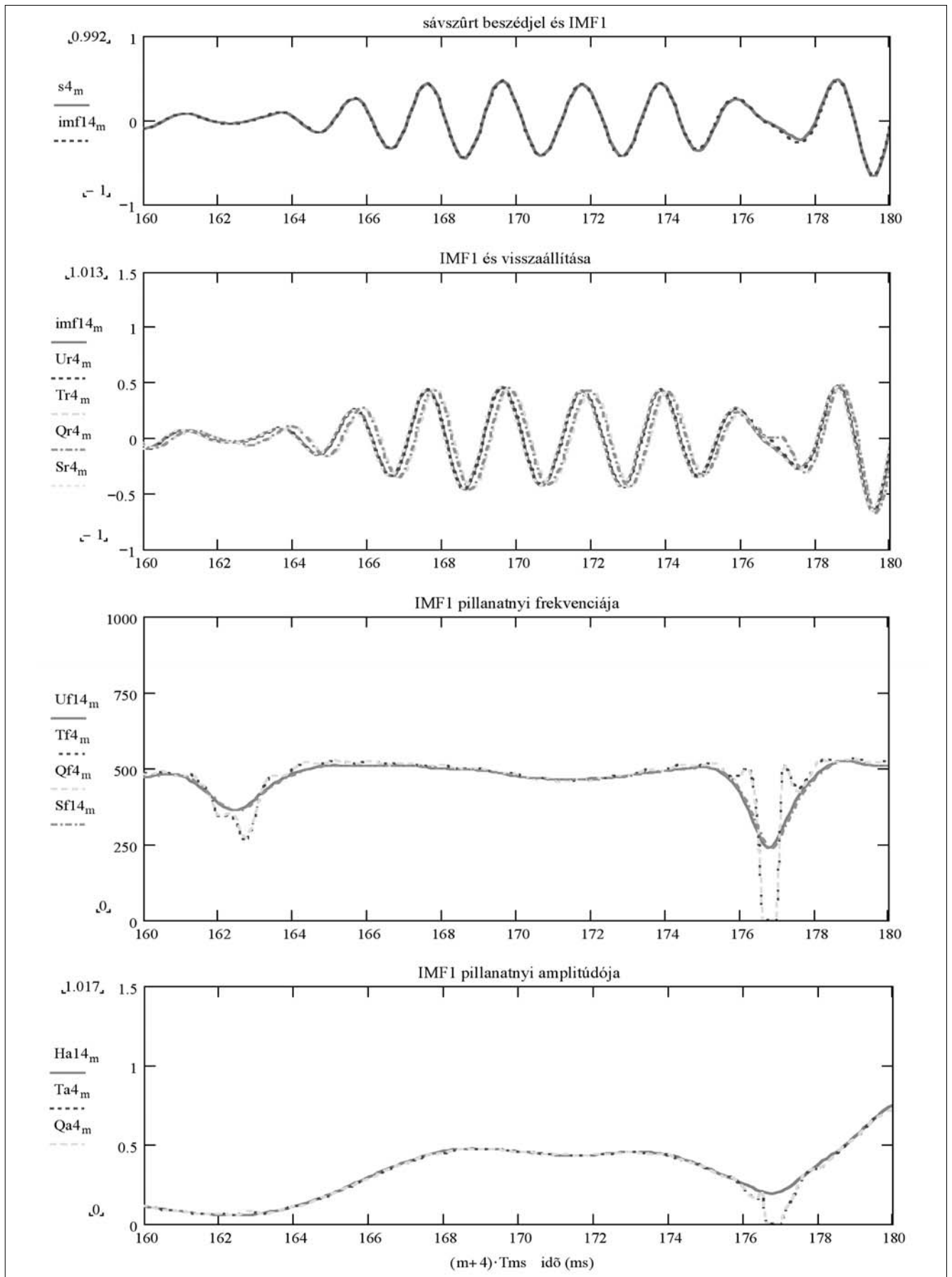
Ebben a pontban sávszűrt beszédjel pillanatnyi jellemzőinek becslését mutatjuk be. A beszédjelminták az igen szó férfi bemondótól származó megvalósításából származnak 8000 Hz mintavételi frekvencia és 16 bites lineáris kvantálás alkalmazásával. Az eredeti bemondást 300 Hz...3400 Hz áteresztő sávú lineáris fázisú FIR-szűrővel sávhatároltuk. A spektrogram megtekintése alapján az 500 Hz körüli erős formáns jelenléte miatt hallásmoell alapú wavelet-szűrősor egyik tagját alkalmaztuk további lineáris fázisú FIR-szűrésre [9]. Az így előállt jel amplitúdóját a  $-1 \leq x(n) \leq 1$  egyenlőtlenség szerint normalizáltuk, ez lett a pillanatnyi jellemzők becslésének alapja.

Megtekintve a sávszűrt jelet, látható, hogy ez is benső módusfüggvény, ezért azt várjuk, hogy az EMD-algoritmus egyetlen lényeges IMF-et ad vissza. Ez itt is így van, amint az a 2. ábrán is látható. A visszaállított jel eltérését mind az eredetitől, mind az IMF-től számszerűen jellemezve a 2. táblázatban látható adatokat kapjuk. A legjobb eredményt a HHT (fázis-differencia) módszer adja.





1. ábra A négy algoritmus-párral számolt eredmények szemléltetése vizsgálójel-részleten:  
 a) a vizsgálójel és az első benső módusűggyvény (IMF1), b) az IMF1 és négyféle becslése,  
 c) az elméleti pillanatnyi frekvencia és négyféle becslése, d) az elméleti pillanatnyi amplitúdó és háromféle becslése



2. ábra A négy algoritmus-párral számolt eredmények szemléltetése sávszűrt beszédjel-részleten:  
 a) a sávszűrt beszédjel és az első benső módustüggvény (IMF1), b) az IMF1 és négyféle rekonstruálása,  
 c) a pillanatnyi frekvencia négyféle becslése, d) a pillanatnyi amplitúdó háromféle becslése

A 2. ábra a számított eredményeket szemlélteti sáv-szűrt beszédjel-részleten. Az ábra b) részén látható, hogy az 1°-os fázisléptetés ellenére több módszernél is van fázishiba. (Megjegyezzük, hogy bár az NSR alapján az eredeti beszédhez képest nagy eltérésre következtetnénk három algoritmusnál is, a rekonstruált beszédet meghallgatva azt jónak találjuk.) A pillanatnyi frekvencia becslésénél együtt fut rendre a két HHT-s és a két Teager-operátoros algoritmussal számolt adatsor. Ez utóbbiaknál a 177 ms-nál lévő beszakadás oka az, hogy a megvalósított program 0 becslült frekvenciaértéket ad vissza, ha negatív számból kellene gyököt vonni (lásd (10),(11),(14),(15)).

Ez a megoldás az algoritmus vizsgálatokor fontos, a gyakorlati alkalmazásban a környező adatokból becslült helyettesítő értékkel élhetünk ilyenkor. A kis pillanatnyi amplitúdót és a jelrészletet megvizsgálva látható, hogy az efféle bizonytalan becslés a 0-hoz közeli jelamplitúdóknál fordulhat elő. Ettől eltekintve a négyféle módszer becslései jól egyeznek.

#### 4.4. A módszerek összehasonlítása beszédjel esetén

Az előző pontban bemutatott eredmények egyrészt megerősítik azt a tapasztalatot, hogy a beszédjel Teager-operátoros feldolgozásához elegendő az egységnyi kritikus sáv szélességű szűrőkből álló szűrősor alkalmazása, másrészt megfigyelhető, hogy a Teager-operátorra alapozott becslések nagyon hasonlítanak a HHT-vel kapható becslésekhez. Felvetődik a kérdés: helyettesítheti-e a szóban forgó szűrést a természetes módusfelbontás, és hogyan alakulnak a becslült pillanatnyi jellemzők? Erre a kérdésre megítélésünk szerint csak nagy beszéd-adatbázison történő részletes vizsgálat eredményei alapján lehet válaszolni. Az alábbiakban egyetlen szó bemondásának elemzésével kapott eredményeinket mutatjuk be.

Ezekben a vizsgálatokban tehát nem szerepel sáv-szűrés. Maga a természetes módusfelbontási eljárás viselkedik sáv-szűrőként, mégpedig az adott beszédjelhez igazodó, adaptív módon. Ugyanis a felső és alsó burkolók egymáshoz igazítása a helyi maximumokhoz és minimumokhoz kapcsolódik, vagyis az első benső módusfüggvény a jelamplitúdóban lévő, egymás szomszédságában található gyors változásokhoz, így a magasabb frekvenciájú spektrális részlethez igazodik. Utána azt a jelből levonva haladunk tovább a következő módusfüggvényekhez, vagyis a kisebb frekvenciájú spektrális részletek felé. (Az EMD-eljárás illetően viselkedése jól nyomon követhető az egyes IMF-ek spektrumján is.) Emiatt az a kérdés, hogy az így megvalósuló adaptív szűrés elégséges-e a Teager-operátoros pillanatnyi jellemző-becslésekhez?

Jelen dolgozatban ezt a kérdést is az előző pontban szereplő *igen* szó bemondásából nyert mintasorozat-

ton vizsgáltuk. A 3. pontban említettük, hogy az EMD-algoritmus alkalmazásakor nincs támpont arra, hogy mennyi a benső módusfüggvények elégséges száma. Numerikus kísérleteink azt mutatják, hogy az első három benső módusfüggvényből (16) alapján az eredeti beszédjel NSR= -22 dB jósággal állítható vissza, ezért a pillanatnyi jellemzőket az első három módusfüggvényre számítottuk ki a négyféle módszerrel, és a visszaállítást is rendre a három benső módusfüggvényre végeztük el, majd a rekonstruált beszédjelet ezek összegzésével határoztuk meg.

Az áttekinthetőség érdekében a 3. ábrán csak a legjobban közelítő algoritmussal kapott beszéd-részleteket mutatjuk be, a 3. táblázat a számszerű eredményeket tartalmazza.

A 3. ábrához tartozó fontos megjegyzés, hogy nem szerepel a másik három módszerrel kapott visszaállított beszédjel, de itt is megfigyelhető a fázisingadozás jelensége, ami a zaj/jel viszonyt lerontja, viszont maga a beszéd jól érthető.

## 5. Következtetések

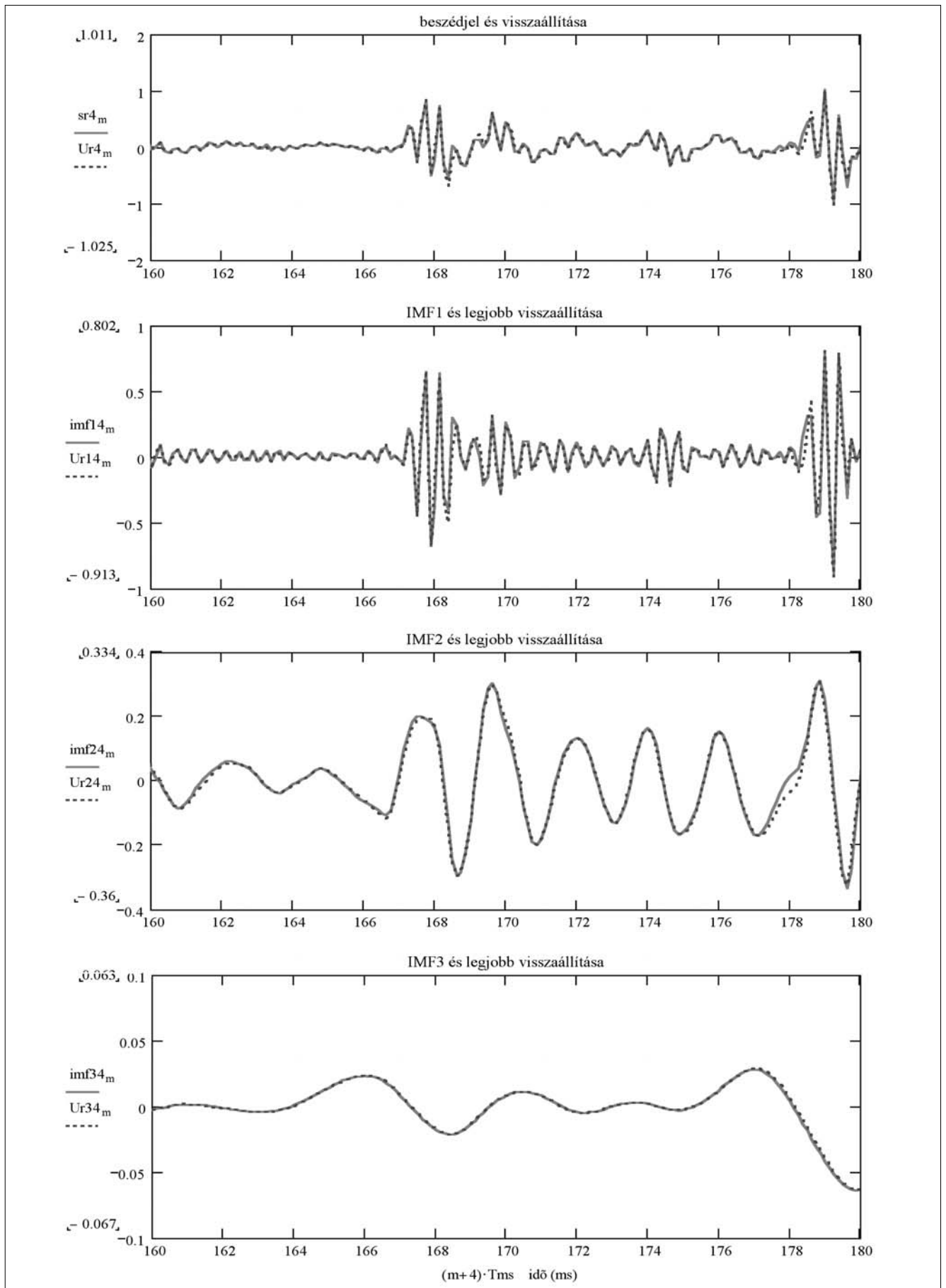
A dolgozatban a beszédjel pillanatnyi amplitúdójának és pillanatnyi frekvenciájának becslésére mutattunk be négyféle módszert. Ezek közül kettő a Teager-operátorra, kettő pedig a Hilbert-Huang-transzformációra épül. A vizsgálójel és beszédjel pillanatnyi jellemzőinek becslési példáin ábrákkal szemléltettük az egyes módszereket, és megadtunk egy visszaállítási eljárást is, amivel a beszédjel a becslült pillanatnyi jellemzőkből rekonstruálható. Ez alapján már alkalmas zaj/jel viszonytal hasonlíthatók össze az egyes módszerek. A munka során szerzett tapasztalatainkat a dolgozatban több helyütt ismertettük.

Legfontosabb következtetéseink az alábbiak:

1. A Teager-operátorra alapozott módszerek esetében lényeges, hogy sáv-szűrt beszédjelen végezzük a becslést. Erre a célra szolgálhat valamely egységnyi kritikus sáv szélességű szűrő, ennek kimenete dolgozható fel tovább a Teager-operátorra alapozott algoritmusokkal. Ez a Teager-operátorra épülő pillanatnyi amplitúdó és pillanatnyi frekvencia becslés esetében is igaz.
2. A beszédjelből a természetes módusfelbontási eljárással kapható benső módusfüggvények pillanatnyi jellemzőire mind a Teager-operátorra alapozott módszerekkel, mind a HHT alapján egymáshoz hasonló eredmények adódnak. Ez nem magától értetődő, és megítélésünk szerint érdemes nagy beszéd-adatbázison részletesen megvizsgálni, hogy általánosabb érvényű-e ez a megfigyelésünk.

3. táblázat  
A legjobb visszaállítást adó  
módszer jellemző adatai

Módszer	Eredeti jel NSR (dB)	IMF1 NSR (dB)	IMF2 NSR (dB)	IMF3 NSR (dB)
HHT (fázis-differencia)	-12	-10	-19	-24



3. ábra A beszédjel valamint az első három módustüggvény visszaállítása a pillanatnyi jellemzőkből: a) az eredeti jel és a becsült IMF-ek összegzésével kapott visszaállítás b-d) rendre az egyes IMF-ek és legjobb becsléseik

3. Az ES-algoritmusnál tapasztalható az a hiba, hogy néha negatív számból kellene a végrehajtás során négyzetgyököt vonni, amit a szomszédos becslések alapján javasolunk kiküszöbölni. Ugyanis – bár kézenfekvő lenne mediánszűrővel simítani a pillanatnyi jellemzőket becslő adatsorokat – tapasztalataink szerint az így simított változathoz visszaállított beszédjel a meghallgatáskor rosszabb minőségű, mind mediánszűrés nélkül.
  4. A visszaállítás során tapasztalható egy fázisingadozási jelenség, mely szerint a nullához közeli jelamplitúdót követő jelerészlet visszaállítása időben elcsúszik. Emiatt jobb visszaállítást várhatunk el, ha a rekonstruáló algoritmust úgy módosítjuk, hogy minden nulla-közeli jelerészlet után keresse meg a legjobb illeszkedést adó kezdőfázist.
  5. Az EMD-algoritmus módosítható annak figyelembe vételével, hogy a vizsgált beszédjel eleve sávhatárolt. Így – például a maradékjel és az eredeti jel megfelelően előírt zaj/jel viszonya alapján – automatikusan kaphatjuk meg a szükséges számú benső módusfüggvényt.
- További feladatunk a jelen dolgozatban tárgyalt algoritmusok alkalmazási lehetőségeinek feltárása.

#### Köszönetnyilvánítás

A szerző ezen a helyen is megköszöni Gordos Gézának, Németh Gézának és Tatai Péternek (BME VIK TMIT) a gépi beszédfeldolgozási célú algoritmusfejlesztési munkái során kapott segítségét, támogatást és biztatást.

#### Irodalom

- [1] Gordos G., Takács Gy.:  
Digitális beszédfeldolgozás. Műszaki Könyvkiadó, 1983.
- [2] Quatieri, T. F.:  
Discrete-time Speech Signal Processing: Principles and Practice. Prentice-Hall, 2001.
- [3] Abbate, A., DeCusatis, M. C., Das, K. P.:  
Wavelets and Subbands: Fundamentals and Applications. Birkhäuser, 2002.
- [4] Chen, S-H., Wang, J-F.:  
„Speech Enhancement Using Perceptual Wavelet Packet Decomposition and Teager Energy Operator”,  
Journal of VLSI Signal Processing 36, pp.125–139.,  
Kluwer Academic Publishers, 2004.
- [5] Huang, N. E., Shen, Z., Long, S. R.,  
Wu, M. C., Shih, H. H., Zheng, Q., Yen, N-C.,  
Tung, C. C., Liu, H. H.:  
„The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”.  
Proc. R. Soc. Lond. A (1998) 454, pp.903–995.
- [6] Valkó P. Vajda S.:  
Műszaki-tudományos feladatok megoldása személyi számítógéppel. Műszaki Könyvkiadó, 1987.
- [7] Gábor, D.:  
Theory of communication.  
J. Inst. Electr. Eng., Vol. 93. (1946), pp.429–457.
- [8] Simonyi E.:  
Digitális szűrők – a digitális jelfeldolgozás alapjai.  
Műszaki Könyvkiadó, 1984.
- [9] Pintér, I.,  
„Perceptual wavelet-representation of speech signals and its application to speech enhancement”,  
Computer, Speech and Language,  
Vol. 10. No.1. pp.1–22., Academic Press, 1996.

# A kibővített Jászberény Rádióállomás új sugárzási rendszere és az optimális üzemi frekvenciaterv meghatározása

STANDEISKY ISTVÁN, DÓSA GYÖRGY, BALLA ÉVA

standi@sze.hu, ballae@ahrt.hu

Lektorált

**Kulcsszavak:** hullámterjedés, rádió-műsorszórás, rövidhullám, optimális üzemi frekvencia

A cikk a 2004-2005 folyamán kibővített Jászberény Rádióállomás történetének és antennáinak rövid ismertetője után bemutatja az optimális üzemi frekvencia meghatározásának folyamatát különböző célterületekre három különböző módszerrel. Az írás egy megrendelésre készült, nagy terjedelmű tanulmány rövidített változata.

## 1. Bevezetés

A jászberényi rövidhullámú műsorszóró rádióállomás 1974. november 4-én kezdte meg rendszeres rövidhullámú műsoradásait. A Magyar Rádió által kívánt sugárzási célterületként az alábbiak voltak megadva:

- Észak-Amerika, Kanada;
- Dél-Amerika;
- Észak-Afrika, Arábia;
- Közél-Kelet;
- Nyugat, Dél és Észak-Európa, másodlagosan;
- Ausztrália, Új-Zéland;

A kiépült Jászberényi Rádióállomás rendszertechnikai tervét a 1. ábra mutatja be.

