

Fonetikai algoritmus a hanghatárok gépi meghatározásának javítására nagyméretű beszédadatbázisokban

Olaszy Gábor

BME Távközlési és Médiainformatikai Tanszék
olaszy@tmit.bme.hu

Kivonat: A beszédtechnológiai kutatásokhoz és gyakorlati alkalmazásokhoz egyre nagyobb méretű beszédadatbázisokat terveznek. Ezek egyik fajtája, amikor előre meghatározott szöveg felolvasásával hozzák létre a több órányi beszédanyagot. Ilyen adatbázis például a BME TMIT időjárás jelentéseket tartalmazó beszédkorpusza, amelyik 5400 mondatot tartalmaz. Ahhoz, hogy gépi módszerekkel lehessen ezt a hanganyagot a későbbiekben feldolgozni (szavakat, hangkapcsolatokat keresni stb.), a szöveget át kell írni hangszintű szimbólumsorozattá, majd jelölni kell a hanghullámon a hangokat, azok határát, a szavak kezdetét, a szüneteket, valamint szinkronba kell hozni a szöveget a hangzó formával automatikus gépi felismerés segítségével. Az így jelölt hangok és hanghatárok csak mintegy 95%-os pontossággal adnak kellően helyes eredményt. Ez azonban nem elegendő a jó minőségű, korpusz alapú beszédszintézishez. Ebben a tanulmányban egy olyan utófeldolgozó algoritmust ismertettünk, amelyikkel növelni lehet a pontosságot, ezzel a szintézis minőségét.

1 Bevezetés

A beszédtechnológiai kutatásokhoz és gyakorlati alkalmazásokhoz egyre nagyobb méretű beszédadatbázisokat terveznek. Ezek egyik fajtája, amikor előre meghatározott szöveg felolvasásával hozzák létre a több órányi beszédanyagot. Ilyen adatbázis például a BME TMIT időjárás jelentéseket tartalmazó beszédkorpusza [2], amelyik 5400 mondatot (95678 szó) tartalmaz. Ahhoz, hogy gépi módszerekkel lehessen ezt a hanganyagot a későbbiekben feldolgozni (szavakat, hangkapcsolatokat keresni stb.), a szöveget át kell írni hangszintű szimbólumsorozattá, majd jelölni kell a hanghullámon a hangokat, azok határát, a szavak határát, valamint a szüneteket. Ezeknek a jelöléseknek elméletileg szinkronban kell lenni az eredeti szöveggel. Ennek elérése gépi feldolgozással csak meghatározott hibaszázalékkal lehetséges. Ebben a tanulmányban a hanghatárok gépi kijelölésével kapcsolatos problémákkal foglalkozunk. Célunk az elméletileg lehetséges legoptimálisabb hanghatár-jelölés közelítése.

A gépi hanghatár-kijelölés egyik módszere a beszédfelismerésre kidolgozott technikát is alkalmazza, amikor a hangok sorozatát alapul véve (fonemikus átírat) irányított felismerést végeznek és bejelölik a hanghullámon a hanghatárokat, valamint a

szüneteket [3]. A módszer helyes működésének kritériuma, hogy a fonemikus átírat és a hanghullám között szoros egyezés legyen, aminek a teljesülését feltételezzük, mint kiindulási alapot. A felismerés eredménye a mondat kezdeti pontjának a megjelenés, továbbá pedig annyi hanghatár bejelölése a hangsorba, ahány hang van a fonemikus átíratban. A BME TMIT beszédkutató laboratóriumában is ilyen módszert alkalmaznak [3]. A beszédfelismerő alapvetően HMM alapú algoritmus. A jelölés eredménye a körülbelüli hanghatár minden megadott hangra. Méréseink szerint az így jelölt hanghatár a humán jelöléshez viszonyítva akár 20ms-os eltérést is mutathat. Ezekre a jelölésekre támaszkodnak az adatbázist különböző célokra felhasználó további algoritmusok (például a korpusz alapú beszéd-szintézis válogató modulja). A beszédfelismerés gyakorlati alkalmazásainál ez a pontosság kielégítő, azonban a beszéd-szintézises alkalmazás során kiderült, hogy a hanghatár helyének jelölését tovább kell finomítani. Például női hang esetén a 20ms-nyi idő körülbelül 4 zöngé periódust fedhet le. Sok esetben ezek a hanghatár-elcsúszások fonetikailag irreleváns helyzeteket hoznak létre (például egy zöngétlen hangban zöngés periódusok lesznek a jelölés szerint), illetve hangelcsúszás is előfordulhat, ami például egy beszéd-szintetizátor hangzásában zavaró. Mindezen hibák kiküszöbölésére felmerült a gondolat, hogy egy utófeldolgozó, finomító algoritmusra van szükség.