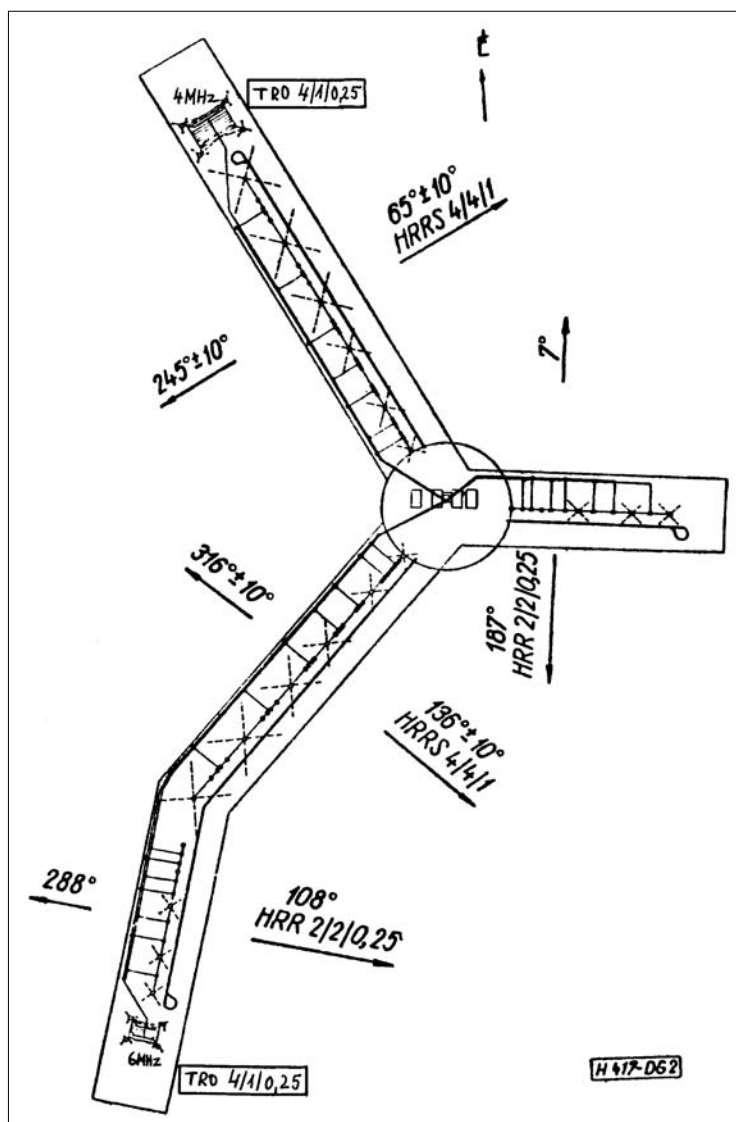
A megadott célterületek besugárzására legkedvezőbbnek bizonyult, ha a síkantennákat magába foglaló antenناسorok háromágú, úgynevezett csillag formában épülnek ki. Az antenناسorok tehát egymáshoz képest körülbelül 120°-os szöveget zárnak be a sugárzási irányok által meghatározott tájolásban.

A nagytávolságú célterületek besugárzására az észak-nyugati és a dél-nyugati antennaágban 7-7 db egysávos HRRS 4/4/1 típusú síkantenna üzemel. A kívánt főirányokban távvezérlés segítségével 180°-os irányváltás eszközölhető (irányváltó kapcsolók). Ezen síkantennák vízszintes iránykarakterisztikáján – a főirányhoz képest – távvezérléssel  $\pm 10^\circ$ -os billentés biztosítható a nagyobb terület besugárzása érdekében.

A HRRS 4/4/1 típusú egysávos síkantennák nyeresége 19 dB, az előre-hátra viszony pedig 20 dB. Kilövellési szögük  $\vartheta_0 = 7^\circ$ ; a vertikális karakterisztika 3 dB-es pontjai  $4^\circ$ , illetve  $11^\circ$ . A középtávolságú célterületek besugárzására a keleti és a dél-nyugati antennaágban szintén a kívánt főirányokban ugyancsak 7-7 db egysávos HRR 2/2/0,25 típusú síkantennák üzemelnek és távvezérléssel szintén 180°-os irányváltás biztosítható.

A HRR 2/2/0,25 típusú egysávos síkantennák nyeresége 13 dB, az előre-hátra viszony pedig 18 dB. Kilövellési szögük  $\vartheta_0 = 17^\circ$ , a 3 dB-es pontjai a vertikális karakterisztikáknak  $9^\circ$ , illetve  $26^\circ$ .

1. ábra  
Az adóállomás eredeti rendszertechnikai terve



Az észak-nyugati és a dél-nyugati antennasorok végén egy-egy TRO 4/1/0,25 típusú közelsugárzó (körsugárzó) antennarendszer üzemel a 4, illetve 6 MHz-es műsorszórási sávokra, melyek a Kárpát-medence és részben Közép-Európa besugárzását biztosítják. Az antennakombinátor 7 kimeneti pontjáról induló 6 főtápvonal (minden csillagághoz két-két főtápvonal), melyek az irányított és a közelsugárzó antennarendszerek táplálását biztosítják, 300 ohm hullámellenállású szimmetrikus légtápvonal kiépítésűek. A síkantennák úgynevezett poligonos felfüggesztéssel épültek ki.

A kilencvenes évek elején Ausztráliából a „Szülőföldünk” rövidhullámú adásainkra beérkező vételmegfigyelések (levelek, illetve kazettafelvételek) megoszlása azt mutatta, hogy Kelet-Ausztrália területéről érkezett a megfigyelések, értékelések hozzávetőlegesen 80%-a (Sydney-Melbourne-Canberra). A nyugat-auztráliai területről (Perth) érkező kevés megfigyelés általában kedvezőtlen eredményeket mutatott, még optimális frekvenciahasználat esetében is.

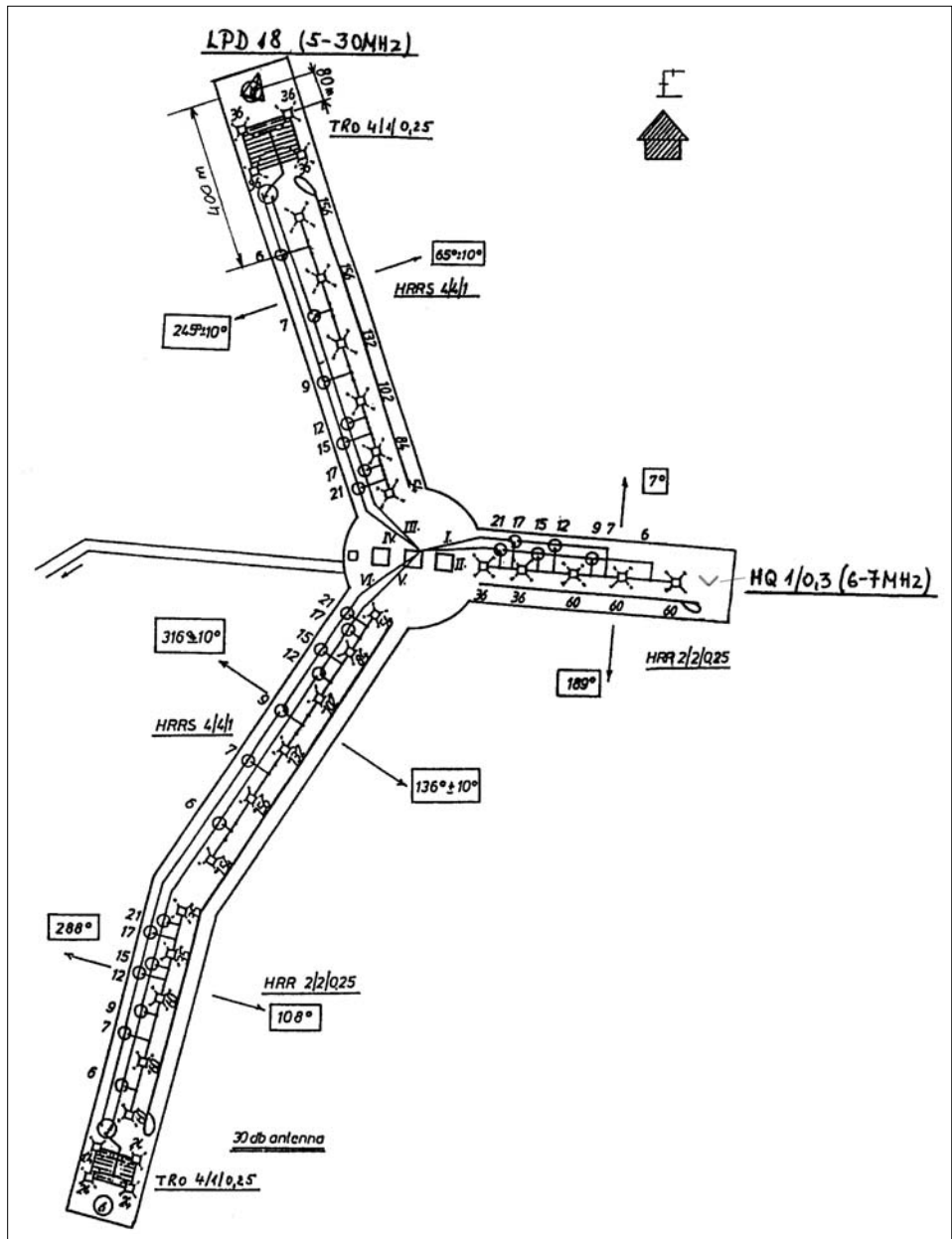
A fő probléma a nyugat-auztráliai terület kedvezőtlen besugárzásánál az volt, hogy a nagynyereségű 65°-os HRR 4/4/1 típusú síkantenna +10°-os vízszintes karakterisztika-billentéssel, tehát 75°-os főirányítással a távoli kontinensnek csak körülbelül 1/3 részét sugározta be, a középső és nyugati terület rész ellátatlan maradt. Felmerült annak szükségessége, hogy feltétlenül javítani kell Ausztrália és Új-Zéland ellátásán. Egyetlen kedvező lehetőség kínálkozott a meglévő 65°-os síkantenna figyelembevételével, mégpedig a vízszintes karakterisztika körülbelül ±22°-os billentésűre történő növelése a fázistolórendszer átalakításával. A fejlesztés arra irányult, hogy a 65°-os 21 MHz-es HRRS 4/4/1 típusú síkantenna-rendszeren az átalakítás mielőbb megtörténjen annak érdekében, hogy +22°-os billentéssel Ausztrália nyugati területe is ellátást kapjon.

A fejlesztési, illetve az átalakítási munkálatok 1996. szeptemberében valósultak meg, de csak a 21 MHz-es síkantennán. A nagyobb billentés esetén a vízszintes karakterisztikák kiszélesedtek, és ez még kedvezőbbé tette

Ausztrália és Új-Zéland lefedését. További jelentős eredmény volt még, hogy +12°-os billentéssel 284°-os főirányú sugárzás esetén egész Közép-Amerika, amely az ideig nem volt besugározható, ezáltal besugározhatóvá vált. Az átalakított 21 MHz-es síkantennarendszer 1997. január elején kezdte meg f=21685 KHz üzemi frekvencián a „Szülőföldünk” műsorát sugározni naponta UTC idő szerint 10-11 óra között. Hullámterjedési szempontból ezen időszak volt a legkedvezőbb Ausztrália felé. A beérkező vételmegfigyelések Sydney és Dunedoo térségéből igen kedvező eredményt mutattak, a SINPO értékek általában 45444 voltak.

A jászberényi rövidhullámú adóállomás antenna- és adórendszere az elmúlt 30 év alatt igen kedvezően és sikeresen biztosította a kívánt célterületekre a Magyar Rádió magyar és idegen nyelvű rövidhullámú adásainak eljuttatását.

2. ábra  
Az adóállomás bővített rendszerteknikai kialakítása



Az elmúlt időszak alatt a célterületekről beérkezett (több ezres nagyságrendű) vételmegfigyelések (levelek és magnófelvételek) bizonyították, és bizonyítják ma is, hogy adásainkat általában az idő 80%-ban kedvező eredménnyel tudták és tudják venni a hallgatók. A SIN-PO átlagértéke rendszeresen 4443 volt.

## 2. A bővített jászberényi rövidhullámú adóállomás rendszer- és sugárzási vizsgálata

A diósi rövidhullámú adóállomás üzemeltetése 1998-ban, a székesfehérvári rövidhullámú rádióállomásé pedig 2004-ben véglegesen megszűnt. A székesfehérvári adóállomás üzemének megszüntetése kapcsán döntés született, hogy az ott üzemelő 2 db 100 kW-os BBC gyártmányú rövidhullámú adóberendezés és az egy darab HQ 1/0,3 típusú kétsávós (6 és 7 MHz) körsugárzó antennarendszer, mely 1987-ben lépett üzembe, valamint az 1995 szeptemberében üzembe állított korszerű Thomcast gyártmányú TH LPD 18 típusú forgatható és dönthető rövidhullámú log-per antennarendszert áttelepítésre kerüljön a jászberényi adóállomásra, és így a kibővült adóállomásról sugároznák a kívánt célterületre a Kossuth, a Szülőföld, valamint az egyéb magyar rövidhullámú műsorokat.

A Thomcast gyártmányú forgatható és dönthető log-per antennarendszer az északnyugati antennaág végére került telepítésre, a HQ 1/0,3 típusú 6-7 MHz-es egy darab körsugárzó antennarendszer pedig a keleti antennaág végére került telepítésre 2005-ben. A TH LPD 18 típusú log-per antenna, valamint a HQ 1/0,3 típusú 6-7 MHz-es körsugárzó antenna rendszer csak a 100 kW-os BBC adóberendezésekkel üzemel.

A bővített jászberényi rövidhullámú adóállomás antennahálózatának rendszertechnikai kialakítását az előző oldali 2. ábra szemlélteti.

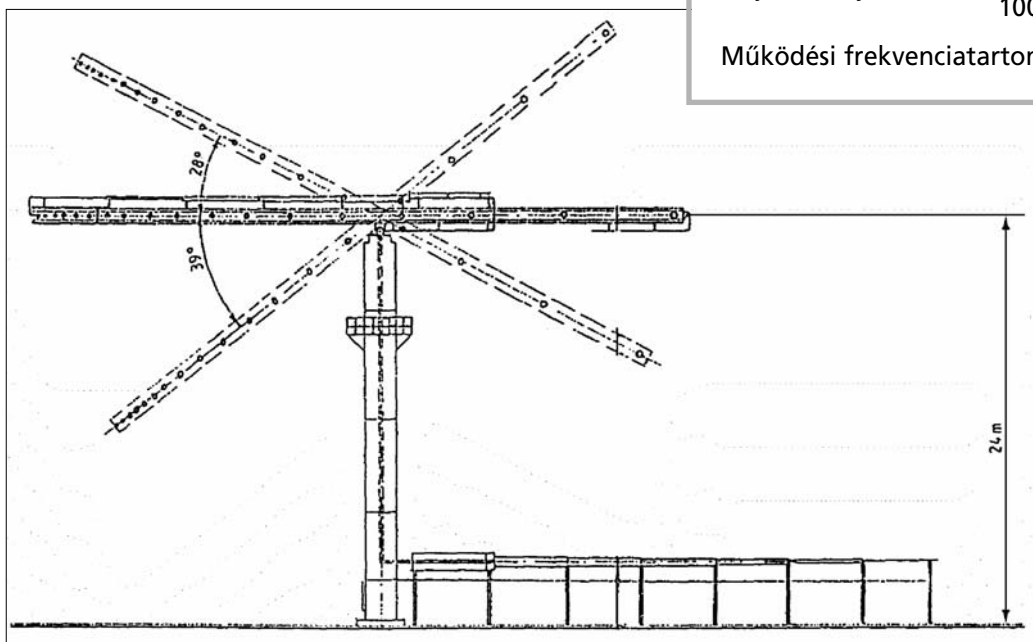
Hangsúlyozni kell, hogy a két áttelepített antennarendszer a HQ 1/0,3 6-7 MHz-es kvadrát körsugárzó, valamint az LPD 18 típusú forgatható és dönthető log-per antennarendszer székesfehérvári üzemeltetésük idején igen kedvező eredménnyel sugároztak a kívánt célterületekre. A körsugárzó antennarendszer európai területre körülbelül 1200...1500 km távolságra kedvező. A log-per antennarendszer pedig – miután billentési szögének változtatásával a kilövellési szöget optimálisan lehet megválasztani – a kívánt célterület besugárzásánál az optimális üzemi frekvencián forgathatóságával bármilyen irányban, főleg középtávolságú területre igen rugalmasan alkalmazható.

Az áttelepítésre került TH LPD 18. típusú forgatható és billenthető log-per antennarendszer kialakítását a 3. ábra mutatja be.

Ezen antennarendszer főbb műszaki adatai:

### Thomcast TH LPD 18

Polarizáció:	horizontális
Nyeresége:	10-12 dB
VSWR:	1,6
Csatlakozási impedancia:	50 $\Omega$ aszimmetrikus
Dipólok száma:	18 db
A főtartó hossza:	36,5 m
A legnagyobb dipól hossza:	26,35 m
A legkisebb dipól hossza:	2,86 m
A billentési szöghatárok:	$\Psi = +28^\circ \sim -39^\circ$
Vízszintes forgatás:	360°
A dipólvégek által bezárt félszög:	$\alpha \cong 20^\circ$ $2\alpha \cong 40^\circ$
Léptékállandó:	$\tau \cong 0,852$
Térközállandó:	$\delta \cong 0,12$
Aktív antennaelemek száma:	$N_A \cong 4$ db
Teljesítmény:	100 KW vivó + 100% AM
Működési frekvenciatartomány:	5...30 MHz



3. ábra

A Thomcast forgatható-billenthető log-per antennája



1. és 2. táblázat

Billentési szög ( $\Psi^\circ$ )	Antenna magasság (m)		Kilövellési szög ( $\Theta$ )		Távolság (km)			
	$f_a$	$f_m$	$\Theta_a^\circ$	$\Theta_m^\circ$	Alacsony frekvencia		Magas frekvencia	
					F2	E	F2	E
-39°	35	12	40°	14°	700	280	1800	800
-20°	29,5	18	36°	10°	800	320	2300	1100
0°	24	24	30°	7°	1000	400	2600	1300
+14°	19,5	28,5	50°	6°	500	200	2700	1400
28°	15	33	60°	5°	350	150	3000	1600

Sávok	$\Psi = 0^\circ$	$\Psi = +28^\circ$	$\Psi = -39^\circ$
		Hatásos terület (km)	
6–7 MHz	400 ~ 1700	250 ~ 1600	300 ~ 1600
9–11–13 MHz	1200 ~ 2400	1300 ~ 2800	200 ~ 1700
15–17 MHz	1800 ~ 3100	1900 ~ 3200	200 ~ 1700
21–26 MHz	2100 ~ 3400	2300 ~ 3500	150 ~ 1300

A különböző billentési szögekhez tartozó kilövellési szögek és a számítással meghatározott sugárzási magasság ismeretében az optimális sugárzási távolság meghatározása került a két szélső frekvenciahatár esetében, a főnyalábra vonatkozóan F2- és E-réteges terjedésre, egyugrásos átvitel esetén.

Az eredményeket összefoglalóan az 1. táblázat tartalmazza. A 2. táblázat pedig összefoglalóan tartalmazza a billentésből adódóan ( $\Psi=0$  alapálláshoz képest) az egyes sávok optimális hatáskörzetének alakulását, növekedését, tehát a billentés előnyét. A log-per antenna alkalmazásával – miután forgatható és dönthető – a kívánt célterületre mindig a hullámterjedési igényeknek megfelelően lehet sugározni.

Az antennarendszer hatáskörzete az adott torony-magasság és az adott billentési szöghatárok esetében a frekvencia függvényében körülbelül 300 km-nél kezdődik, és többugrásos F-réteges terjedést figyelembe véve 11...12000 km-ig terjed, tehát e tartományban használható optimálisan.

### 3. A kedvező frekvenciasávok meghatározása különböző célterületekre

A kibővített jászberényi rövidhullámú adóállomás antennarendszerének típusait és fő paramétereit a 3. táblázat foglalja össze.

Feltétlenül szükségessé vált tehát, hogy az állomás antennahálózatára, a kívánt célterületekre (főirányokra) optimális frekvenciahasználati és besugárzási vizsgálat, illetve terv készüljön az ionoszféra szokásos átlagértékei alapján napfoltmaximumos, -minimumos és -közepes értékekre; téli-nyári, valamint átmeneti (őszi-tavaszi) időszakra,

napokra, napi felhasználási lehetőséggel. A frekvenciaszükséglet számításoknál a napfoltmaximumra  $R=160$ , a napfoltminimumra pedig  $R=10$ , illetve 20 értéket vettünk figyelembe az IFRB, CCIR és a CRPL ajánlásai, adatai alapján.

A nemzetközi rövidhullámú műsorszóró szolgálat üzembiztonságának előfeltétele – a teljesítményigény mellett – a lehető legkedvezőbb frekvencia használata. A terjedési adatok számításánál a napfoltmaximumos és -minimumos nyári és téli időszakok figyelembevétele megadja az üzemi frekvenciasáv alsó és felső értékeit. Az ionoszféra részletes, hosszú évek alatt folytatott vizsgálatai alapján dolgoztak ki különböző intézetek, szervezetek (CRPL; SPIM; IPS, RRS, NBS, IZMIRAN) a hullámterjedési mechanizmuson alapuló frekvencia-előjelzési módszereket, vizsgálatokat.

A rövidhullámok terjedését elsősorban az ionoszféra viselkedése befolyásolja, és ebben igen jelentős szerepe van a Napnak, mely az ionizáló sugárzások fő forrása. Ezért érthető tehát, hogy milyen fontos az ionoszféra és a naptevékenységek ismerete, előrejelzése a kívánt célterület besugárzástervezése szempontjából. További befolyásoló tényező még a Föld mágnesessége, valamint a talaj is, amelynek csillapító hatása van a visszaverődési pontokban többugrásos terjedés esetén. Ezek a befolyásoló tényezők szabják meg a használható legnagyobb frekvenciát (MUF), a használható legalacsonyabb frekvenciát (LUF) és az optimális üzemi frekvenciát (FOT vagy OWF), valamint az átviteli út csillapítását.

3. táblázat

Típus	P [kW]	f sáv [MHz]	$\varphi$ [°]	s [°]	$\zeta$ [°]	Gd [dB]
HRR 2/2/0,25	250	6-7; 9-12-15-17-21	$\pm 25$	-	17	13
HRRS 4/4/1	250	6-7; 9-12-15-17-21	$\pm 12$	$\pm 10$	7	19
TRO 4/1/0,25	250	4-6	360	-	90	9
THLPD 18	100	5-30	60	-	5-60	12
HQ 1/0,3	100	6-7	360	-	50	6

A rövidhullámok tehát távoli, nagy felületű területek besugárzására használhatók. A távolsági ellátottságot főleg az ionoszféra F2-rétege biztosítja a réteg és a földfelszín közötti visszaverődéssel. Az átvitel hatótávolságát azonban jelentősen korlátozzák azok a veszteségek, amelyek a terjedés útján fellépnek. A kozmikus és földmágneses hatások a rövidhullámú terjedés erős zavaraihoz vezethetnek. A térhullám terjedése során olyan fadingjelenségek léphetnek fel, amelyeket különböző okok hozhatnak létre (például polarizációs elfordulás, az elnyelés ingadozásai az ionoszférán, többutas terjedés stb.) A naptevékenység által előidézett erős ionizáció határozza meg a rétegek refrakciós viszonyait. A rövidhullámú jel refrakciója, azaz törése a terjedési rétegen függ a jel beesési szögétől, a rétegben való ionkoncentrációtól és a jel frekvenciájától. A terjedési feltételeket nem lehet befolyásolni, de célirányos antennákkal és legkedvezőbb meghatározott frekvencia alkalmazásával a nagy hatótávolságra történő sugárzás biztosítható.

A bővített jászberényi rövidhullámú rádióállomásnál a várható kedvező frekvencia használati értékeket három módszerrel határoztuk meg a kívánt főirányokban lévő célterületekre:

- a.) a klasszikusnak nevezhető CRPL-módszerrel;
- b.) az újabban alkalmazott statisztikai módszerrel;
- c.) számítógépes programmal.

### 3.1. A CRPL-módszer alkalmazása

Alkalmazása, számítása nagyon hosszadalmas és időigényes; számos segédlet, diagram és mellékszámítás elvégzése szükséges (világövezet-térkép; nagykörös térkép; F2(0) MUF; F2 (4000) MUF térképek különböző zónákra; a napfoltmaximum és minimum értékeire; F2 MUF és FOT diagramok; napfoltszámot és napzenitszöget meghatározó diagramok; geomágneses veszteségi térképek, csillapítás diagramok stb. alkalmazása szükséges). A frekvenciaszámítások mind az öt főirányra, tehát Európa; USA-Kanada; Dél-Amerika;

Ausztrália-Új-Zéland; Közel-Kelet, Szaud-Arábia irányokra napfoltmaximumos és napfoltminimumos nyári és téli időszakra elkészültek. A táblázatok utolsó oszlopa (RH-sáv) a várható kedvezően használható üzemi frekvenciasávot tartalmazza.

A 4. táblázat a számítások alapján bemutatja nyári időszakra és napfoltmaximumos időszakra, az USA-Kanada felé (316°-os főirány) legkedvezőbb használható frekvenciasávokat.

### 3.2. Statisztikus módszer alkalmazása

Újabban a frekvenciaszámításoknál alkalmazzák a statisztikai módszert. Ezt csak abban az esetben lehet alkalmazni kedvezően, ha legalább három napfoltmaximumos és -minimumos időszakra visszamenőleg ismeretes a kívánt célterületre a használt üzemi frekvenciák értéke és a célterületekről igen jelentős (nagy számú) vételmegfigyelés, értékelés is rendelkezésre áll. Ezek alapján a jövőben várható napfoltmaximumos és napfoltminimumos időszakra, bizonyos kisebb kiegészítő, korrekciós számításokkal az új üzemi frekvenciaértékek meghatározhatók.

Ez a módszer részünkre kedvező, mert 30 évre visszamenőleg rendelkezésre állnak az üzemi frekvenciák, és a célterületekről igen sok vételmegfigyelés (levél, magnófelvétel, személyes tájékoztatás) a Magyar Rádióon keresztül szintén rendelkezésre állt. A nevezett célterületekre az üzemifrekvencia-értékek ismertek és a vételkiértékelések a Magyar Rádió Külföldi Adások Főszerkesztősége által évente kiadott, úgynevezett Vételértékelés (nyári és téli időszakokra a hallgatóktól beérkezett vételmegfigyelések) kiadvány révén szintén rendelkezésre állnak, tehát a várható kedvező üzemifrekvencia-értékeket meg lehetett határozni.

A frekvenciaszámítások a fentiek alapján szintén megtörténtek az öt főirányban lévő területekre, téli-nyári időszakokra, napfoltminimumos és -maximumos időszakokra, és táblázatos formában kerültek összefoglalásra.

4. táblázat  
USA-Kanada felé optimálisan kedvező frekvenciasávok (d=10.000 km, R=160 max., nyári időszak)

GMT [óra]	Budapesti oldal A			USA oldal B			A MUF [MHz]	B MUF [MHz]	ut MUF [MHz]	ut FOT [MHz]	RH sáv [MHz]
	Nap szög	F2(4000) MUF [MHz]	F2(4000) FOT [MHz]	Nap szög	F2(4000) MUF [MHz]	F2(4000) FOT [MHz]					
	A'	A''	B'	B''							
00	-	22	18,7	-	25	21,3	22	25	22	17,7	17
02	-	20	17	-	26	22,1	20	26	20	15,3	15
04	-	22	18,7	-	24,6	20,9	22	24,6	22	17,7	17
06	-	21,5	18,3	-	21	17,9	21,5	21	21	17,7	17
08	-	21	17,9	-	17	14,5	21	17	17	11,9	11
10	-	20,5	17,4	-	16	13,6	20,5	16	16	11,9	11
12	-	21	17,9	-	16	13,6	21	16	16	11,9	11
14	-	20	17	-	18,5	15,7	20	18,5	18,5	15,3	15
16	-	21	17,9	-	20,5	17,4	21	20,5	20,5	15,3	15
18	-	23	19,6	-	21,5	18,3	23	21,5	21,5	17,7	17
20	-	22,5	19,2	-	22,5	19,2	22,5	22,5	22,5	17,7	17
22	-	22	18,7	-	24	20,4	22	24	22	17,7	17

Sugárzási terület		Adás helye: Jászberény Rádióállomás						
Dél-Amerika	Helyi idő	UTC idő	Jan./Febr. és Nov./Dec.		Márc./Ápr. és Szept./Okt.		Május - Augusztus	
			Max.	Min.	Max.	Min.	Max.	Min.
			160	10	160	10	160	10
D = 13.000 km	00-04	04-08	11	7	9	9	11	7
	04-08	08-12	17	17	11	11	21	11
	08-12	12-16	21	17	17	17	21	17
	12-16	16-20	17	11	17	15	21	17
	16-20	20-00	15	9	17	15	15	15
	20-00	00-04	9	7	11	9	15	9

5. táblázat

Optimálisan használható üzemi frekvenciasávok (MHz) Dél-Amerika részére (Földrajzi körzetek: 12-13-14-15)

Az 5. táblázat a számítások végeredményei alapján bemutatja a például Dél-Amerika felé történő adások kedvező üzemi frekvenciasávok használatát a nap 24 órájában.

Rövidhullámú műsoradásoknál a vételi területeket úgynevezett zónatérképek alapján adják meg a Nemzetközi Rádiószabályzat szerint, amelyet a 4. ábrán látható térkép tartalmaz.

#### 4. Térerősség átlagértékének számítása

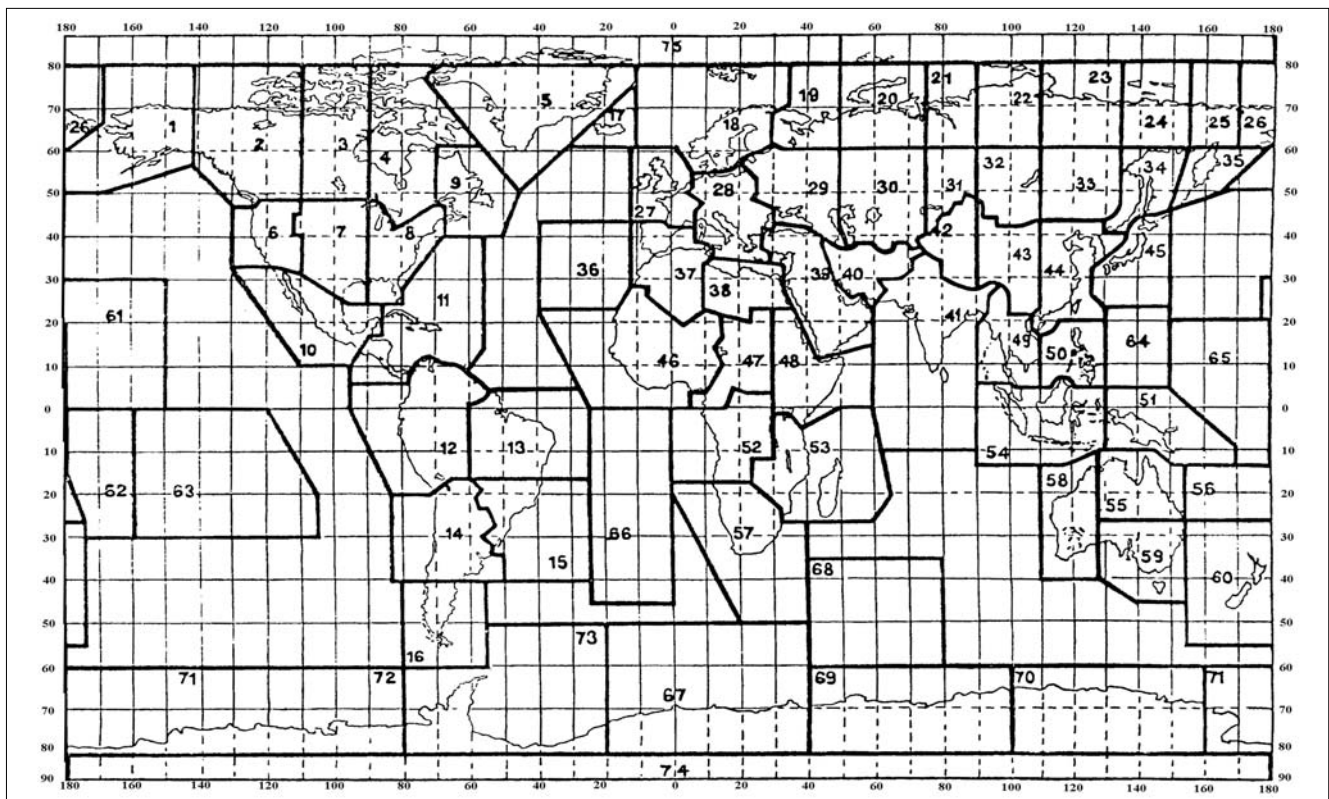
A számítás a CRPL-módszer, valamint a CCIR 252-2 (New Delhi, 1970) összefoglaló anyagán alapul, amely 10 különböző módszerre épül fel. A térerősségszámítás fő problémája abból adódik, hogy a Föld két pontja között azonos kilövellési szög esetében több terjedési út lehetséges egyidejűleg. Az ionoszféra egyes ré-

tegei a rádióhullámokat nemcsak reflektálják vagy át-eresztik, hanem el is hajlítják, s ezzel az ionoszférában való hosszabb-rövidebb haladásra készítetik, majd – a helyi jellemzőktől függően – reflektálják vagy át-eresztik. Mindezek miatt számos, úgynevezett kevert út valósulhat meg, különböző ugrástávolságokkal.

Ahhoz, hogy megállapítható legyen, hogy a lehetséges terjedési utak közül melyik valósul meg, vagy melyik úton áll elő a legkisebb csillapítás, megbízható hosszútávú előrejelzés számítás szükséges, illetve meg kellene határozni minden számításba jöhető terjedési útvonal csillapítását. Ilyen számítások csak speciális számítógépes programmal lehetségesek. A hosszú távú előrejelzések nem állnak mindenben rendelkezésre. A manuális számítási eljárásokat egyszerűsíti az igen sok diagram, nomogram és táblázatok használata.

4. ábra

Földrajzi övezetek a rádió-műsorszórásnál



Az egyes célterületekre meghatároztuk a térerősség átlagos értékét, különböző antennák figyelembevételével, 250 kW, illetve 100 kW adóteljesítménnyel. A számítások részleteit terjedelmi okok miatt e cikkből ki kellett hagynunk.

A térerősség-számítások eredményeiből látható, de az eddigi üzemi tapasztalatok (a vételmegfigyelések és a kiértékelések) is igazolják, hogy a HRRS 4/4/1 síkantennák a 250 kW-os adóberendezésekkel az időszakoknak megfelelő üzemi frekvencián biztosítják kedvezően a három nagytávolságú célterület besugárzását (USA-Kanada; Dél-Amerika; Ausztrália, Új-Zéland).

## 5. A frekvencia- és térerősség-érték számítógépes meghatározása

Az öt főirányú célterületre a használható kedvező frekvenciák, valamint a várható térerősségek alakulása számítógépes programmal is meghatározásra kerültek napfoltminimumos és napfoltmaximumos időszakokra, melyek eredményeit több mint 160 térkép tartalmazza. Sajnos, ezekből helyhiány miatt jelen cikkben nem tudunk közreadni.

## 6. Összefoglalás

- a.) Feltétlen szükséges – a korábbi tapasztalatok ezt alátámasztják –, hogy a várható aktuális sugárzási időszakra minden évre előzetesen, a rövidtávú előrejelzések ismeretében a nyári-téli és tavaszi-őszi időszakokra a célterületekre frekvenciahasználati terv készüljön a várható ionoszféraállapot és a pontosabb napfoltszám ismeretében, korrigálva a hosszútávú frekvenciahasználati tervet. A sugárzási időszakokra a Rádió Igazgatásoknak a Rövidhullámú Koordinációs Konferenciákra (HFCC-WARC '97 alapján) az aktuális rövidhullámú sugárzási frekvenciaóraterveket előzetesen el kell készíteni, és egyeztetni kell nemzetközi szinten, tehát feltétlen szükséges előtte a célterületekre vonatkozó sugárzási idő szerinti frekvenciaterv elkészítése, véglegesítése.
- b.) A frekvenciaszámítások eredményei, de a korábbi vételi megfigyelések eredményei alapján bizonyos időszakokban az európai területre igen kedvező a 9-11 MHz-es sáv használata. Ezért feltétlen célszerű és kedvező lenne, ha a jászberényi állomásra a 9-11 MHz-es körsugárzó szögletantenna-rendszer is áttelepítésre kerülne a közeljövőben, mert ezáltal nagymértékben növelni lehetne az európai terület besugárzásának hatékonyságát. A két HQ 1/0,3 típusú körsugárzó szögletantennával (6-7 MHz, illetve a 9-11 MHz-es) 100 kW teljesítménnyel körülbelül 1600 km-es terület, a TRO 4-6 MHz-es körsugárzó antennával 250 kW-os adóval egy 500...600 km sugarú terület igen kedvezően besugározható. A forgatható és dönthető log-per antennával a 100 kW-os adórendszerrel pedig bármilyen újonnan felme-

rülő európai, észak-afrikai vagy közel-keleti sugárzási igény nagyon eredményesen és azonnal biztosítható (rugalmas sugárzási lehetőség).

- c.) Ausztrália és Új-Zéland sugárzási ellátásának javítása érdekében célszerű lenne, hogy az észak-nyugati síksorban lévő 65°-os főirányítású HRRS 4/4/1 típusú síkantennánál legalább a 17 és 15 MHz-es sugárzók esetében a  $\pm 22^\circ$ -os vízszintes karakterisztikabillentés megvalósuljon, mert ezáltal a napfoltmaximumos időszakokban Nyugat-Ausztrália területére kedvezőbb, illetve hatékonyabb lenne sugárzásuk, sőt  $+20^\circ$ -os billentéssel  $284^\circ$ -os főirányú sugárzással Közép-Amerika felé is használni lehetne az antennát.
- d.) A jászberényi rövidhullámú rádióállomáson a meglévő nagy- és középtávolságú síkantennákkal, valamint a kibővült körsugárzó és a forgatható, dönthető log-per antennákkal igen hatékony sugárzást lehet biztosítani a kívánt célterületekre. Feltétlen mielőbb meg kellene kezdeni a DRM-sugárzást is az állomáson. A DRM-adások gazdasági megtakarítást jelentenének, és jobb, nagyobb ellátottságot is biztosítanának.
- e.) A WRC'03 – vagyis a DRM hivatalos elindulása – óta a rövidhullámú műsorszóró sávban a digitális sugárzásra bejelentett igények száma évről évre, pontosabban a nemzetközi koordinációnak megfelelően félévről félévre növekszik. A cikk írásakor már több mint 100 rendszeres DRM adás van világszerte, ezek igen nagy arányban a rövidhullámú sávban zajlanak. A frekvenciák nemzetközi koordinációja során egyre több a DRM-re igényelt rövidhullámú spektrum. 2006-ban megjelennek a tömeggyártott vevők, ezáltal bizonyosak lehetünk abban, hogy nemcsak az adások, de a hallgatók száma is növekedni fog.

A székesfehérvári rádióállomás Jászberénybe történt áttelepítése révén Magyarországnak ez az egyetlen RH adóállomása maradt, ez azonban – az előbb leírtak alapján – igen kedvező lehetőségeket egyesít mind a Kárpát-medence, mind Európa, valamint az Európán kívüli területek lefedésére.

Az állomással megvalósítható feladatok három csoportra oszthatók:

- Magyarország területének lefedése tropikus sugárzóval
- Külföldön élő magyarok ellátása hazai műsorról
- Külföldi műsorstúdiók jelének sugárzása külföldi területre (bérsugárzás).

Igen kedvező elméleti lehetőség adódik az ország területének lefedésére a 4, illetve 6 MHz-es TRO 1/0,3 antennák használatával. Az újonnan áttelepített BBC (ma Thales) adók már PSM üzeműek, tehát egy DRM modulátor beiktatásával és egy minimális átalakítással digitális sugárzásra alkalmassá tehetők. Rövidhullámon a B, C vagy a D üzemmód alkalmazható; B mód, 10 kHz RF sávzélesség esetén 21 kbit/s, D módban 11 kbit/s érhető el – ez monó FM minőségnek felel meg – természetesen fading nélkül.



5. ábra Thales DRM modulátor

Ahhoz, hogy az elméleti lehetőségből gyakorlati lehetőség váljon, az szükséges, hogy a sugárzási rendszerben az egyik 100 kW-os BBC adó mindig szabad legyen, valamint hogy az adó bemenete elé kerüljön egy Thales típusú DRM modulátor. Egy ilyen modulátor képe látható az 5. ábrán.

A többi, sokrétű feladat ellátásához Jászberény rádióállomás rugalmasságát lehetne növelni azáltal, hogy a jelenlegi kombinátormátrixot egy olyan mátrixra cserélik, amely lehetővé teszi, hogy bármely adó bármely antenára üzemeljen. Ebben a kiépítésben ugyanis gyakorlatilag a két állomás rendszere egy helyen, de egymástól függetlenül működik, tehát nincs lehetőség arra, hogy például egy 100 kW-os adó valamelyik függőny-antennával sugározzon.

Abban az esetben, ha a jelenlegi – már több mint 30 éves! – 250 kW-os adóberendezések cseréjére is sor kerülhetne, igazán korszerű és rugalmas rövidhullámú állomás láthatná el az országot, a Kárpát-medencét, Európát, Afrikát, Ausztráliát, Ázsiát és Amerikát célterületeit.

Noha az Internetet és a műholdat sokan emlegetik mint a rövidhullám konkurenciáját, de ne feledkezzünk meg arról, hogy az internetes rádió csak korlátozott számú hallgatóság ellátására képes, a műholdas rádiózás pedig (az USA-t kivéve) helyhezközött tevékenység, tehát a rádiózás egyik legnagyobb előnye veszne el a rövidhullám elvetésével: a hordozható és a mobil vételi lehetősége.

## Irodalom

- [1] A Magyar Rádió külföldre szóló adásainak értékelése. 1970. nyári időszak.
- [2] A Magyar Rádió külföldre szóló adásainak értékelése. 1971. téli időszak.
- [3] HF Transmitting and receiving antennas characteristics and diagrams. ITU Doc., 10 D (Temp) 7 1990. 1994.
- [4] Kenneth Davies: Ionospheric Radio Propagation. National Bureau of Standards Monograph 80., 1965.
- [5] Dr. Szokolay Mihály: Rádió rendszertechnika. BME Villamosmérnöki Kar, MTI Kiadványa, Tankönyvkiadó Budapest 1970.
- [6] Ionoszféra Adatok – 1987. június, Békéscsaba. Magyar Meteorológiai Szolgálat
- [7] Ionospheric Predictions 1971. jan-febr. National Bureau of Standards C.RPL.
- [8] Műsorszóró szolgálatnak kiosztott RH-sávok tervezésével megbízott Igazgatási Rádió Világértekezlet (HFBC-87). Záróokirat, Genf 1987., KDV Budapest 1988.
- [9] CCIR 1966 Oslo Report 340. CCIR Atlas of Ionospheric Characteristics. UIT Genf 1967.
- [10] Dósa György: Rövidhullámú műsorszóró adásainak vételi lehetősége a világ különböző területein. PRTMIg Modulátor, 1996. XI.
- [11] Dr. Turi-Kováts Attila, Molnár Béla: Rádió-hullámterjedés és hálózattervezés. Közdok Budapest 1969.