2 Anyag és módszer

A gépi címkézéshez és fonemikus hangjelöléshez HMM alapú beszédfelismerőt [3] alkalmaztunk, irányított felismerési technikával. Ennek lényege, hogy segítjük a felismerőt a döntéshozatalban a fonetikai átírat szerinti hangsor megadásával. A beszédhullámban tehát annak kezdetét jelöli meg az algoritmus, majd pontosan annyi hangot jelöl, ahány hang van az fonetikai átírat szerint a mondatban. Azt azonban a program nem tudja garantálni, hogy minden jelöléshez korrekt hang tartozik. Ez a kényszer – ritkán – helytelen hangazonosításhoz is vezethet. Ha például nem talál olyan akusztikai tartalomnak megfelelő részt a hullámformában, mint amilyen hang az átíratban szerepel, akkor kényszerből betesz egy hanghatárt a saját döntése alapján, csak azért, hogy a hanghatárok száma megegyezzen az előre megadott számmal. A szintézis során csak a bejelölt hanghatár jelekre támaszkodunk a szóválasztásnál, azok alapján vesszünk ki a hullámforma részeket a hanghullámból. Rossz helyzetű hanghatár rossz hangot eredményez.

A szöveg fonemikus átírását a beszédfelismerő végezte (a hangjelölések az ábrákon a számítógépes program belső kódjai szerintiek, ezeket az adott helyen magyarázzuk).

A hibafeltáráshoz a teljes szöveg- és hangkorpusz 5400 mondatából válogattunk ki 110 címkékkel ellátott mondatot, és manuálisan elemeztük azokat. Egyrészt a hangjelölések és az akusztikai tartalom megegyezését vizsgáltuk, másrészt pedig ellenőriztük a gépileg bejelölt hanghatárok pozíciójának helyzetét a hanghullámhoz viszonyítva. A hanghatárok humán megállapítása is nehézségekkel teli, különböző pontossággal tudjuk meghozni a döntést [4]. Viszonylag pontos meghatározás lehetséges, ha gerjesztésváltás van a két hang kapcsolódási pontjában. A hangintenzitás viszonyok hirtelen változása is jó támpont (például magánhangzó és zöngés zárhang találkozá-

sánál). A legnehezebb a hanghatár kijelölése az olyan hangkapcsolódási pontokon, ahol a hangok természetéből adódóan nincs jelentős akusztikai szerkezeti változás (például magánhangzó kapcsolatokban). Korrekt manuális hanghatár-jelölésekre példákat találunk a <http://fonetika.nytud.hu/cccc> honlapon.

Elemzéseinkhez az eredeti mondat hangzó változatát és az azzal szinkronban lévő gépi hangjelöléseket használtuk. Az analízis során a Praat 4.6 fonetikai programmal [1] egymás alatt megjelenítettük a mondat rezgésképét, spektrogramját, valamint a hozzá elkészített hangjelöléseket és hanghatárokat. Hangonként elemeztük a jelölés és az akusztikai tartalom helyességét, illetve hibás voltát. Az elemzésekhez felhasználtuk az 1. táblázat szerinti hangosztályozást. Négy olyan jellemző paraméter szerint osztottuk fel a hangsor elemeit, amelyek a hibajavító algoritmusban egyértelműen elkülöníthetők és felhasználhatók. Ezek: a gerjesztés formája, a képzési üreg, a hang intenzitása és végül a hang szerkezeti felépítése. A gerjesztés formája a fonemikus jelölés alapján is és az akusztikai elemzésből is rendelkezésre áll (a zöngés periódusok meg vannak jelölve a hanghullámban). Az üreget a jelölt hang alapján definiáljuk. A hang intenzitását pontosan tudjuk mérni, ennek a kategorizáláshoz három szintet használunk (intenzív, közepes, gyenge). Az intenzitás-szinteket a hangok akusztikai felépítéséből ismert adatok alapján rendeljük hozzá a vizsgált elemhez. A szerkezeti felépítést a fonemikus jelölés alapján két kategória szerint osztottuk meg: egyszerű és összetett. A beszédhangokat besoroltuk ezekbe a paraméter-csoportokba és 11 osztályt különböztettünk meg a kategorizáláshoz. Minden osztályon belül ugyanazok a paraméterek jellemzik a beszédhangokat azok artikulációs meghatározásától függetlenül.