## A hazai rádiózás születése és kezdetei

DÓSA GYÖRGY

*A 19. század végén az első sikeres rádiókísérletek – szikra-adások – jelentős érdeklődést keltettek a világon, így Magyarországon is. A Magyar Posta szakemberei már az első nemzetközi kísérletek eredményeinek ismertté válásakor (Marconi, Braun, Popov, Slaby, Arcó) felfigyeltek a szikratávíróra és már 1896-ban a Magyar Posta tudományos intézetében, a Posta Kísérleti Állomáson megkezdődött a drótnélküli átviteli rendszer tanulmányozása, vizsgálata. A posta szakemberei tanulmányutakon és nemzetközi konferenciákon vettek részt és ismerkedtek az új fejlesztésekkel, tapasztalatokat szereztek, hogy a tudományos munkájukban felhasználhassák.*

1903-ban az első kísérleti összeköttetést a vásárolt Slaby rendszerű szikraadóval és a koherens vevőkészülékkel Csepel és Újpest között eredményesen bonyolították le. A Kolozsváry Endre, Hollós József és Tolnai Henrik által vezetett szakember csoport 1904-ben a módosított, illetve tökéletesített szikraadó és vevőberendezéssel Budapest-Bécs között folytattak távíró összeköttetési kísérleteket, melyek igen sikeresek voltak.

A szerzett tapasztalatok alapján 1906-ban már a Posta Kísérleti Állomás irányításával az Adria tengeren hatottak végre újabb rádiókísérleteket az előre hajóra szerelt mozgó állomás és a fiumei parti állomás között, jórészt hazai gyártású berendezésekkel (elektrolitikus detektor alkalmazásával). Az áthidalt távolság a parti állomás és a hajó között – Ancona térségében – körülbelül 250 km volt. Az összeköttetés (távíratváltás) igen kedvező eredményt adott. Feltétlen meg kell említeni, hogy az olajban csillapított szikraközzel együtt jelentős új megoldás volt a Szvetics Emil cége által gyártott nagy átütési szilárdságú papírkondenzátorok alkalmazása a szikraadónál.

A megbízhatóan működő adó és vevő berendezéseket a posta szakemberei tovább korszerűsítették és különböző helyeken kipróbálták s ennek eredményeként a tízes évek legelején már az ország bármely két pontja között biztosítani tudtak rádióösszeköttetést.

A nemzetközi és hazai szikraadó- és vevő-kísérletek egyértelműen bizonyították, hogy az elektromágneses hullámok alkalmazásával egy új hírközlési rendszer jött létre, az úgynevezett szikratávíró adásrendszer. A rádiózás első évtizedei a rádiókommunikáció kezdetét jelentette.

A szikraadók – szikrainduktoros adók – voltak a legegyszerűbbek, melyeknél gyorsan csillapodó rezgések keletkeztek. A sok harmonikus, széles sávban zavaró hatású volt. Az alapproblémát a túl gyorsan csillapodó rezgések jelentették. Javítást jelentett, hogy egyetlen szikraköz alkalmazása helyett több szikraközt alkalmaztak sorbakapcsolással, légmentes kialakításban, ezáltal az ionizációs viszonyok jelentősen javultak

és a szikra hosszabb ideig maradt fenn. A szikraadók számának és a teljesítményének növekedésével azonban egyre jobban mutatkoztak meg a csillapodó rezgések miatti problémák, a széles frekvenciasáv, melyet egyes állomások elfoglaltak és ezért erősen zavarták a többi állomás vételét. Emiatt a csillapodó rezgések helyett a csillapítatlan rezgések vizsgálata és a alkalmazása került előtérbe.

A csillapítatlan rezgésű adórendszerek első típusa az ívfény adó volt, ahol az ívlámpához kapcsolt rezgőkörben csillapítatlan rezgések keletkeztek és ezekkel nagyobb távolságra lehetett távíró összeköttetést biztosítani. A bonyolult üzemeltetésen túl a fő probléma az volt, hogy 300-600 kHz-nél rövidebb hullámokat nem lehetett ezen berendezésekkel előállítani. További problémát jelentett, hogy az adóteljesítmény növelése miatt az elektródák megfelelő hűtése is megoldatlan volt. Ezért az ívfény adót a korszerűbb kialakítású nagyfrekvenciás, úgynevezett gépadó rendszer váltotta fel. Ez esetben a rezgőkör táplálására a nagyfrekvenciás áramot külön e célra alkalmas forgógéppel állították elő.

A rádiós kísérletek eredményei alapján a katonai vezetés is egyre nagyobb érdeklődést mutatott a rádiózás iránt. A tízes évek elején már több ország hadserege fel volt szerelve rádióadó-vevő berendezésekkel. Ez különösen az egyazon szövetségi kötelékbe tartozó, de egymástól földrajzilag távollévő országok számára volt nagyon fontos követelmény. Ilyen helyzetben voltak a központi hatalmak országai, nevezetesen Ausztria, Magyarország és Németország, amelyek el voltak vágva két szövetségesüktől, Törökországtól és Bulgáriától. Sürgősen felmerült tehát az igény, hogy a Monarchia Törökországgal és Bulgáriával távíró útján teremtsen állandó kapcsolatot. Erre a leggyorsabb megoldást a rádiótávíró rendszer nyújtotta. A magyar politikai és katonai vezetés döntött, hogy Magyarországon sürgősen meg kell kezdeni egy drótnélküli rádiótávíró állomás kiépítését.

Az első helyhez kötött magyar rádióállomás konkrét felépítéséhez tehát az első világháború kitörése szolgáltatott okot. Az állomás elsősorban katonai rendelte-

tésűnek indult, de a diplomáciai levelezés is tervezve volt, továbbá külön feladatául tűzték ki, hogy a Dunán közlekedő hajókkal is kell rádióforgalmazást biztosítani. Az egyesített hadvezetésnek feltétlen érdeke volt tehát, hogy más összeköttetés hiányában csak rádiótávíró útján teremtsen gyors kapcsolatokat a szövetségesekkel. Ennek érdekében az első világháború kitörésekor a magyar kormány megbízásából oltószikrás távíró adót és szintén Telefunken gyártmányú kristálydetektoros vevőberendezést rendelt meg. Az új rádióállomást – mely a hazai rádiózás bölcsője lett – Csepel községtől északra, a Csepel szigetén jelölték ki, mintegy 60 kataszteri hold területen.

Az állomás építése 1914. július 28-án kezdődött meg az antennahuzalokat kifeszítő segédárbcok (12x45 m magas faoszlopok) felállításával. Ezek tartották az ernyő-antenna egyes szálait a középső, 120 méter magas „Rendal” rendszerű vasszerkezetű toronnyal. A teljes antennarendszer és a földhálózat 1914. szeptember 6-ára készült el. Közben folytak az adó- és vevőkészülékek belső szerelési munkálatai is. A villamosenergiát a Fővárosi Elektromos Művek szolgáltatta, az akkumulátor telepeket a Tudor Gyar, a kapcsoló táblát és a világítási berendezéseket a Siemens Művek szállították. 1914. október közepére üzembe állt az antenna-

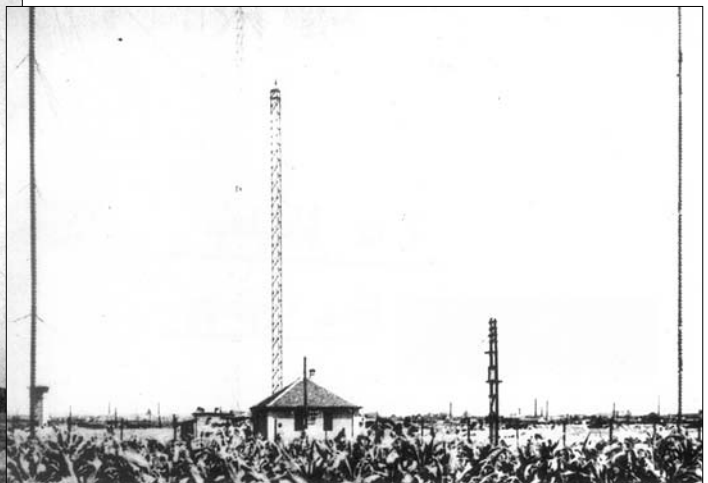
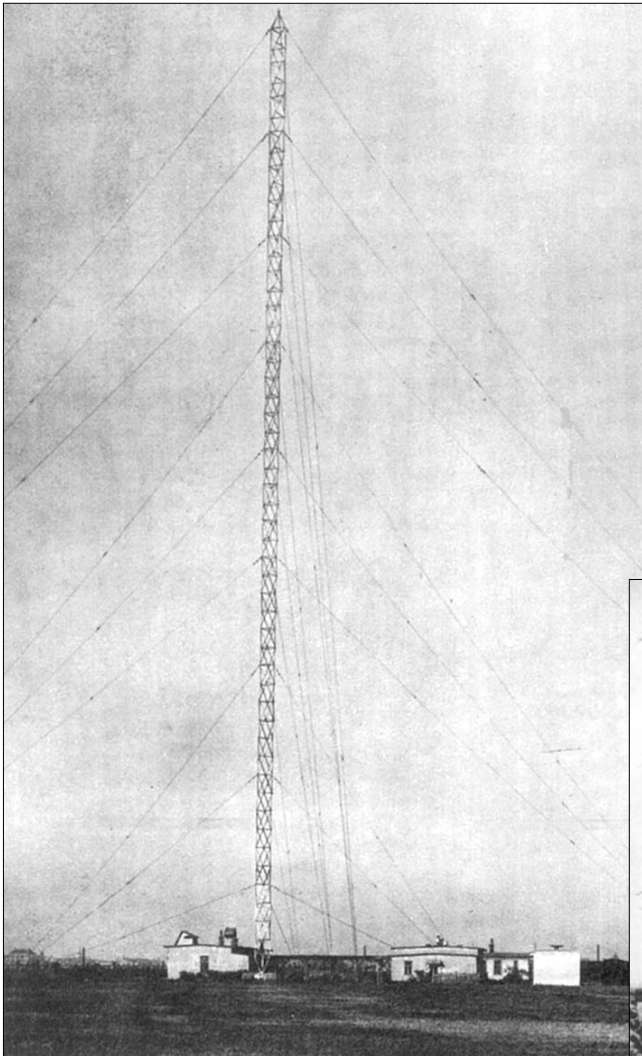
rendszer és a Telefunken rendszerű kristálydetektoros vevő és adókészülék is. Ezt követően, 1914. október 14-én, három hónappal a háború megkezdése után megkezdte működését az első magyar rádióállomás. Az állomás vezetője rádiózásunk egyik úttörője, Tolnai Henrik lett. A katonai és diplomáciai távíró forgalmazás 1914. november elején indult meg erőteljesen.

A vevőkészülék 1400-4000 méteres hullámsáv vételére volt alkalmas tekercs-csere nélkül. A vétel javítása érdekében 1915-ben a vevőrendszert egy Lieberens rendszerű erősítővel is felszerelték. A 120 méter magas, úgynevezett ernyőantennát felváltva lehetett vételre és adásra kapcsolni. Az ernyőantenna 24 szál, 150 m hosszú, 3 mm átmérőjű bronzhuzalból állt. A földhálózat 72 db sugárirányú, 300 m hosszú és 3 mm átmérőjű huzalból készült, mintegy 80 cm mélységbe fektetve.

A lenti fényképek mutatják be az állomás épületeit és az antennarendszert. Az állomáson megindulás után még egy 30 kW-os Pulsen rendszerű ívlámpa adó-vevő berendezést is üzembe állítottak. A csepeli rádióállomás forgalma az üzembe helyezés után folyamatosan növekedett, szinte valamennyi európai állomással létesített rádió kapcsolatot, de főleg Béccsel, Konstantinápolyal, Moszkvával, Szentpétervárral, sőt Barcelonával, Madriddal és Aranjuez spanyol állomással is. 1916. október 15-től pedig a svédországi Boden állomással rádió távíró kapcsolat is létrejött. 1917. februárjától pedig a párizs és Lyon-i adók hírnap-táviratait vették fel.

Hamarosan megjelentek a hadszínterek állásáról szóló hivatalos hadijelentések, az úgynevezett Hófer-jelentések is. Ezeket naponta meghatározott időben sugározta a csepeli állomás „CQ”-ban, azaz mindenkinek szóló adásban. A távirati forgalom lebonyolításán kívül venni kellett az állomásnak a német Nauen-i, majd később a Kóningswurstenhausen-i rádióállomások által leadott német hadijelentéseket, továbbá az Eiffel-torony rádióállomása által sugárzott francia, majd később a Poltanó-i állomás által sugárzott olasz hadijelentéseket is. Ezek vétel szempontjából jelentősen növelték a csepeli állomás forgalmát.

Igen jelentős volt az 1916. esztendő második felében meginduló hadifogoly-rádióforgalom. A Nemzetkö-



zi Vöröskereszt útján bárki kereshette eltűnt hozzátartozóját. A nevezett személyi adatokat és a kérést tartalmazó táviratot a csepeli állomás útján a semleges svédországi Boden-i rádióállomásnak sugározta, mely a táviratokat Szentpétervárnak továbbította. A válaszok ugyanezen az úton érkeztek Magyarországra. 1917. novemberétől a Csepel rádióállomás jelentős forgalmat bonyolított le Szentpétervárral a béketárgyalások megkezdése céljából is. A háború végén pedig rendszeres összeköttetés alakult ki Moszkvával és nyugati városokkal, Párizssal és Münchennel.

Az állomás indulásakor az adóberendezés billentyűzése az állomáson történt, később azonban az állomás és a Központi Távirda között, úgynevezett Hughes-összeköttetés épült ki. Magyar szempontból történeti jelentőségű volt az állomásnak a párizsi béketárgyalásokkal kapcsolatos feladata, szerepe. A magyar kormány és a párizsi, illetve a trianoni béketárgyalások vitelére kiküldött kormánydelegációval a csepeli rádióállomás útján tartott állandó kapcsolatot és így adta utasításait a küldöttségnek.

Az állomás ezzel kapcsolatos forgalma 1920. január 8-án kezdődött. Duplex üzemből dolgoztak, vagyis egyidejű adás és vétel történt különböző hullámhosszon. A párizsi állomás ettől kezdve állandó levelezésben (összeköttetésben) állt a csepeli állomással.

A Magyar Posta vezetése a háború végén hozott döntése alapján csillapítatlan hullámú adók alkalmazásával tervezte a csepeli állomást továbbfejleszteni. Így 1919. januárjában a Telefunken cégnél megrendelt egy korszerű 5 kW-os „lámpa”-adót és egy keretantennás vevőkészüléket is. Az 5 kW-os adóberendezés három egységből állt. A nagy higanygőz egyenirányító táplálta a 12 db párhuzamosan kapcsolt 500 W teljesítményű adócsőből felépített erősítőt, amelyben a csövek külön-külön egységbe építve működtek. A harmadik egység a közbeeső kört és az antenna hangoló elemét tartalmazta. Egy átkapcsoló rendszerrel a régi 24 szál as ernyőantenna erre az új adóra, vagy a régi szikraadóra volt kapcsolható.

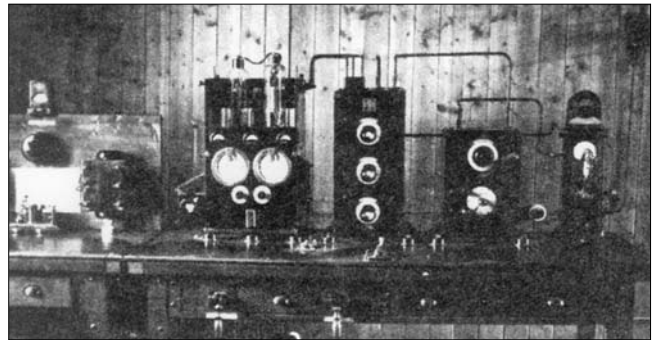
Ezen új 5 kW-os lámpaadó 1921. október 15-én lépett üzembe és igen rövid időn belül Európa valamennyi nagyobb rádióállomásával forgalmazott. Az 5 kW-os lámpaadó átlagosan napi 18-22 órát üzemelt és 3000 km-es távolságra is kedvező összeköttetést biztosított.

A kisebb távolságú forgalom biztosítására – 1000 km-es zóna – a Magyar Posta vezetése terve vette, hogy egy kisebb teljesítményű lámpaadót is üzembe állít a csepeli rádióállomáson. Alátámasztotta ezt az igényt az is, hogy az időközben megalakult Magyar Légiforgalmi Rt. (MALÉRT) is igényt tartott repülőtéri légiforgalmi rádiószolgálatra. A 7,5 kW-os első szikraadó a lámpaadó üzembe állítása után még jó ideig meteorológiai híreket sugárzott, míg 1924-ben üzemét megszüntették.

A meginduló hazai légiforgalom tehát szükségessé tette, hogy a repülőterekkel és esetleg a repülőgépekkel megfelelő hírcsoporsó jöjjön létre. Ennek érdekében 1923-tól a Mátyásföldi Repülőtér posta hivatalának egyik helyiségében rádió-kirendeltséget hoztak lét-

re. 1923. júniusától az egyik csepeli 250 W teljesítményű HUTH adó és a Mátyásföldi Repülőtér kirendeltségén üzembe állított vevőberendezés bonyolította le a rádiótávíró kapcsolatokat a környező államok repülőtereivel. A növekvő légiforgalom (a légiforgalmi meteorológia és a szolgálati táviratok) miatt csepelen egy új 1 kW-os Telefunken távíró adót kellett a repülőtéri összeköttetések szolgálatába állítani.

Az adóberendezést az *alábbi fénykép* mutatja be. Az adó antennája 40 méter magas T antenna volt a nagy antenna oldal árbócára történő felfüggesztéssel.



A forgalom fokozatos növekedése miatt szükségessé vált, hogy az adás és vétel ebben az esetben is különválasztásra kerüljön. Csepelen ugyanis adás alatt a vétel szünetelt, illetve vétel alatt az adás szünetelt. 1919-ben a Kísérleti Állomáson egy önálló vevőállomást és üzemközpontot hozott létre a Magyar Posta és így a vételi szolgálat fokozatosan áttevődött a Kísérleti Állomás új vevőközpontjába. A csepeli adóállomás és a Gyáli úti vevő és távíró központ külön HUGHES betűnyomós gépek útján érintkezett.

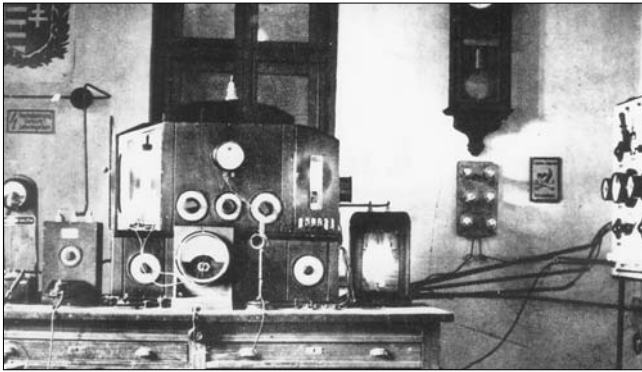
A húszas évek elején mind jobban előtérbe került, hogy nemcsak távíró jeleket, hanem emberi hangot is továbbítsanak rádió útján, és megkezdődtek az úgynevezett rádiótelefon vizsgálatok, fejlesztések. Ezen a területen a legjelentősebb volt, hogy 1906-ban megjelent Lee de Forest amerikai fizikus által feltalált három elektrodás cső, a trióda. A trióda tette lehetővé, hogy létrejöttek a nagyobb teljesítményű lámpaadók és csillapítatlan rezgéseket lehetett stabilan létrehozni.

Magyarországon 1923-ban megkezdődtek az ilyen irányú kísérletek, ezért a csepeli 5 kW-os lámpaadót a Magyar Posta a Telefunken céggel alkalmassá tette rádiótelefon üzemre és több kedvező kísérleti összeköttetést, forgalmazást folytattak Berlinnel. A kísérletek kedvező eredményei alapján Magyarország megvásárolt a berlini HUTH gyártól 2 db 250 W teljesítményű távíró és távbeszélő üzemre is alkalmas adóberendezést. Az adóberendezések műszaki átvételét és hazai üzembe állításukat Erdős Gyula és Jurcsek Viktor mérnökök bonyolították le. Az egyik adóberendezés a csepeli rádióállomáson, a másik adóberendezés pedig a soproni Postaigazgatóság épületében került üzembe helyezésre.

A csepeli 250 W-os HUTH adóberendezéssel 805 m hullámhosszon kezdődött meg 1923. november 6-án a rendszeres MTI hírszolgálatának a sugárzása a hazai MTI kirendeltségek – Debrecen, Pécs, Szombathely,

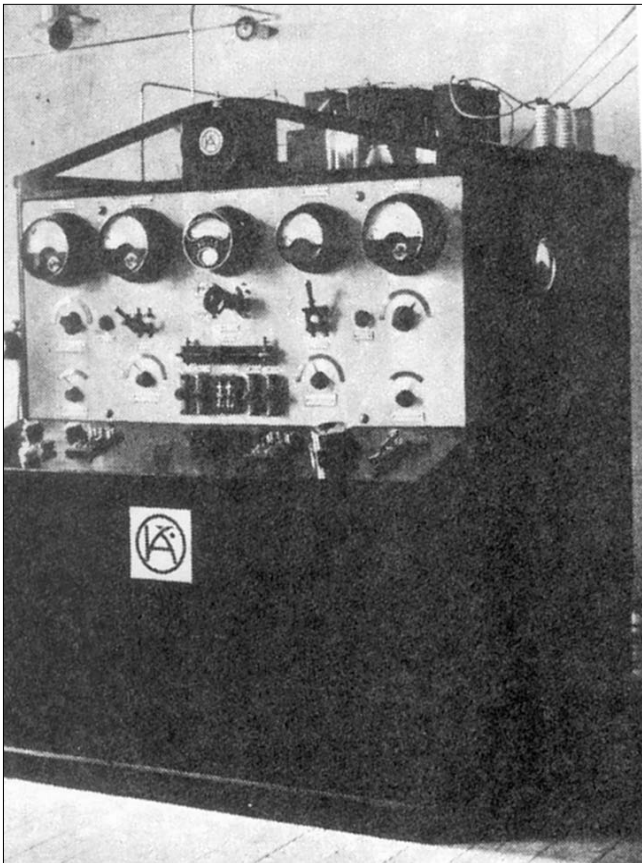


Miskolc és Győr – részére, majd későbbi időpontban a tőzsdei jelentéseket és az MTI nemzetközi híreit is sugározta. A fényképen a 250 W-os HUTH adó látható.