1. Táblázat: a hangsor elemeinek kategorizálása 11 hangosztály, 4 paramétercsoport és ezeken belül összesen 10 paraméter szerint. A hangokat a betűjelükkel jelöltük. V=magánhangzó, h*=zöngés h hang.

	OSZTÁLY	1	2	3	4	5	6	7	8	9	10	11
	HANG	V	b d g gy	ny	p t k ty	m n	j h* v l	h f	z zs	sz s	c cs	r
PARAM. CSOPORT	PARAMÉTER											
1	zöngés	x	x	x		x	x					x
	zöngétlen				x			x		x	x	
	vegyes								x			
2	nazális			x		x						
	orális	x	x		x		x	x	x	x	x	x
3	intenzív	x										
	közepes			x		x	x		x	x	x	x
	gyenge		x		x			x				
4	egyszerű szerk.	x				x	x	x	x	x		
	összetett szerk.		x	x	x						x	x

3. Az analízis eredményei

Az elemzés során 4300 hanghatárt ellenőriztünk, ezekből 232 esetben kellett módosítást elvégezni. A talált hibákat két fő kategória szerint osztályoztuk: helytelen

hangjelölés, vagyis a jelölt hang teljes egészében nem felel meg a hangsorbeli akusztikai tartalomnak (H1). Ebbe a kategóriába tartozik a beszédhangokon kívül a szünet (sil) helyének vizsgálata is. A második kategóriába (H2) tartoznak azok a hibák, amikor a jelölt hanghatár nem a tényleges kapcsolódási ponton van, hanem attól lényegesen (több mint 10ms-mal) eltér. Ez utóbbi esetben a kapcsolódó hangok mindegyikéből szerepel egy-egy rész a két jelölt hanghatár közötti hangsorrészben.

A (H1) hiba kétféle okból keletkezhet: egyszerű szöveghibából, illetve beszédfelismerési tévesztésből. Szöveghiba, amikor például a felolvasandó szövegből egy betű kimaradt (1. ábra). A bemozdó az ilyen betűhibákat nem érzékeli, korrektül olvassa fel a szöveget. Ha a hibát nem veszi észre az első ellenőrzést végző személy sem, akkor a hullámforma és a szöveg fonemikus átírása között nem lesz hangszinkron egyezés. A hullámformában eggyel több a hang lesz, mint a fonemikus átiratban. Ilyen esetben a kényszerített beszédfelismerési algoritmusnak el kell döntenie, hogy a hangsor mely pontján „spórol meg” egy hangot.

Példa: *Az eleinte jellemző lehűlést az év utolsó napján a magasban melegedé váltja fel.*

A példából látható, hogy a felismerő úgy döntött, hogy a réshangot szünetként jelöli. Ilyen hiba detektálása esetén manuálisan ki kell javítani az adott mondat szöveges és fonemikus formáját. A szinkronitás így helyreáll. Egyébként, ha a program az így jelölt szünetet választja ki a szintézis során, akkor egy erős réshang fog hallatszani a szünet helyett. A szünet-jelzést a beszédfelismerő olyan helyeken is beiktathatja, amikor a beszélő glottalizál a hangkapcsolódás folyamán (2. ábra).

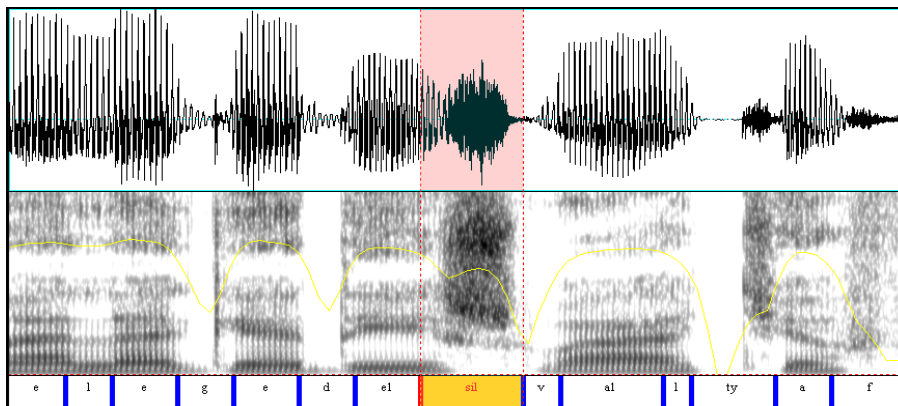


Fig 1. A fenti példamondat végén a *melegedés* szóban a réshang helyén szünet (sil) jel szerepel (szürke mező), mivel az eredeti szövegből kimaradt az s betű. (e1 = é; a1 = á). Az ábrán példát látunk a (2) hibakategóriára is, mivel a szünet hanghatára beelőz az előző magánhangzó végébe

Ezek valójában nem igazi szünetek, ezért nem szabad őket szünet funkcióra felhasználni (tehát ezeket is jelöléssel kell ellátni).