A HUTH adó kis teljesítménye szükségessé tette, hogy az MTI szolgáltatás javítására egy nagyobb teljesítményű adóberendezés kerüljön üzembe Csepelen. A PKÁ (Posta Kísérleti Állomás) két mérnöke – Magyar Endre és Tomcsányi István – vezetésével megtervezte és megépítette az 1 kW-os első teljesen hazai kivitelezésű táviró-távbeszélő adóberendezést, mely 1925. elején lépett üzembe a csepeli állomáson. A *lenti fényképen* az első magyar gyártású 1 kW-os telefon-távíró adóberendezés látható.

Az MTI híreit a rádióamatőrök kristálydetektoros vevőkkel jól vették és a vételi tapasztalataikról a PKÁ szakembereinek is beszámoltak, ezért felmerült az a lehetőség, hogy a 250 W-os HUTH adóval műsoradási kísérletek is történjenek a postai szakemberek irányítá-



sával. Az első ilyen konkrét kísérlet 1924. márciusában történt a 250 W-os HUTH adóval, mely kedvezően sikerült. Ez azonban már egy új korszak kezdetét jelentette.

A csepeli rádióállomás nemcsak a rádiótávírónak, hanem a magyar rádió távbeszélőnek és a rádió műsorszórás kísérleti kezdetének is, tehát a hazai rádiózásnak a bölcsője. 1914-től kezdve két évtizeden keresztül állt a hírszolgálat, a közgazdaság, a kultúra és a honvédelem szolgálatában, míg 1934-ben a Magyar Posta vezetősége üzemét megszüntette. Új, korszerű rádióállomások léptek üzembe, mint Székesfehérvár, Lakihegy és Tárnok vevőállomás.

A hazai rádiózás a húszas évek elejére teljesen felzárkózott a nemzetközi élvonalhoz és kimagasló eredményeket ért el a rádiózás fejlesztés gyakorlati területein.

Végezetül kegyelettel emlékezzünk azon szakemberekre, akik a Csepel Rádióállomás létrehozásában, munkálataiban és üzemeltetésében kimagaslóan részt vettek: Kolozsváry Endre, Hollós József, Jurcsek Viktor, Erdős Gyula, Magyar Endre, Tolnai Henrik, Gasparik László, Novák Károly, Léber József, Lászlófi István, Kénoszt Rezső.

#### Irodalom

- [1] Fata Lajos, Kénoszt Rezső, Marczal János: Személyes elbeszélések alapján, 1955-1965 között.
- [2] Tízéves a Magyar Rádió. Magyar Királyi Posta – Magyar Telefon Hírmondó és Rádió Rt. kiadványa, 1935.
- [3] Postamérnöki Szolgálat 50 éve (1887–1937) Magyar Királyi Posta, 1938.
- [4] A Rádió Kincsesára. Rádió Újság Lapvállalat KT, 1935.
- [5] Baczynski István: A magyar posta rádióhírszóró szolgálatának első 10 éve, Magyar Posta, 1935. december.
- [6] Dósa György: Egy évforduló emlékére. AH Rt, Modulátor, 1996/5.
- [7] Dósa György: Megemlékezés egy évfordulóról, I. rész HTE Hírlevél, 2000. szeptember.
- [8] Dósa György: Paskay Bernáth munkássága és élete. Napkiadó, 2000. október.
- [9] Dósa György: Száz éve kezdődtek meg Magyarországon az első rádiótávíró kísérletek, Híradástechnika, 2004/7.
- [10] Koós Árpád: Tomits Iván munkássága és élete. Napkiadó, 2001. december.
- [11] Holldonner László: Hollós József munkássága és élete. Napkiadó, 2001. december.

# A 2005-ös KDD kupa feladatának megoldása a Fürkész algoritmussal

KARDKOVÁCS ZSOLT, TIKK DOMONKOS, BÁNSÁGHI ZOLTÁN

Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék  
{kardkovacs,tikk}@tmit.bme.hu, empzooli@gmail.com

Lektorált

**Kulcsszavak:** KDD kupa, internetes keresőkifejezések osztályozása, taxonómiák leképezése egymásra, szövegosztályozás

A 2005-ös ACM KDD kupa versenyfeladata internetes keresőkifejezések kategorizálása volt. Jelen tanulmányban ismertetjük a problémára adott megoldásunkat, amellyel a pontossági és kreativitási versenyben is második helyezést értünk el. A megközelítésünk túlmutat a konkrét feladat megoldásán: általános eszközt nyújt olyan rosszul specifikált osztályozási feladatokra, ahol nem áll közvetlenül rendelkezésre elegendő tanulóadat. Az algoritmus, amely az Internetet használja a szükséges tanulóadatok forrásaként, három részből áll: 1. probléma-specifikus adatszűrő; 2. webrobot konfigurálása az adatok szüretelésére; 3. hatékony osztályozó algoritmus alkalmazása. A módszerünkben kiemelt fontosságú a probléma megoldása során kifejlesztett általános algoritmusunk, amely képes különböző kategóriarendszereket egymásra leképezni.

## 1. A feladat

2005-ben kilencedik alkalommal írta ki az ACM szervezet KDD<sup>1</sup> szakcsoportja a KDD kupát<sup>2</sup>. A verseny az adat- és szövegbányászattal, valamint a gépi tanulásal foglalkozó szakemberek (mind akadémiai, mind ipari területről) legrangosabb megmérettetése, amelyen évről évre egyre több kutatócsoport vesz részt a világ minden tájáról. Versenyfeladatként általában olyan problémát tűznek ki, amely a szakterület legaktuálisabb megoldatlan kérdéseit érinti. Az évek során a KDD kupákon számos nagy kihívást jelentő gyakorlati probléma lett kitűzve, melyek megoldásai hozzájárultak a tudományterület növekvő sikereihez.

A 2005-ös feladat az internetes keresések értelmének (kontextusának) meghatározásához kapcsolódott, ami szintén egy nehéz gyakorlati probléma. A legtöbb internetes keresés csak néhány szóból áll, azaz nagyon kevés információt tartalmaz a felhasználó keresési igényére vonatkozóan. Ha ez utóbbi rendelkezésre állna, akkor a keresőszolgáltatások hatékonysága nagymértékben javítható lenne.

A feladat megoldásának egyik számítástudományi megközelítése az, hogy megbecsüljük a keresőkifejezésnek egy adott taxonómia kategóriáihoz való hozzátartozását. Például a *jaguar* szóhoz egyaránt hozzárendelhetjük a *zoológia* és az *autó* kategóriákat is. Ezzel a módszerrel a keresési igény meghatározásának kérdése a legvalószínűbb kategóriák kiválasztására redukálódik – egy tetszőleges, de adott taxonómia alapján.

A KDD kupa 2005-ös feladata egy konkrét taxonómia feletti osztályozási probléma volt: 800.000 internetes keresőkifejezést kellett tartalmuk szerint 67 előre

megadott kategóriába besorolni<sup>3</sup>. Minden keresőkifejezéshez legfeljebb 5 kategóriát kellett rendelni rangsorolás nélkül.

A probléma megoldását külön nehezítette a rendkívül nagy méretű és rossz minőségű adattömeg, valamint hogy egyáltalán nem állt rendelkezésre tanulóadat, ami az osztályozás típusú gépi tanulási problémák esetén feltétlenül szükséges. További nehézséget jelentett, hogy a szervezők menet közben megváltoztatták a kategóriarendszert. A feladat nehézségére jellemző, hogy a több mint 140 regisztrált résztvevő közül csak 32 csapat adott be megoldást.

A cikkben az alábbi terminológiát fogjuk használni:

• **céltaxonómia:**

A szervezők által meghatározott 67 kategória halmaza, amelybe be kell sorolni a keresőkifejezéseket. Itt azért használjuk a „taxonómia” kifejezést, mert a megoldásunk során a kategóriákat kétszintű hierarchiába rendeztük, úgy, hogy egyes összetartozó kategóriákat egy közös felső szintű kategória alá soroltunk. Így a *Computer*, *Entertainment*, *Living* stb. lettek az új csúcskategóriák és az eredeti *Computer/Hardware*, *Computer/Software* stb. kategóriák lettek a levélszintű kategóriák.

• **célkategória:**

A céltaxonómia egy levélszintű kategóriája.

• **keresőkifejezés:**

Általában a 800.000-es keresőkifejezés egy elemét értjük alatta, ha másképp nem állítjuk; a szövegosztályozási paradigma alapján esetenként használjuk erre a *dokumentum* terminológiát is.

• **szótövezett keresőkifejezés:**

Ugyanaz mint előbb, de a keresőkifejezés szavai szótövezve vannak.

<sup>1</sup> Association of Computing Machinery, <http://www.acm.org>; Knowledge Discovery and Data Mining

<sup>2</sup> <http://www.acm.org/sigs/sigkdd/kdd2005/kddcup.html>

<sup>3</sup> A cikkben a rögzített kategóriarendszerbe történő besorolás feladatára egyaránt használjuk a besorolás, osztályozás, kategorizálás megnevezéseket.

• **forrástaxonómia:**

Internetes keresőszolgáltatás által nyújtott kategóriarendszer, amely segíti a felhasználókat dokumentumok keresésében és témák közti navigálásban.

**2. Bevezetés**

A dokumentumok rögzített kategóriarendszerbe való besorolását szövegosztályozásnak nevezik. A KDD kupa 2005-ös versenykiírása is nyilvánvalóan ebbe a feladatkörbe tartozik. A szövegosztályozás tipikus felügyelt gépi tanulással feladat; a tanuló algoritmus ismert mintaadatok segítségével „megtanulja” a kategóriarendszer sajátosságait, majd ezután ismeretlen, korábban nem látott dokumentumokat a megtanult jellemzők figyelembevételével osztályoz.

A kupa kiírása azonban tartalmazott néhány olyan lényeges eltérést a standard feladattól, amely miatt a feladatot nem lehetett egyszerűen, valamely ismert algoritmus alkalmazásával megoldani:

1. A dokumentumok (itt: keresőkifejezések) nagyon rövidek voltak, a 90%-uk 5 szónál rövidebb.
2. A korpusz nagyon zajos volt, legalább 30%-ában rossz karakterkódolású nem-angol szövegeket, illetve teljesen értelmetlen szövegeket tartalmazott.
3. A leglényegesebb eltérés az volt, hogy nem állt rendelkezésre tanulóadat: a mellékelt 111 mintapélda csak a céltaxonómia szemantikájának illusztrálását szolgálta.

Ebből következően a mintapéldákat csak validálási célra lehetett korlátozott mértékben felhasználni. Ennek ellenére úgy véltük, hogy a feladatot csak felügyelt tanulással érdemes megközelíteni, mivel

- A keresőkifejezések nagy számú tulajdonnevet tartalmaztak, amelyek hatékony figyelembevétele csak a szótárak vagy tanulóadatok segítségével lehetséges.
- Ha létezett volna a céltaxonómiához hasonló forrástaxonómia megfelelő dokumentumokkal, azt csak akkor lehetett volna felhasználni, ha valamilyen leképezés rendelkezésre állt volna a forrástaxonómia és a céltaxonómia kategóriái között. Mivel azonban a célkategóriák szemantikája nem volt ismert, ezzel bármely ilyen leképezés érvényessége megkérdőjelezhetővé vált.

Összegezve úgy véltük, hogy a megoldást a keresőkifejezések szemantikailag indokolt kibővítése és tanulóadatokként való felhasználása jelentheti, amely lépés után a felügyelt tanulás paradigmája már alkalmazható.

A módszert „Fürkész algoritmusnak” neveztük el, amely a következő lépésekből áll (1. ábra):

1. **Forrás megtalálása:** Határozzuk meg a céltaxonómia kategóriáinak szemantikáját a keresőkifejezések tartalmából kinyert kiindulási szótár létrehozásával, majd adjunk meg ezen szemantika alapján érvényes leképezést internetes kereső-

szolgáltatások forrástaxonómiája, valamint a céltaxonómia között.

2. **Dokumentumok szótővezése:** Hajtsunk végre szó-tővezést az összes keresőkifejezésen, így kapjuk a szótővezett keresőkifejezéseket.
3. **Keresés az Interneten:** A cél- és forrástaxonómiák közti kapcsolat alapján a szótővezett keresőkifejezéseket küldjük el a forrástaxonómiához kapcsolódó keresőszolgáltatások felé tanulóadatok gyűjtése céljából.
4. **Eredmények feldolgozása:** Dolgozzuk fel a 3. pontban kapott eredményeket. Az egyes keresőkifejezésekhez kapott eredményoldalakból és a taxonómiák közti leképezésből meghatározzuk a keresőkifejezés kategóriáját. Az eredményoldal tartalmát tanulóadatként az adott kategóriához rendeljük.
5. **Osztályozó betanítása:** A 4. pontban kapott tanulóadatokkal tanítsunk be egy tetszőleges szövegosztályozót. Munkánk során a HITEC hierarchikus osztályozót használtuk [1,6,7], részleteket lásd a 3.4. szakaszban.
6. Futtassuk az osztályozót a keresőkifejezésekre. Rangsoroljuk az Internet keresés és az osztályozás eredményeit, és határozzuk meg a legjobb 5 kategóriát minden keresőkifejezésre.

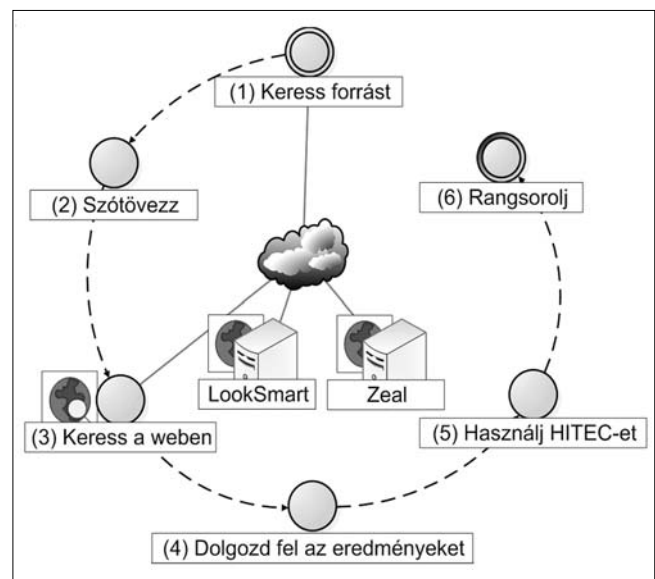
A következő szakaszokban a fenti lépéseket részletesen bemutatjuk.

**3. A Fürkész algoritmus**

**3.1. Az Internet mint tudásbázis**

A Fürkész algoritmus lényege, hogy az Internet-et használja tudásbázisként azért, hogy különböző keresőszolgáltatásokat használ fel a keresőkifejezések szemantikájának meghatározására. Tehát az algoritmust egy metakeresőnek is tekinthetjük.

1. ábra  
A Fürkész algoritmus lépései



A KDD kupa feladatának megoldása során a LookSmart (<http://search.looksmart.com>) és a Zeal (<http://www.zeal.com>)<sup>4</sup> keresőszolgáltatásokat használtuk. A választásunk azért esett erre a kettőre, mivel némi párhuzam felfedezhető a céltaxonómia és a keresőmotorok forrástaxonómiai közt, illetve a két forrástaxonómia is hasonló, ezért megoldható feladat volt köztük leképezést meghatározni. A keresőszolgáltatások által adott eredményeket lokálisan dolgoztuk fel.

Mindazonáltal a LookSmart és Zeal teljes forrástaxonómiáinak újraépítése nem tűnt célszerű feladatnak, mivel ezek nagyon nagyméretű és szövevényes rendszerek. Ezek a taxonómiák ugyanis lehetőleg teljesen le kívánják fedni a felhasználók érdeklődési körét, hiszen a szolgáltatások fő célja, hogy hatékonyan segítsék a meghatározott céllal kereső felhasználót internetes oldalak és regionális információk megtalálásában. Ezért a Fürkész algoritmussal csak azt a részgráfját térképeztük fel a forrástaxonómiáknak, amelyhez a keresőkifejezések által meghatározott kategóriák tartoztak. Ezt a szolgáltatások publikus keresési felületének alkalmazásával tettük meg, leszűretelve az eredményoldalak egy részét a későbbiekben történő feldolgozásra.

Szintén jól felhasználható forrás volt a fenti szolgáltatások esetén, hogy minden kategóriához egy rövid meghatározás is tartozott. Ez alapján hoztuk létre a céltaxonómia szemantikáját leíró kiinduló szótárát.

### 3.2. Kategóriák szemantikája

Mivel meg voltunk győződve arról, hogy a KDD kupa feladatát felügyelt tanulással célszerű megoldani, ezért a tanulóadatokat begyűjtése elsődleges céllá vált. Ezt a következőképpen valósítottuk meg. Először minden keresőkifejezést a Porter-eljárással [2] szótőveztünk, majd ezeket elküldtük a keresőszolgáltatásoknak.

Az LZ eredményoldalak két fő részből álltak<sup>5</sup>: ha a keresésnek van eredménye, akkor az egyik rész a találat rövid kivonata (*kontextus*), illetve ha a találat a keresőmotor szerkesztői által be lett sorolva a forrástaxonómia valamelyik kategóriájába (*forráskategória*), akkor a találat másik része a kategóriának a neve.

Abban az esetben, ha vannak még jellemző kategóriák a keresőkifejezésre, akkor ezekből a legfontosabbakat egy külön szekcióban jelzik. Emellett a fontosabb kategóriákhoz a kategória tartalmát jellemző rövid leírás is adott. A forrástaxonómia szerkezetét, illetve az egyes forráskategóriák gyökérkategóriából való elérhetőségét gyerek-szülő relációk alapján térképeztük fel.

Ezzel a módszerrel megkaptuk a két forrástaxonómia feladat megoldása szempontjából releváns részstruktúráját, valamint néhány szavas szemantikai leírást a forráskategóriákra. A leírások összességét kiindulási forrásszótárnak ( $BC$ ), a kapcsolódó kategóriákat pedig  $BC$ -kategóriáknak nevezzük.

### 3.3. A taxonómiák közti leképezés meghatározása

A taxonómiák közti leképezés meghatározásának alapja a céltaxonómiahoz tartozó maximális releváns szóhalmaz meghatározása.

Feltesszük, hogy a céltaxonómia neve a lehető legjobban leírja az adott kategóriát. Ha ez fennáll, akkor a céltaxonómia nevében szereplő szavak szinonimái szintén jól írják le a kategóriát. Ezt a lépést a szavak WordNet<sup>6</sup> szinonimáival való kibővítésével végeztük, és az eredményt a céltaxonómia szemantikus lezárójának neveztük.

$$\text{Legyen } W(0) = \bigcup_i w_i(0)$$

a céltaxonómiát leíró kiinduló célszótár, ahol  $w_i(0)$  jelöli az  $i$ -edik kategória szemantikus lezáróját.

A  $W(0)$  halmazból kiindulva kerestünk releváns  $BC$ -kategóriákat, vagyis ahol a  $BC$ -kategória szemantikai leírása és a kiinduló célszótár közös elemeket tartalmaz. Formálisan, legyen  $C_i^0 \subseteq BC$  azon része a forrásszótárnak, amelyet a  $w_i(0)$  céltaxonómia szemantikus lezárója meghatároz, vagyis amely kategórialeírások metszete  $w_i(0)$ -lal nem üres. A Fürkész algoritmus a jól ismert TF-IDF mértéket számolja ki (pl. [3]) a  $C_i^0$  leírások szavaira, amelyeknek relatív gyakorisága legalább  $\omega$  legalább egy  $C_i^0$ -beli leírásban, és legfeljebb  $\alpha$  számú  $C_i^0$ -beli leírásban fordul elő. (Ezzel a túl ritka és túl gyakori szavak kiszűrését tudjuk parametrikusan megvalósítani). Legyen  $A^0$  azon szavak halmaza, amelyre ez a tulajdonság fennáll. Ekkor a következő rekurzív formulát alkalmaztuk:

$$w_i(n+1) = w_i(n) \cup \{a \mid a \in A^n \cap C_i^n\} \quad (n = 0, 1, \dots) \quad (1)$$

Ezt a lépést nemcsak az eredeti, a céltaxonómiában levél szinten lévő 67 kategóriára végeztük, hanem a felső szintű kategóriákra is. Ekkor úgy jártunk el, hogy a gyerekkategóriák halmazainak unióját képeztük, és arra alkalmaztuk a fenti eljárást. Ez azért fontos lépés, mert az IDF tényező egyes gyakori szavakat – amelyek például a kategóriák egy csoportjára jellemzőek – kiszűrhet, de ekkor ezeket még felső szintű kategóriára jellemző szóként figyelembe vehetjük. A felső szintű céltaxonómia kategóriák meghatározásánál is szerepük van: ha egy dokumentum csak a *Computer* kategóriához tartozik, de egyik alkategóriájához sem, akkor mint egyéb lesz osztályozva (*Computer/Other*).

A fenti (1) képlettel leírt rekurzív algoritmus a leírásban szereplő szavak végeessége miatt nyilván terminál. Azok a forráskategóriák

$$C^- = BC \triangleleft \bigcup_i C_i^N$$

(itt  $\triangleleft$  a halmazok közti különbségképzés jele),

melyhez nem rendeltünk céltaxonómiát, úgy kerülnek felhasználásra, hogy a forrástaxonómia relációi mentén a legközelebbi olyan szülőkategóriához rendeljük őket, amelynek már a céltaxonómiában van párja.

<sup>4</sup> Együtt a kettőre LZ-ként hivatkozunk a továbbiakban.

<sup>5</sup> A letöltéseket 2005 júliusában végeztük, azóta a Looksmart kereső honlapja és eredményoldalainak szerkezete megváltozott.

<sup>6</sup> <http://wordnet.princeton.edu/>

Az ilyen  $C^+ = BC \triangleleft C^-$  halmazbeli forráskategóriákat *megjelölt kategóriának* nevezzük. Vegyük észre, hogy az algoritmussal megkapjuk mind a célkategóriák szemantikus leírását,  $w_i^N$ -t, mind a cél- és forrástaxonómia közti leképezést; ezt minden  $i$  célkategóriára a  $C_i^N$  halmazban lévő forráskategóriák adják meg. Az algoritmus folyamatábrája és pszeudokódja rendre a 2. és 3. ábrákon látható.

2.-3. ábra

A taxonómiák közti leképezést meghatározó algoritmus pszeudokódja és a leképezés meghatározásának lépései

**bemenet:** egy részfa struktúra és annak gyökere

**kimenet:** egy leképezés a forrás és célkategóriák között

```
function katLeképezés( Csomópont Gyökér, Gráf Fa ) begin
  let gyakorisági_küszöb := 0.05;
  let ismétlődési_küszöb := 3;
  if ( Gyökér is Levél )
    return nil;
  foreach S in Gyökér Fa-beli gyermekei
  begin
    let ÚjSzók[S] := { S morfológiai alaptöve és szinonimái };
    let KatJelöltek[S] := Üres;
    let Szók[S] := Üres;
  end
  while ( ÚjSzók != Szók )
  begin
    let Szók := ÚjSzók;
    foreach S in Gyökér Fa-beli gyermekei
    begin
      let ÚjKatJelölt[S] := KatJelöltek[S] + { forráskategóriák,
      amelyek tartalmazzák Szók[S] egy elemét };
      let ÚjSzók[S] := { ÚjKatJelölt[S] leírásában szereplő azon
      szavaknak morfológiai tövei, amelyre a tf > gyakorisági_küszöb };
    end
    let ÚjSzók := azon szavak halmazainak tömbje, amelyek legfeljebb
    ismétlődési_küszöb kategóriában fordulnak elő: idf < ismétlődési_küszöb;
    let KatJelöltek := azon kategóriahalmazok tömbje, amelyre igaz,
    hogy egy kulcs csak egy kategória alá lett besorolva;
  end
  return KatJelöltek;
end
```

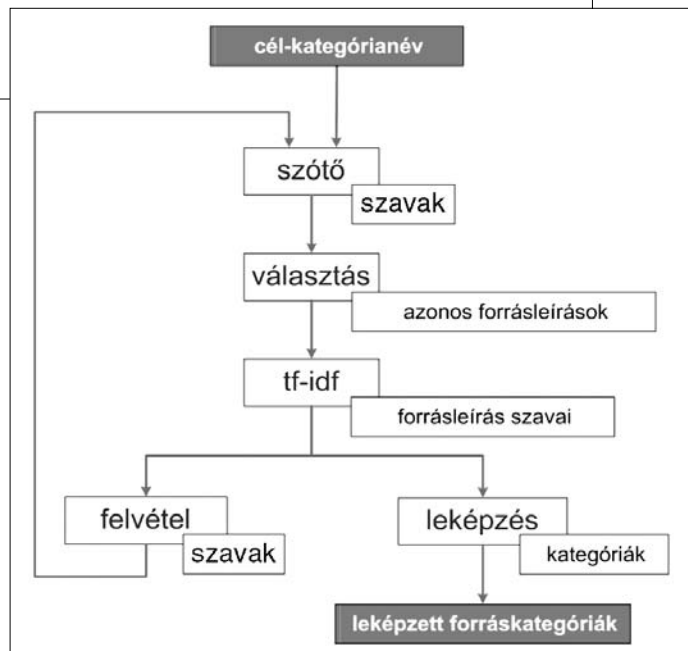
Vegyük észre, hogy a 3. ábrán ismertetett algoritmus tetszőleges forrás- és céltaxonómia összekapcsolására alkalmas, ha a forrástaxonómia kategóriáinak van szemantikus leírása. Ez a feltétel könnyen teljesíthető, ha a forrástaxonómiához adottak tanulmányok, ekkor ugyanis a szemantikus leírást a kategóriákhoz rendelt tanulmányokban szereplő leggyakoribb szavak összességeként kaphatjuk. A kategóriáknak szóprofilja, vagy prototípus vektora szintén tekinthető szemantikus leírásnak.

### 3.4. Tanítás és osztályozás

Miután a leírt módon a kategóriákhoz tanulmányokat gyűjtöttünk, a HITEC [1,6,7] osztályozót alkalmaztuk a keresőkifejezések osztályozására. Erre azért volt szükség, mert az összes keresőkifejezésnek csak mintegy feléhez sikerült a fenti módszerrel

kategóriát rendelnünk. A HITEC neurális hálózat alapú osztályozó, amely a tanulmányok alapján minden kategóriához egy prototípus vektort készít. Amikor ismeretlen dokumentumokat osztályoz, akkor a prototípus vektorokhoz való hasonlóság alapján határozza meg a taxonómiában lefelé haladva a dokumentum releváns kategóriáit. A HITEC-nek számos dokumentumfeldolgozást, tanulást és következtetést szabályozó paramétere van. Ez utóbbiak főleg a taxonómiában való keresés szélességét és mélységét befolyásolják.

További részletek a hivatkozásokban találhatóak.



### 3.4.1. A tanulóadatok

Mivel a keresőkifejezések eleve nagyon kevés szót tartalmaztak, ezért a csak ezek tartalmából épített szótár mérete is kicsiny, és nem bír elegendő leíróképeséggel az ismeretlen dokumentumok osztályozására. Ezt javítandó, 4 alternatív tanulóhalmazt alakítottunk ki, amelyeket az eredeti dokumentumoknak az LZ eredményoldalak tartalmával való bővítésével kaptunk meg. Az eredményoldalak szerkezetének leírását az Olvasó a 3.2. szakaszban találja.

Az első három alternatíva forráskategóriákhoz rendel tanulóadatokat az LZ eredményoldalak alapján. A tanulóadat maga a keresőkifejezés, vagy annak a lent leírt módon való valamilyen bővítése. Olyan forráskategóriákhoz rendeltünk ily módon tanulóadatokat, amelyek legalább egyszer előfordultak a keresőkifejezések eredményoldalán. A forrás- és célkategóriák közti kapcsolatot a korábban ismertetett leképezés (lásd a 3.3. szakaszt) segítségével hoztuk létre. A negyedik alternatív tanulóhalmazt közvetlenül a forráskategóriák alapján határozzuk meg, a keresőkifejezések használata nélkül.

- **K** – KERESŐKIFEJEZÉS:

Ez a legegyszerűbb eset, amikor tanulóadatként magát a szótővezett keresőkifejezést alkalmaztuk. Ezt minden olyan forráskategóriához (és a leképezésen keresztül célkategóriához) hozzárendeltük, amely az adott kereséskor eredményként legalább egyszer előfordult.

- **SK** – SÚLYOZOTT KERESŐKIFEJEZÉS:

Az előzőhöz hasonló, de ekkor a szótővezett keresőkifejezést annyiszor rendeltük a forráskategóriához, ahányiszor az adott kereséskor eredményként előfordult. Vagyis ha a  $q$  keresőkifejezés eredményoldalán a  $c$  kategória kétszer fordult elő, akkor a  $q$  szövegét kétszer egymás után összefűzve rendeltük  $c$ -hez.

- **T** – TEXT:

A szótővezett keresőkifejezés szövegét az alábbi módon bővítettük. Az eredményoldalak letöltése után kinyertük a szöveges ASCII információt a HTML oldalakból, és evvel kibővítettük a keresőkifejezést. A kategóriákhoz analóg módon rendeltük az így kapott tanulóadatot.

- **C** – KATEGÓRIALEÍRÁS:

A keresőszolgáltatások által adott rövid szemantikai leírás szövegét rendeltük a forráskategóriákhoz. Vegyük észre, hogy ez független a keresőkifejezésektől. A leírás címből és általában egy mondatból áll, amelyeket külön mezőként kezelve különböző súlylyal láttunk el (lásd 1. táblázat). Ezt a módszert csak megjelölt kategóriákra alkalmaztuk.

### 3.4.2. A jellemzők kiválasztása

A tanulás egyik legfontosabb része a megfelelő jellemzők – szövegosztályozás esetén az optimális szó-

Futás	$d_1$	$d_2$	$w_t$	$w_d$	$ D $	Tanulóhalmaz
R1	2	0,7	5	3	888.565	K+T+C
R2	5	1,0	10	3	149.792	K+C
R3	5	1,0	10	3	809.165	K+T+C
R4	0	1,0	10	3	1.088.171	SK+T+C
R5	1	1,0	10	3	763.872	T+C

1. táblázat

A szótár mérete és a tanulóhalmazok közti összefüggés

tárméret – meghatározása. A cél a lényegtelen, ritka szavak elhagyása és a megkülönböztető szavak megtartása közti optimális egyensúly elérése.

A HITEC két egyszerű paraméterrel rendelkezik a szótár méretének ( $|D|$ ) szabályozására:

- $d_1 \in [0, \infty)$  a szavak korpuszban való minimális előfordulására vonatkozó alsó küszöbérték;
- $d_2 \in (0, 1]$  pedig a szavak korpuszon való teljes eloszlására vonatkozó felső küszöbérték.

Ezek a paraméterek a TF-IDF súlyozási sémával kapcsolatosak:  $d_1$  és  $d_2$  rendre a TF-re vonatkozó alsó, illetve az IDF-re vonatkozó felső korlátnak tekinthető.

Az adott feladat esetén, amikor K, T vagy C tanulóhalmazokat (és ezek kombinációit) alkalmaztuk, nagyon alacsony  $d_1$  (2-5) és nagyon magas  $d_2$  (~0,5) értékekkel számoltunk, mivel az egész korpusz csak kevés szót tartalmazott, és jellemzően azok eloszlása is ritka volt. A  $d_1$  paraméter alacsony értékét az is indokolta, hogy ebben sok értékes, megkülönböztető jelleggel bíró szó csak 1-2-szer fordult elő. A C tanulóhalmaz használatakor két súlytényezőt alkalmaztunk:  $w_t$ -t a címre, és  $w_d$ -t a leírásra.

A legnagyobb tanulóhalmaz esetén (K+T+C), a teljes szótár szótővezés után 1080 ezer szót tartalmazott. Jellemző-kiválasztással ennek méretét 809 ezerre csökkentettük. A K+C tanulóhalmaz esetén csak 149 ezer szó volt a szótárban. Az 1. táblázatban feltüntetünk néhány jellemző-kiválasztási futás eredményét.

### 3.4.3. A tanulás paramétereit

Az iteratív tanulás során az iterációk számát 5-nek rögzítettük, ami korábbi nagy korpuszokon való osztályozási feladatoknál (pl. Reuters Corpus Volume 1) kapott kvázi-optimális érték. A tanulás során minden tanulóadatot felhasználtunk.

A HITEC-ben két fontos paraméter szolgál az osztályozási következtetés eredményhalmaza méretének szabályozására. A maximális varianciával ( $v_{\max} \in (0, 1]$ ) megadható, hogy a legnagyobb konfidenciaszintű csomópontokhoz képest milyen arányú eltérést engedünk a taxonómia egy adott szintjén, amikor a további keresésnél figyelembe vett csomópontokat határozzuk meg.

Ha ezt az értéket alacsonyra állítjuk (kb. 0,5), akkor minden szinten több kategóriát választunk ki, és így a következtetés több szálon fut. A küszöbérték,  $\theta$  segítségével a kiválasztáshoz szükséges minimális konfidenciaérték adható meg. Ha ez az érték alacsony (0,05~

0,15), akkor ismét több kategóriát kapunk eredményül. Ezzel a két paraméterrel tehát a felidézés és a pontosság duális mértékek közti egyensúlyt lehet beállítani; alacsony paraméterértékek esetén az előbbi nő, az utóbbi viszont csökken.

A KDD kupa feladatánál ezeket az értékeket alacsonyra állítottuk, hogy a lehető legtöbb keresőkifejezésre megkapjuk a szükséges öt eredménykategóriát:  $v_{max} = 0,5$  és  $\theta = 0,1$ .

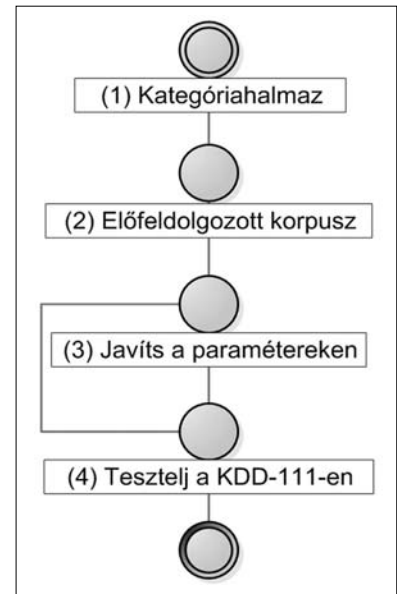
A HITEC lehetővé teszi, hogy a szótárkészítéshez, és a tanuláshoz különböző tanulóhalmazokat használjunk. Ezt kihasználva a legtöbb futás során a tanuláshoz figyelmen kívül hagytuk a K tanulóhalmazt, mivel azt tapasztaltuk, hogy jelenléte rontja a tanítást. Ez annak köszönhető, hogy ezek a dokumentumok túl rövidek és zajosak. Másrészt viszont K felhasználása a szótár létrehozásánál hasznos, mivel több fontos szó előfordulási értékét növeli –, például a K+T+C tanulóhalmaz alkalmazása esetén.

Egyes beállítások tanulási kapacitását – azaz, hogy a HITEC mennyire volt képes a tanulóadatokat megtanulni – a 2. táblázatban ismertetjük. A 4. ábrán az iterációk során elért tanulási hatékonyságot ábrázoljuk a HITEC belső kiértékelő függvényének segítségével. A bemutatott hatékonysági mértékeket a tanulóadatokon történt teszteléssel kaptuk meg.

#### 3.4.4. Kiértékelés

A különböző jellemzőkiválasztási verziókat és tanulási beállításokat a megadott 111 mintaadaton teszteltük (lásd még az 5. ábrát). A kiértékelés során az LZ eredményoldalakat közvetlen feldolgozásával kapott kategóriák hatékonyságát is vizsgáltuk. Itt hasonló módon

– a taxonómiák közti leképezés alapján – jártunk el, mint ahogy a keresőkifejezésekhez tartozó célkategóriákat meghatároztuk. Az összehasonlítás során megállapítottuk, hogy a 111 mintaadaton a HITEC által adott következtetés lényegesen rosszabb eredményt adott, mint amit az LZ oldalak közvetlen feldolgozásával kaptunk, függetlenül a HITEC paramétereinek beállításától. Ez egyáltalán nem meglepő, hiszen a HITEC eredményeiben kétféle hiba kumulálódik: egyrészt a tanulóhalmaz szintén nem elhanyagolható mértékű hibája, másrészt a tanulás és az osztályozás hibája. Ráadásul a 111 mintapéldában számos olyan szó van, amely az egész korpuszban nem fordul elő egyszer sem (pl. *aldactone*), amire tehát az osztályozó nem tud értelmes következtetést adni.

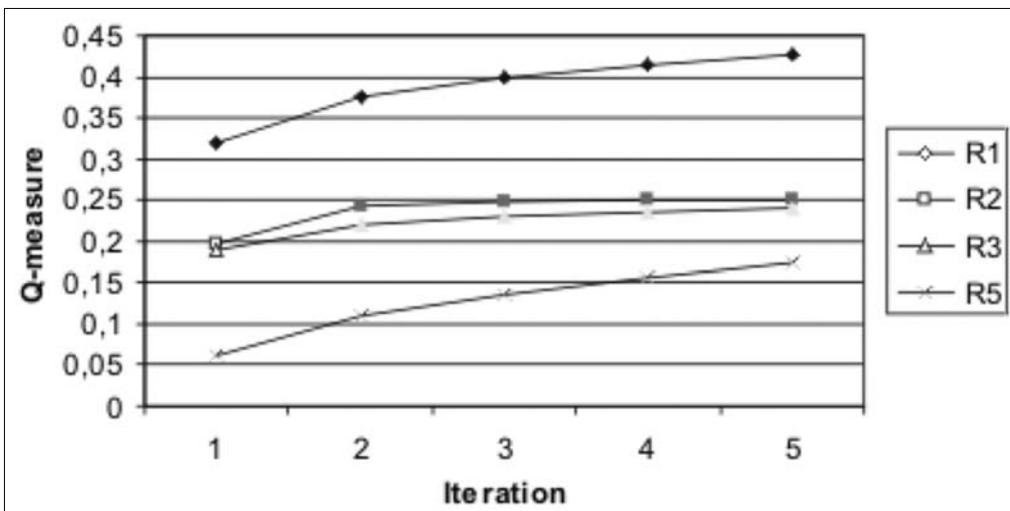


5. ábra  
Kiértékelés a HITEC-kel

Futás	$v_{max}$	$\theta$	$F_1$	Pontosság	Felidézés	Adatok
R1	0,5	0,1	0,655	0,586	0,741	T+C
R2	0,5	0,1	0,532	0,459	0,634	K+C
R3	0,5	0,1	0,517	0,446	0,616	T+C
R5	0,7	0,2	0,326	0,280	0,391	K+T+C

2. táblázat

Validációs eredmények egyes futásokra (jelölések az 1. táblázat alapján)



4. ábra

A tanulás hatékonysága a HITEC belső kiértékelő függvénye alapján

A tanulás hatékonyságát két tényezővel ellenőriztük. Egyrészt a 111 mintaadaton, másrészt a validációs értékek  $F_1$  mértéke alapján. Ez alapján azt találtuk, hogy a legígéretesebb eredményt a (T+C) tanulóhalmaz adta a  $d_1 = 2$ ,  $v_{\max} = 0,5$  és  $\theta = 0,1$  értékek mellett. Idő hiányában nem volt lehetőségünk a HITEC optimális beállításának meghatározására. (Az R1 futás már a beadási határidő után ért véget, ezért az R2 futás eredményét adtuk be.)

#### 4. A beadott eredmények

A beadott eredmények két forrásból származtak:

1. Azokra a keresőkifejezésekre, amelyekre az LZ keresések adtak forráskategóriát, eredményként a leképezés által megadott célkategóriá(ka)t adtuk be. Ezáltal mintegy 400 ezer keresőkifejezésre kaptunk eredményt.

2. A többire a HITEC betanított modelljének legnagyobb 5 konfidenciaértékű következtetését adtuk be. Ezzel további mintegy 320 ezer keresőkifejezésre kaptunk eredményt.

3. A maradék körülbelül 80 ezer keresőkifejezésre nem adtuk be eredményt. Ezek túlnyomórészt olyan keresőkifejezések voltak, amelyek nem tartalmaztak értelmes angol szavakat, vagy tulajdonneveket.

A beadott megoldásokat a szervezők mindössze 800, három szakértő által osztályozott adaton tesztelték. Az általunk beadott eredmények 0,340883 pontosságot és 0,34009  $F_1$ -mértéket értek el. Az előbbi értékkel, illetve a Fürkész algoritmus ötletességével második helyezést értünk el a pontossági és a kreativitási versenyben.

Utólag megvizsgáltuk, hogy az algoritmusunk egyes komponensei milyen mértékben járultak hozzá az elért eredményhez. A 800 adatból 665-re volt válasz a tanulóadatokban (1. csoport). A maradék 135 keresőkifejezésre (2. csoport), amennyiben volt válasz, a HITEC által szolgáltatott eredményt adtuk be.

Akad 800 között néhány olyan keresőkifejezés is, amire a HITEC sem adott megoldást, tehát a 3. csoportba tartoztak (lásd 3. táblázat „Nincs címke” oszlop). Az ezek nélkül számolt értékek a 4. táblázatban találhatóak, ahol nyilván csak a felidézés és az  $F_1$  értékek változnak.

A HITEC válaszaira megnéztük, hogy a legjobb három futás (R1, R2, R3) milyen eredményt adott. Vizsgálatunk érdekes eredményt adott: a legjobb eredményt az R3 futás adta, míg a legrosszabbat az R1, ami éppen a tanulóadatokon mutatott viselkedés fordítottja (részleteket lásd a 3. és 4. táblázatban).

Ezt két okkal magyarázhatjuk. Egyrészt a tanulóadatokon való kedvező viselkedés azokon a keresőkifejezéseken való hatékony tanulást jelenti, amelyekre az 1. csoportból adtuk be eredményt, tehát az ezekre való tanulási képesség a kiértékelésünkben nem játszott szerepet. Másrészt a jobb validációs eredményű tanuláshoz túl tanulás lép fel, ezért az így betanított osztályozó általánosító képessége kisebb. A hatékonyság alacsony számértéke meglepő: ezek pont a legnehezebb, tanulóadatokkal nem rendelkező keresőkifejezésekre adott válaszok, tehát itt ennél lényegesen jobb eredmény nem is várható el.

#### 5. A versenyen díjazott további módszerekről

##### 5.1. Osztályozók kombinációja

Az első helyezést mindhárom kategóriában a hongkongi HKUST egyetem csapata érte el [5]. Megoldásukban egy többkomponensű osztályozót készítettek, melynek a vázlata a 6. ábrán látható.

A keresőkifejezések és a céltaxonómia elemei közti hozzárendelés megvalósításának első lépése a keresőkifejezések internetes keresőmotorok (SE) segítségével való kibővítése. Három keresőmotort használtak fel erre a célra: a Looksmartot, a Google-t, és egy általuk konfigurált ODP taxonómián működő Lemur-alapú

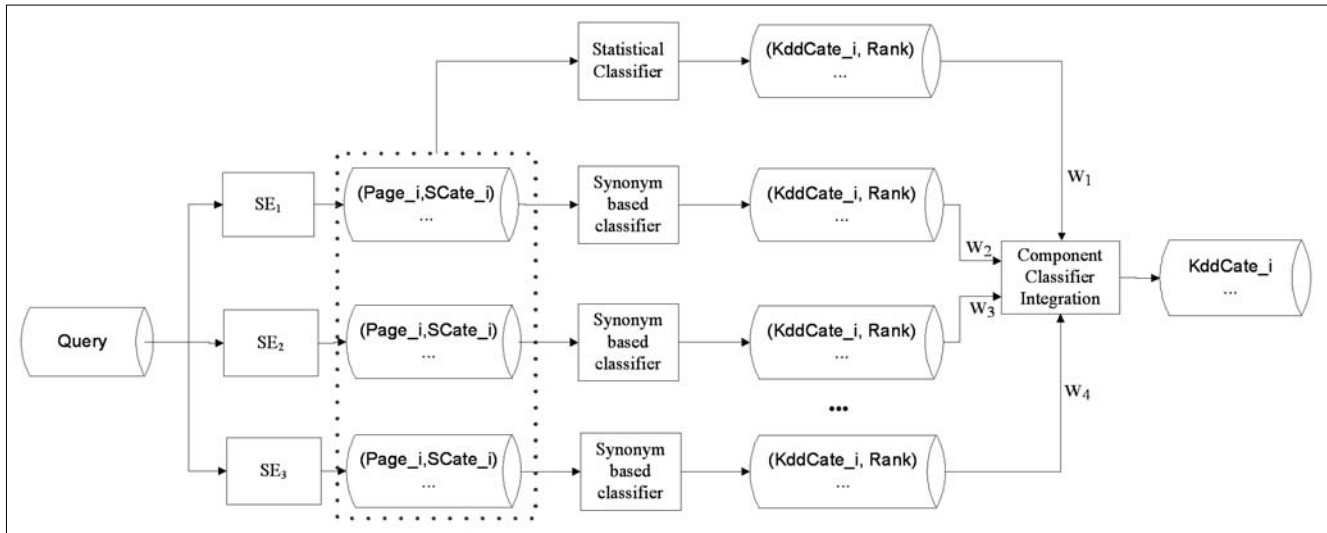
3. táblázat HITEC futások hatékonysága a 135 keresőkifejezésen

Futás	$F_1$ 1. címkéző	$F_1$ 2. címkéző	$F_1$ 3. címkéző	Átlag pontosság	Átlag felidézés	Átlag $F_1$	Nincs címke
R1	12,47%	10,03%	10,82%	13,55%	9,58%	11,11%	39
R2	16,61%	14,48%	14,22%	16,95%	13,90%	15,10%	31
R3	17,43%	15,43%	16,39%	19,36%	14,53%	16,42%	32

4. táblázat HITEC futások hatékonysága azon keresőkifejezéseken, amelyekre adott választ

Futás	Válaszok száma	$F_1$ 1. címkéző	$F_1$ 2. címkéző	$F_1$ 3. címkéző	Átlag pontosság	Átlag felidézés	Átlag $F_1$
R1	96	15,19%	11,68%	13,06%	13,55%	13,35%	13,31%
R2	104	19,33%	16,13%	16,49%	16,95%	18,06%	17,32%
R3	103	20,50%	17,41%	19,18%	19,36%	19,10%	19,03%





6. ábra Az HKUST osztályozójának vázlatja

keresőt<sup>7</sup>, összesen mintegy 40 millió weboldalt és 50 GB-nyi adatot letöltve. Az ismertett megoldásunkhoz analóg módon, az eredeti keresőkifejezéseket a keresőkhöz elküldve, a kapott eredmény oldalainak feldolgozásával készítették a következő komponens, az osztályozók számára tanulóadatokat. Nyilván itt is szükség van a három forrástaxonómia és a céltaxonómia kategóriái közötti leképezés meghatározására ahhoz, hogy az összegyűjtött tanulóadatokat alkalmazni lehessen a feladatra.

Ezt a leképezést két lépésben hozták létre. Elsőnek használtak egy, az általunk bemutatott kulcsszó illeszkedési technikához hasonló (bár nem iteratív) algoritmust, ahol szintén a Wordnetet használták fel a célkategóriák nevének bővítésére. Ez a módszer nagy pontosságú leképezést biztosít, azonban a fedés (másképp felidézés) szempontjából kedvezőtlen, hiszen kevés olyan forráskategória van, amelyet közvetlenül a neve alapján egy célkategóriához lehet rendelni. Ezért a második lépésben SVM (szupport vektor gép) alapú tanulómódszert alkalmaztak, ahol a letöltött oldalakhoz az első lépésben rendelt célkategóriák jelentették a tanulóadatokat. Ez a korpusz összesen 15 millió weboldalt tartalmazott. Az így felépített vektortér modell már elegendően teljes lett, és így a fedés is kielégítőnek bizonyult.

A leglényegesebb különbség az általunk javasolt módszerhez képest a következő lépésben rejlik. Nyilván az előző két lépésben létrehozott különböző leképezések különböző osztályozó függvényeket és eltérő eredményeket adnak, amelyek bizonyos kategóriákon jobb teljesítménnyel működnek. Célszerű tehát ezeknek az *osztályozóknak* valamilyen *kombinációját* tekinteni oly módon, hogy az adott keresőkifejezésre a lehető legjobb eredményt kapjuk. Az osztályozók jóságának kiértékelését a 111 validációs adat segítségével valósították meg, amely alapján súlyfaktorokat rendeltek az

osztályozókhoz. Amennyiben egy keresőkifejezést helytelenül osztályozott a rendszer, akkor a megfelelő súlyok állításával elérték, hogy a hibát kiküszöböljék. Ezzel a boosting jellegű [4] iteratív technikával az  $F_1$  értéke jelentősen javítható. Mivel azonban a 111 adat nagyon kevés, ezért ez a módszer rendkívül hajlamos a túlfitelésre. Ezt kiegyensúlyozandó olyan osztályozókat is bevettek a kombinációba, amelyek uniform súlyokkal rendelkeztek (statisztikai osztályozó). A végső eredményt a két típusú osztályozók eredményeinek kombinációjaként állították elő.

## 5.2. Osztályozás logikai regresszióval

Az  $F_1$  alapú osztályozási hatékonyság versenyének második helyezését a floridai A.I. Insight, és MEDai cégek, valamint a berlini Humboldt Egyetem közös csapata érte el.

Módszerük [8] első lépéseként a Google keresőjét alkalmazták az ODP taxonómian<sup>8</sup> némileg módosított keresőkifejezéseken. Itt kihasználták a keresőnek helyesírási hibák javítására vonatkozó szolgáltatását („*did you mean:...*”). Az ODP taxonómia és a céltaxonómia kategóriái közti leképezést manuálisan valósították meg, általában az ODP taxonómia felső két szintjének csomópontjait rendelték célkategóriákhoz, de ahol a finomítás megkívánta, akár a negyedik szintre is lementek a forrástaxonómiában. Az ennél is lejjebb lévő kategóriák besorolásához egy ajánló rendszert fejlesztettek ki, aminek eredményeképpen a manuális leképezést kiegészítették. A végső leképezés  $n:m$  típusú volt, ahol egy ODP kategóriához legfeljebb 3 célkategóriát rendeltek.

Az osztályozást az A.I. Insight logikai regresszióval alapuló szoftverével végezték<sup>9</sup>. A modellben három paramétertípus van: a kategóriákra vonatkozó súlytényező, a kategória rangja az adott célkategóriák sorrendjében, illetve a két legjobb kategória közti súlytényező

<sup>7</sup> Lemur: <http://www.lemurproject.org>, ODP (Open Directory Project): <http://dmoz.com>

<sup>8</sup> <http://www.google.com/dirhp>

<sup>9</sup> Mitch (Multiple Intelligent Tasking Computer Heuristics)

különbsége. A szoftver ezeket a paramétereket egyetlen valószínűségi változóvá kombinálja, és ez alapján végzi a következtetést. A módszer előnye, hogy lehetőséget biztosít a pontosság és a felidézés közti optimalizásra.

## 6. Összefoglalás

Jelen munkánkban bemutattuk a KDD kupa 2005-ös kiírására kifejlesztett algoritmusunkat. A megoldásnak két lényeges tényezője volt: egyrészt az Internet alapján megtalált és feldolgozott forrástaxonómiák, valamint a céltaxonómia közti leképezés meghatározása, és ily módon a tanulóadatok biztosítása; másrészt pedig a HITEC osztályozó hatékonysága. A Fűrész algoritmus más problémák megoldásában is használható, ahol különböző taxonómiák közt kell leképezést meghatározni, például különböző besorolási rendszert használó szabadalmi hivatalok alkalmazhatják a taxonómiák összehangolására.

Úgy érezzük, hogy eljárásunk sikeresen szerepelt, tekintve hogy először vettünk részt a KDD kupán, kezdetben a HITEC-en kívül nem állt rendelkezésre más segédeszköz, valamint hogy a díjazott csapatok közül a miénknek volt a legkevesebb tagja.

### Köszönetnyilvánítás

Tikk Domonkost az MTA Bolyai János kutatói ösztöndíja támogatta. Jelen anyag elkészítését a Mobil Innovációs Központ is támogatta.

### Irodalom

- [1] HITEC categorizer online.  
<http://categorizer.tmit.bme.hu>
- [2] M. F. Porter:  
An algorithm for suffix stripping.  
Program, 14(3):130–137., July 1980.  
[http://telemat.det.unifi.it/book/2001/wchange/download/stem\\_porter.html](http://telemat.det.unifi.it/book/2001/wchange/download/stem_porter.html)
- [3] G. Salton, M. J. McGill:  
An Introduction to Modern Information Retrieval.  
McGraw-Hill, 1983.
- [4] R. E. Schapire, Y. Singer, A. Singhal:  
„Boosting and Rocchio applied to text filtering”,  
Proc. of SIGIR-98, 21st ACM Int. Conference on  
Research and Development in Information Retrieval,  
pp.215–223., Melbourne, Australia, 1998.  
<http://citeseer.ist.psu.edu/schapire98boosting.html>
- [5] Shen et al:  
An ensemble search based method for  
query classification.  
<http://q2c.cs.ust.hk/q2c/Readme.pdf>
- [6] D. Tikk, Gy. Biró, J. D. Yang:  
„A hierarchical text categorization approach and  
its application to FRT expansion”,  
Australian Journal of  
Intelligent Information Processing Systems,  
8(3):123–131., 2004.
- [7] D. Tikk, Gy. Biró, J. D. Yang:  
„Experiments with a hierarchical text categorization  
method on WIPO patent collections”,  
In: N. O. Attok-Okine and B. M. Ayyub, editors,  
Applied Research in Uncertainty Modelling and  
Analysis, no.20, International Series in  
Intelligent Technologies, pp.283–302., Springer, 2005.
- [8] D. Vogel et al,  
„Classifying search engine queries using the web  
as background knowledge”,  
SIGKDD Explorations (megjelenés alatt).  
[http://www.medai.com/publications/pdf/vogel\\_kddcup\\_2005.pdf](http://www.medai.com/publications/pdf/vogel_kddcup_2005.pdf)

# A WTC 2006 konferencia szubjektív értékelése

LAJTHA GYÖRGY

*lajtha.gyorgy@t-com.hu*

*A World Telecommunications Congress (WTC) Budapesten tartotta ez év május 1-3. között első kongresszusát. A WTC ugyanis csak körülbelül másfél évvel ezelőtt alakult meg, amikor két korábban működő szervezet, az ISS (a Nemzetközi Kapcsolástechnikai Szimpózium) és az ISSLS (a Nemzetközi Kapcsolás és Helyi hálózati Szimpózium) elnöksége úgy döntött, hogy egyesülnek és a jövőben közösen tartják szimpóziumukat. Az új szervezet a címben szereplő új nevet választotta.*

Az egyesülés hátterében két világméretű jelenség húzódik meg. Az első szakmai jellegű és az eszközök egységesítésével, integrációjával függ össze. Nehéz ugyanis ma már élesen megkülönböztetni a hálózati síkokat, a különböző hálózati eszközöket, mert felhasználótól felhasználóig digitális bitfolyamok haladnak a hálózaton. Ezeket szabványos routerek irányítják és a bitfolyamokat egységes berendezésekkel fogják össze, majd a végponton bontják szét. Érdeemes ezért a hálózati technikai kérdéseit közösen vizsgálni.

A másik tényező, hogy az éles versenyben kevesebb az olyan újdonság, amit szívesen tesznek közzé a gyártó cégek, vagy a szolgáltatók. A konkurrencia határok nélküli, az egész világon versenyben van és szeretné a legtöbb felhasználót megnyerni. A szolgáltatókat sem védik már az országhatárok. Ennek következtében a kiemelkedően új gazdaságos megoldások évek óta nem szerepelnek a konferenciák programjában, hanem csak a már bevezetett eljárásaikat ismertetik.

E két tendencia együttesen azt eredményezte, hogy a konferencia résztvevői nem visznek haza olyan ötleteket, megoldásokat, vagy eljárásokat, melyeket saját távközlési iparuk, vagy szolgáltatásuk közvetlenül hasznosítani tudna. Az eredmények áttételesen jelentkeznek, tanulságokat lehet leszűrni és személyes kapcsolatok révén lehet megismerni a különböző távközlési vállalatok elképzeléseit. A külföldi kapcsolatok megteremtése azonban továbbra is jelentős, mert a személyes tapasztalatcsere még mindig sokkal gyorsabb, mint a kiadványokban megjelenő ismertetések.

E problémák vezettek oda, hogy több nemzetközi konferenciasorozat megszűnt, mások szövetkeztek vagy összeolvadtak, hogy nagyobb érdeklődési kört fedjenek le. Ezen fejlődési irányok eredményezték a WTC megszületését is és ennek jegyében tartották Budapesten első kongresszusukat. A következőkben néhány főbb irányzatot igyekszünk bemutatni abban a reményben, hogy azok a hazai kutatók, fejlesztők, szolgáltatók célkitűzéseit és terveit kedvező irányba terelik.

## QoS

A szolgáltatás minőségével világszerte foglalkoznak, így például a PKI is ebben a témában szervezett ez év februárjában workshop-ot. A jelen konferencián is há-

rom szekcióban ez volt a központi téma. Ezen kívül a plenáris ülések előadói is szívesen tértek ki a távközlés, a szolgáltatás minőségére.

Az előadások egyik csoportja a mobilitás és a minőség kapcsolatát vizsgálta. Itt előre kell bocsátani, hogy az ITU ajánlásokban a mobil rendszerek kedvezményt kapnak, mert a felhasználók a mobilitás érdekében esetleg gyengébb minőséggel is megelégszenek. Ez a gondolat volt a központi témája az egyik szekciónak, ahol a mobil rendszerek, az ad-hoc hálózatok és a mindenki által elérhető csomópontok (Hot Spots) minőségi követelményeit és az elérhető optimumot vizsgálták a szerzők.

A téma bonyolult, ez abban is megmutatkozott, hogy általában nem egy-egy kutató vagy fejlesztő eredményeit mutatták be, hanem kutató csoportok közös munkája hozott eredményt. Az első csoport, aki a témában jelentkezett, az Eurescom zászlaja alatt végzett kutatást, melyben német, spanyol és az európai NEC kutatói vettek részt. Módszereikkel azt igyekeztek elérni, hogy mobil rendszerekben a haladás folyamán, vagy amikor egy-egy bázisállomás határait átlélik, ne legyen észrevehető minőségromlás. Az eredmények előremutatóak és a leendő 4G mobil rendszerek tervezésénél és üzemeltetésénél használhatók.

Az előadók néhány szóval elmondták, hogy mit értenek a minőség alatt és milyen tartalommal használják a következőkben az elterjedt rövidítést, a QoS-t. A definíciók helyesnek tűnnek, de nem minden szerző használta azonos értelemben. Véleményem szerint szükséges lenne ezért, hogy a QoS tartalmát széles körben publikálják. Ennek során a QoS-t befolyásoló paraméterek elégséges és kiváló értékeit kellene megadni, vagyis rögzíteni kellene a bithibaarány, a szótévesztés, a késleltetés és a dszitter értékeit. Erre vonatkozóan már vannak kiinduló adatok, azonban a következő 1-2 évben a korábbi paramétereknél kialakult referencia összehasonlítás bevezetésével ezek felosztására is érdemes lenne javaslatot tenni.

A WTC során számos tanulságos előadás hangzott el, melyek a QoS alkalmazását mobil rendszerekben változó hozzáférésű összeköttetéseknel igyekeztek megadni. Ezen kiinduló nézetek alapján elképzelhetőnek látszik, hogy a jelenlegi hálózati struktúrákra is mind fix,

mind mobil szakaszokat figyelembe véve elkészüljön a referencia összeköttetéshez rendelt paraméterek várható és szélső értékének nemzetközileg elfogadható listája.

### NGN

A címben szereplő betűszó kifejtése filozófiai problémákat vet fel, mert bár a *G* a Generation és a második *N* a Network egyértelmű és jól meghatározható, az első *N*, amelyik lehet akár Next, akár New nem ad egyértelmű támpontot arra, hogy mikortól és mire vonatkozik az elkészült javaslat, vagy előírás. Tétélezzük fel, hogy mind a két jelző egyértelműen azt jelenti, amire a köznapis olvasó gondol, vagyis a legközelebbi jövőben üzembe helyezendő eszközökre vonatkozik. Ilyen formában a jelenleg fejlesztési stádiumban, vagy gyártásban lévő eszközök specifikációját igyekszik az NGN rögzíteni.

Az előadók az NGN megvalósításának időtávatát és esetleges végleges kiteljesedését különböző meghatározások segítségével igyekeztek megadni. Valamennyien kizárták a nagyobb távlatokat az NGN fogalmából, mert 5-10 év alatt olyan lényegi változások következhetnek be, melyek jelenleg nem láthatók és hatásuk nem jósolható. A meglévő és már telepített eszközöknél pedig nincs értelme új követelmények megfogalmazásának, hanem a most fejlesztendő eszközöknek kell alkalmazkodni a működő hálózathoz.

Ezek a gondolatok legtöbb NGN előadásban tükröződtek. A vendég előadó a Deutsche Telecomtól elkerülte a technológiára vonatkozó nehéz kérdéseket. Előtérbe helyezte viszont az NGN-en létrehozható új szolgáltatásokat, az illeszkedést a különböző alapáramkörökhez és a kapcsolódást a meglévő hálózathoz. Hangsúlyozta, hogy az NGN az IP világában igyekszik garantálni a minőséget. Erre azért van szükség, mert az új hálózat gyors elterjesztése érdekében mindenütt próbálják az árakat csökkenteni, ami bár dicséretes törekvés, de nem mehet a szolgáltatások rovására.

Az NGN-nel kapcsolatos szolgáltatások megjelentek azokban a szekciókban is, melyek a konvergenciára, vagy a forgalomirányításra vonatkoztak. A WTC mindhárom napját áthatotta az új generációval kapcsolatos különböző gondolatok bemutatása. Nem volt teljes a konvergencia ezek között az előadások között, hiszen különböző kiindulópontokból, különböző technológiákra építkezve eltérő módon juthatunk el a következő években távlatilag kialakuló hálózatokhoz.

### Mobilitás

A harmadik központi gondolat a mindenütt, mindenkor, mindenhol elérhető távközlés volt. Ennek alapja már mindenütt meg van és a mobil szolgáltatást a lakosság nagy része mind Európában, mind Észak-Amerikában használja. Ismeretesek adatok a távol-keleti terjedésről is.

A rendszer széleskörű alkalmazása ellenére voltak újdonságok, melyek a következő évek fejlesztését meghatározzák. Ezek közül első a mobilitás és a szélessáv összekapcsolása, valamint a címezéssel, irányítással kap-

csolatos problémák bemutatása. Különösen érdekes volt az utolsó napon a Molnár Sándor által vezetett Routing szekció. Ennek előzményei korábbi szekciókban már megjelentek, amikor az Internet Protokoll alapján kialakuló hálózatok és ezen belül az irányítás és a számlázás problémája előtérbe került.

A mobilitás kérdése szorosan összefüggött a sáv-szélesség és a hibavédelem problémáival. A jelenlegi technológiával mindezek megoldhatók és számos országban a tapasztalatok azt mutatják, hogy ezen többszolgáltatások ellenére a megszokott és elvárt minőség biztosítható.



### Összefoglalás

A háromnapos előadássorozat jól illeszkedő gondolatokkal mutatta be a távközlésfejlesztés aktuális kérdéseit. A sok érdekes témát négy párhuzamos szekcióban tárgyalták meg. A hallgatóságnak csak az a problémája volt, hogy mikor és melyik szekciót látogassa? Talán remélhető, hogy egy összefoglalóban a szervezők bemutatják a különböző fejlesztések együttes hatását.

Visszatekintve jól látszik, hogy három irányban fejlesztik az IP hálózatokat:

- általánossá válik a QoS iránti igény,
- szélessávú rendszereket telepítenek,
- az igények általános kielégítésére fényvezetős hálózatokat építenek ki.

Ezek együttesen eleget tesznek az üzleti igényeknek és kialakul egy jól felügyelhető infrastruktúra. Ezen különbözőnek tűnő tendenciák együttese vezethet el valamilyen eddig nem létező új hálózathoz.

Itt is, mint eddig többször hangsúlyozni kell, hogy a meglévő hálózat olyan jelentős értéket képvisel, mely az új gondolatokkal szemben igyekszik konzerválni a különböző távközlési rendszereket és eszközöket.

Talán sok résztvevőben felmerült a gondolat, hogy mindezen újdonságokat gyorsan be kellene vezetni, közben azonban nem szabad elveszteni azokat az értékeket, melyek a hálózatban már megvalósultak. A WTC három napja felhívta a figyelmet arra, hogy sok új szolgáltatást lehet felkínálni, ezeket azonban esetleg gazdasági megfontolások fékezhetik. Ha hasznosíthatók a meglévő eszközök, akkor hamarosan mindenki számára szélessávú, nagy megbízhatóságú összeköttetéseket lehet kínálni. Ezek a gondolatok más konferenciákon is előtérbe kerülnek és a Mbit/sec sebességű hálózatok kiépítése a cél.

**Collection and research of processing methods of European multi-lingual news database for automatic processing of multimedia broadcasts**

*Keywords: digital signal processing, speech processing, speech databases*

This paper investigates the problem of the transcription of Broadcast News (BN) database. The research has been prepared as a member of the Broadcast News Interest Group of COST278. The Broadcast News Interest Working Group is involved 10 institutions from Europe. It is known that the interpretation of the annotation and transcription rules can vary from annotator to annotator; therefore the first aim of the workgroup is to join together and discuss the rules of the transcription methods and their interpretation. The evaluation of the transcribed data has been prepared with tools developed by the Portugal and Belgian members of the BN Interest Group of Cost278 (INESC or ELIS). The second aim of the Working Group was to develop different language independent methods: for speech detection, for speaker clustering, for gender detection, for jingle detection, etc.

**Algorithm for tracking, modifying and synthesis of formants**

*Keywords: formant tracking, formant synthesis, linear prediction, speech character modification*

Precise formant tracking has been a challenge for researchers in speech processing for long. In this paper, the authors present a method to track and modify formants in speech signal. It is an efficient tool for analyzing and altering the spectral content of speech, furthermore it provides an opportunity to modify timbre and voice quality. The method is based on the linear prediction model.

**The role of the acoustic company image in the evaluation of infocommunication service providers**

*Keywords: acoustic company image, IVR, quality of service in infocommunications, number and date reading*

This paper is an attempt to extend the well known „company image“ with regard to the acoustic appearance. The author introduced the term „acoustic company image“ in 1998. When creating their acoustic image, organizations should systematically plan and implement the features and quality of acoustic appearance toward the outside world. The paper provides a general overview of the topic, introduces general definitions and illustrates the importance of the acoustic image by specific examples of infocommunication services.

**An MPEG-4 based model for mouth animation**

*Keywords: audio-visual speech processing, face animation, multimodal communication*

In this paper the coding principles of the MPEG-4 standard is summarized especially that part which is re-

lated to the face and body animation. The open source coded LUCIA decoding model is discussed and one of its special applications. In this application the motion of the face model is controlled by speech signal parameters. This application serves as communication aid for deaf persons. The results of communication tests are also discussed.

**Estimation of the instantaneous parameters of speech signals based on the Teager operator and Hilbert-Huang transform**

*Keywords: Teager operator, Hilbert-Huang transform, instantaneous amplitude and frequency*

To investigate the fine structure of speech signals, methods are needed to determine the non-linear and non-stationary properties. This paper introduces Teager operator and Hilbert-Huang transform (HHT) as signal processing methods suitable for estimating the instantaneous amplitude and frequency. The estimates obtained by using HHT and Teager operator are compared for the case of test signals and real speech signals.

...

**The new radiation system of the extended Jászberény Radio Station and determination of the optimal operating frequency plan**

*Keywords: wave propagation, radio broadcasting, short waves, optimal operating frequency*

This paper presents the short story of Jászberény SW station and its antennas, after that it shows how to determine the optimal operation frequency for different target areas by three different methods.

**The birth of the radio communications in Hungary**

...

**Solution of the 2005 KDD cup problem with Ferrety-algorithm**

*Keywords: classification of internet user queries, taxonomy mapping, text categorization*

The goal of the 2005 ACM KDD cup contest was classification of internet user queries. This paper presents our solution to the problem, which achieved precision and creativity runner-up awards at the competition. Our approach consists of three main parts: (i) a central knowledge filter (ii) an on-demand web crawler and (iii) a very efficient categorizer system. The outlined solution proposes a general tool for ill-posed categorization problems with insufficient training data. This problem is solved via taxonomy mapping using the Internet as a knowledge base.

**Subjective evaluation of the WTC2006 Conference**

## Az üvegszálalás átviteli hálózat változó világunkban

Manapság már világméreteket ölt a távközlés összefonódása. Ez az átalakulás egyre intenzívebbé válik. Mind a hagyományokkal rendelkező, mind a most induló távközlési cégek Észak-Amerikától és Európától Ázsiáig nagy erővel dolgoznak azon, hogy üzleti stratégiájukat úgy alakítsák ki, hogy minél nagyobb teret tudjanak hódítani a világban, figyelembe véve a versenyminták változásait is.

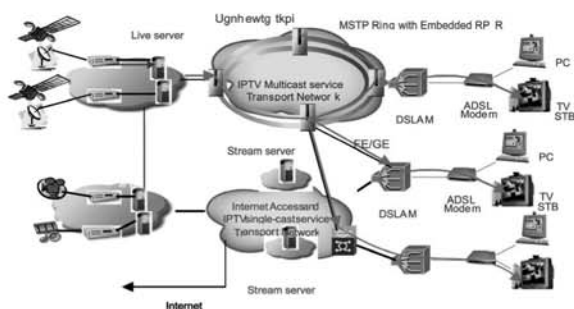
Hosszú fejlődési folyamat után a vonalas telefonhálózat telítődött, az előfizetők száma és a bevételek mértéke csökkent; a pusztán szélessávú előfizetői lehetőség már nem elégséges a multimedia szolgáltatások megnövekedett igényei számára. Ehhez társul még az a tény is, hogy az Internet szolgáltatások hatása a hagyományos távközlési szolgáltatások esetében csökkentette a távközlési cégek bevételeit és profitját is. Ahhoz, hogy véget vessenek ennek a kedvezőtlen helyzetnek, miszerint "a felhasználók száma növekszik, a bevétel viszont nem", a távközlési cégek aktív szerepet vállalnak abban, hogy együttesen kifejlesszék a hang-, adat- és képi szolgáltatásokat. Folyamatosan újabb és újabb szolgáltatásokat dobnak piacra, fontos hivatkozási helyre vonatkozó szolgáltatási típusokat, beleértve a 3G, NGN, IPTV, SAN, VIP magánhálózatot is.

Mivel az infrastrukturális hálózat állandóan újabb és újabb lehetőségekkel és kihívásokkal találja magát szembe, az üvegszálalás átvitelnek egyértelműen alkalmazkodnia kell a szolgáltatás típusának változásaihoz annak érdekében, hogy a szolgáltatás hatékonyan megvalósulhasson.

Az új átviteli szolgáltatás fejlesztésének követelményei

Milyen az átvittel szemben támasztott követelményeket kell a szolgáltatónak teljesítenie ebben az állandóan változó világban? Ahhoz, hogy jobban megértsük az átviteli hálózat átalakításának követelményeit, alábbiakban röviden elemzünk néhány olyan szolgáltatást, mely ebben a folyamatban felmerül:

**IPTV:** az IPTV szolgáltatás olyan media szolgáltatás, amely szélessávú hálózatra épül, képes a többfunkciós átvitelre és a VoD-re. Az átviteli hálózatnak képesnek kell lennie nagy kiterjedésű szétosztó hálózat működtetésére (CDN) az alaphálózattól a DSLAM-ig. Az IPTV magas sávszélességet követel meg, egy csatornával 2M-ig. Többszörös programcsomagokat lehet egyidejűleg továbbítani, miközben a szolgáltatás QoS követelményrendszere maximálisan teljesül.



**3G:** a harmadik generációs mobil kommunikációs technológia (3G) képes arra, hogy magasabb sávszélességet nyújtson a mobiltelefonos előfizetők számára, hang-, adat- és képi szolgáltatásokkal egybekötve. A Nagysebességű Összekötő Csomag Hozzáférési (High-Speed Downlink Packet Access - HSPDA) technológia bevezetésével a 3G hozzáférési hálózat magasabb átviteli sávszélességet igényel, az

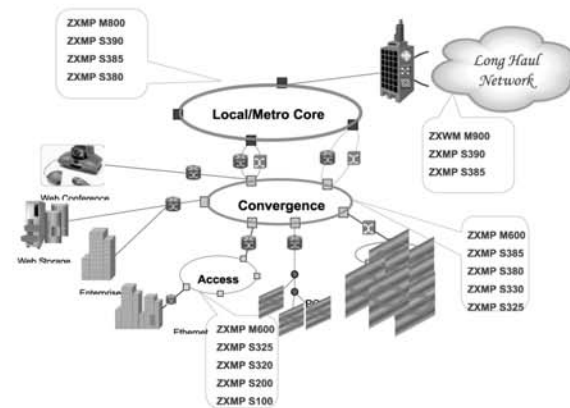
adatt forgalom megszakadása, illetve bizonytalansága egyre inkább nyilvánvalóvá válik. Ez az oka annak, hogy a 3G hozzáférési réteg átviteli hálózatának először a sávszélesség kihasználtságának rátáját kell megnövelnie, azzal egyidejűleg, hogy QoS garanciát nyújt a hangszolgáltatásra.

**SAN:** a SAN szolgáltatás egyfajta adattárolási szolgáltatás, ami távoli katasztrófa-elhárítási adathárteret biztosít nagy sebességgel, fontos intézmények számára. Miután magas szolgáltatási sávszélességgel és szigorú szolgáltatásbiztonsági követelményekkel bír, a SAN szolgáltatás olyan átviteli hálózatot igényel, ami adatcsatornákat és csatlakozók sorát biztosítja nagy fokú megbízhatósággal párosulva.

**VIP magánhálózat:** a VIP a szolgáltató cégek figyelmének középpontjában van. A VIP magánhálózatban, mivel a sávszélességi követelmények változnak a többfokozatú QoS-sal, az átviteli hálózatnak képesnek kell lennie arra, hogy magasfokú titoktartást biztosítson és megfeleljen a CoS, QoS különféle követelményeinek.

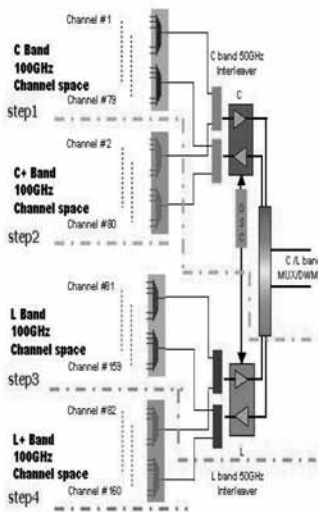
### A ZTE újgenerációs üvegszálalás hálózati megoldása

Az átviteli hálózat a szolgáltatások fejlődésével együtt fejlődik. Az új technológiák, mint például a RPR, MPLS, ASON, ROADM bevezetésével az újgenerációs hálózat egyre érettebb lesz. Erős R&D csapattal a háta mögött, egy ügyfél-orientált szolgáltatói rendszerrel a ZTE Cég továbbra is haladó megoldásokat kínál annak érdekében, hogy tökéletes és kitenő szolgáltatásokat nyújtson ügyfelei számára, nagy hangsúlyt fektetve a hálózatépítés költségeire és a működés hatékonyságára.



### \* A különböző hálózati rétegekre vonatkozó különféle megoldások

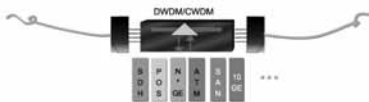
A MAN-ban és a hozzáférési réteg átviteli hálózatban, a számos új szolgáltatás megjelenésével a hozzáférési sávszélesség gyors és állandó használata megnövekedett, ami odáig vezetett, hogy a nagy teljesítményre és a magas sávszélesség használatára egyre nagyobb szükség van. A 155MHz és a 622MHz berendezéseket a jelenlegi hálózat már nem tudja fenntartani. A ZTE Cég Compact 2.5GHz MSTP üvegszálalás átviteli berendezése valamint a CWDM berendezések alkalmasak ennek a problémának a megoldására. A konvergencia és az alapréteg hálózat érdekében a legalapvetőbb követelmény a nagy kapacitás. A ZTE DWDM rendszer képes 1.6T-os vagy még nagyobb sávszélességi átvitelt nyújtani annak érdekében, hogy teljesen kielégítse az egyre bővülő szolgáltatásokhoz kapcsolódó követelményeket.



A nagyon hosszú távra épülő technológiát főleg amiatt kell alkalmazni, hogy kiküszöböljük a non-REG távolság miatti korlátozásokat és nagyban csökkentjük a regenerátor megépítésének és karbantartásának költségeit. Az ipar mindig a nagyon hosszú távolságra összpontosított, az olyan technológiák felbukkanása mint a Raman amplifier (hangosító), Hi-power (magasfeszültségű) OA, Remote pump(távolsági pumpa), RZ code (kód), erősítő/fokozó FEC lehetővé teszik a szétszóródást és az átvitelt, ami rendkívül nagy kapacitást igényel. A ZXMP M900 a ZTE Cég nagyon nagy távolságú és nagy kapacitású DWDM rendszere, amely meg tud valósítani egy több mint 3000km non-REG átvitelt, valamint nagyvolumenű hálózati módokat tud kiszolgálni.

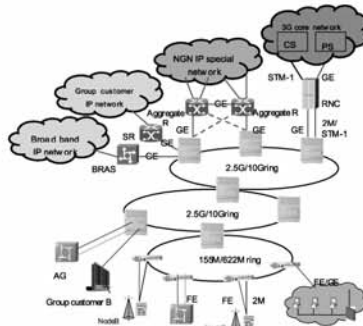
**\*Többfunkciós szolgáltatások és hatékony megoldások**

A MSTP és a DWDM rendszernek nagyon sok csatlakozó típusa van, melyek támaszt jelentenek az új, jövőbeli szolgáltatások csatlakozói számára. Különösen a DWDM rendszer jelent kitűnő, többfunkciós szolgáltatási vivőrendszert a teljes átláthatósága miatt.



A ZTE MSTP magába foglalja a beágyazott RPR/MPLS technológiákat, intelligens adatszolgáltatási folyamatokat valósít meg, nagyban csökkenti a fázisú manuális konfigurációt és szórás, ily módon gyors és rugalmas adatszolgáltatást biztosít. A sávzélességeket kívánság szerint osztja szét és igazítja a rendszerhez, amivel nagyban hozzájárul a hálózat működésének hatékonyságához. Ráadásul, az ASON mechanizmus bevezetésével az átviteli hálózat és a sávzélesség kívánság szerint elhelyezhető irányítás útján és jelrendszerek

segítségével a jobb adatszolgáltatás érdekében.



**\*Széles skálájú és többfunkciós hálózati megoldások**

Ahogy a hardware technológia fejlődik, az átviteli berendezések erős keresztkapcsolási és hálózati kapacitást nyújtanak. Jelenleg az SDH berendezés központjának egyetlen alfogása 640G keresztkapcsolási kapacitással bír (egészen 2.56T-ig többszörös lépcsőzetes alfogás útján), egyetlen berendezés 32 10G gyűrűt tud megformálni, ily módon teljes egészében megfelel a mai követelményeknek, ami a keresztkapcsolási és hálózati szolgáltatásokat illeti. Ahogy az ASON technológia egyre inkább elterjed, az olyan fejlett technológiák mint az automata topológia felfedezése széles körben el fog terjedni annak érdekében, hogy még inkább csökkentse a kézi konfigurálást a hálózati munkában és még inkább kiszolgálja a rugalmas és gyors hálózati munkát.

**\*Ügyfél orientált, differenciált szolgáltatások megoldása**

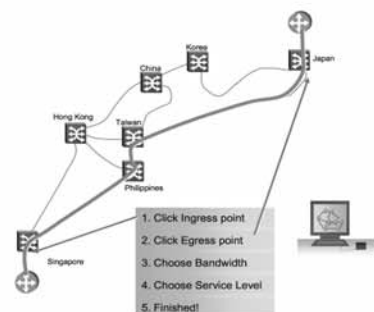
A MSTP hálózatba beágyazott RPR/MPLS 50ms gyűrűvédelmet, megfelelő sávzélességi elhelyezést, az adatszolgáltatás statisztikai multiplexing funkcióját, kényelmes jelölési kapcsolást, MESH-alapú, gyors átirányítást, erőteljes forgalom lebonyolítást, illetve CoS/QoS garanciákat nyújt, ily módon teljesen megfelel a jelenlegi QoS szolgáltatási követelményeknek.

Az ASON rendszer erős SLA (Service Level Agreement-Szolgáltatási Szint Egyezmény)

funkciót is ellát, amelyben különböző COS-t lehet elosztani. Különböző védelmi és hibaelhárítási stratégiákat lehet meghatározni a különböző COS-ok segítségével. A legmagasabb szintű szolgáltatások számára 1+1 pótló védelmi és hibaelhárítási szolgáltatást kell létrehozni; az alacsonyabb szintű COS szolgáltatásoknak is létezik a leghatékonyabb védelmi és hibaelhárítási stratégiája. Ily módon differenciált szolgáltatásban részesülnek az ügyfelek annak érdekében, hogy a jövőben lépést tudjanak tartani a kommunikációs hálózat fejlődésével.

**\*Kényelmes és gyors szolgáltatást nyújtó és karbantartó megoldások**

A ZTE üvegcszál hálózati termékei saját fejlesztésű protokollal működnek annak érdekében, hogy végponttól végpontig terjedő szolgáltatást nyújtsanak a saját berendezéseik segítségével a TDM szolgáltatás korlátozásai nélkül. Az RPR/MPLS új generációi, melyek MSTP berendezésekbe vannak beágyazva meg tudják valósítani a végponttól végpontig terjedő adatszolgáltatást, megkönnyítik az új szolgáltatások szállítását. Ahogy az ASON technológia egyre inkább terjed, az ASON termékeknek, melyeket több cég forgalmaz, meg kell tudni valósítani a végponttól végpontig terjedő szolgáltatást minden ASON hálózatban. Az ágazat közös érdekei és erőfeszítései lehetővé teszik azt, hogy elérjük ezt a célt.



A ZTE folyamatosan dolgozik az átviteli hálózat újabb és újabb generációinak kifejlesztésén annak érdekében, hogy a kommunikációs szolgáltatás adott irányzatait követni tudja, azonkívül teljes és mindig megújuló hálózati megoldásokkal kell előállnia ügyfelei számára.



# Contents

<i>FOREWORD</i>	2
<b>Csaba Teleki, Klára Vicsi</b> Collection and research of processing methods of European multi-lingual news database for automatic processing of multimedia broadcasts	3
<b>Tamás Bőhm, Géza Németh</b> Algorithm for tracking, modifying and synthesis of formants	11
<b>Géza Németh</b> The role of the acoustic company image in the evaluation of infocommunication service providers	17
<b>György Takács, Attila Tihanyi, Tamás Bárdi, Gergely Feldhoffer, Bálint Srancsik</b> An MPEG-4 based model for mouth animation	22
<b>István Pintér</b> Estimation of the instantaneous parameters of speech signals based on the Teager operator and Hilbert-Huang transform	28
<b>István Standeisky, György Dósa, Éva Balla</b> The new radiation system of the extended Jászberény Radio Station and determination of the optimal operating frequency plan	38
<b>György Dósa</b> The birth of the radio communications in Hungary	46
<b>Zsolt Kardkovács, Domonkos Tikk, Zoltán Bánsághi</b> Solution of the 2005 KDD cup problem with Ferrety-algorithm	50
<b>György Lajtha</b> Subjective evaluation of the WTC2006 Conference	59

Cover: Main building of the Hungarian Academy of Science, Budapest

---

## Szerkesztőség

HTE Budapest V., Kossuth L. tér 6-8.  
Tel.: 353-1027, Fax: 353-0451, e-mail: info@hte.hu

## Hirdetési árak

1/1 (205x290 mm) 4C 120.000 Ft + áfa  
Borító 3 (205x290mm) 4 C 180.000 Ft + áfa  
Borító 4 (205x290mm) 4 C 240.000 Ft + áfa

## Cikkek eljuttathatók az alábbi címre is

Szabó A. Csaba, BME Híradástechnikai Tanszék  
Tel.: 463-3261, Fax: 463-3263  
e-mail: szabo@hit.bme.hu

## Előfizetés

HTE Budapest V., Kossuth L. tér 6-8.  
Tel.: 353-1027, Fax: 353-0451  
e-mail: info@hte.hu

## 2006-os előfizetési díjak

Közületi előfizetők részére: bruttó 30.450 Ft/év  
Hazai egyéni előfizetők részére: bruttó 6.800 Ft/év  
HTE egyén tagok részére: bruttó 3.400 Ft/év

## Subscription rates for foreign subscribers:

12 issues 150 USD,  
single copies 15 USD

www.hte.hu

Felelős kiadó: NAGY PÉTER  
Lapmenedzser: Dankó András

---

HU ISSN 0018-2028

Layout: MATT DTP Bt. • Printed by: Regiszter Kft.