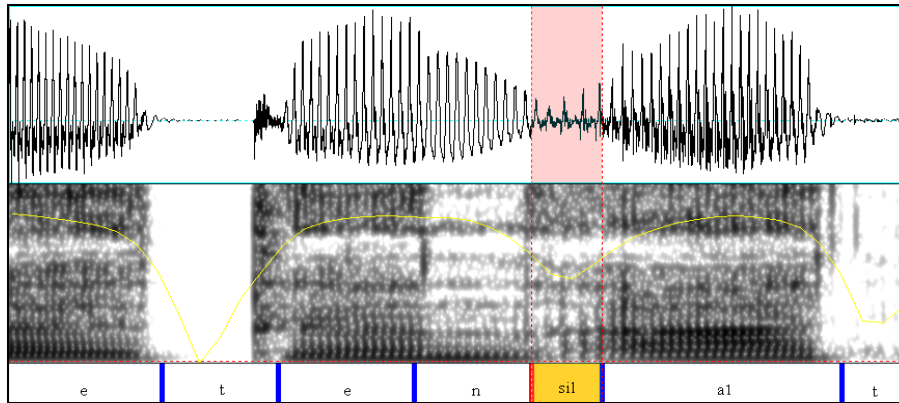


Fig. 2 A *Keleten átmeneti...* kezdetű mondat szóhatárán a bemondó glottalizált. Ezt a szakasz (szürke rész) a beszédfelismerő szünetnek jelölte (a1=á)

A (H1) típusú hibát okozhatja még bizonyos hangkapcsolódásokra jellemző hangszerkezeti változás. Erre példa az *ismét* szó réshang-nazális hang kapcsolatának hibás feldolgozása (3. ábra).

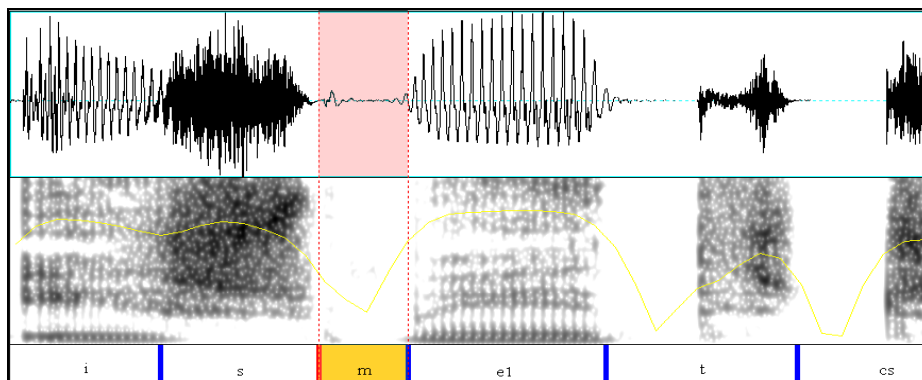


Fig 3. A koartikulációs néma fázis megzavarja a felismerést, ezért a gép erre a szakaszra sorolja be a teljes nazális hangot (szürke jelölés)

A felismerő a réshang végén kialakuló koartikulációs néma fázishoz [5] azonosította a nazális mássalhangzót, így ez a hang – a jelölés szerint – a réshang belsejébe került. A hiba következménye az is, hogy a hangsor szerinti nazális hang és az őt követő magánhangzó egyetlen hangként lesz jelölve. További példa a (H1) hibatípusra, amikor a beszélő egy svá töltelék elemet is ejt egy hang végén (4. ábra). Ez utóbbi a színészi ejtésnél gyakori (*csütörtökönö, pénteken*). Ekkor a felismerő a svára is tehet szünet jelölést (mást nem tehet rá, mivel a hangok száma kötött). Ezt a szünetet sem szabad felhasználni, mivel egy zöngés hang szól alatta.

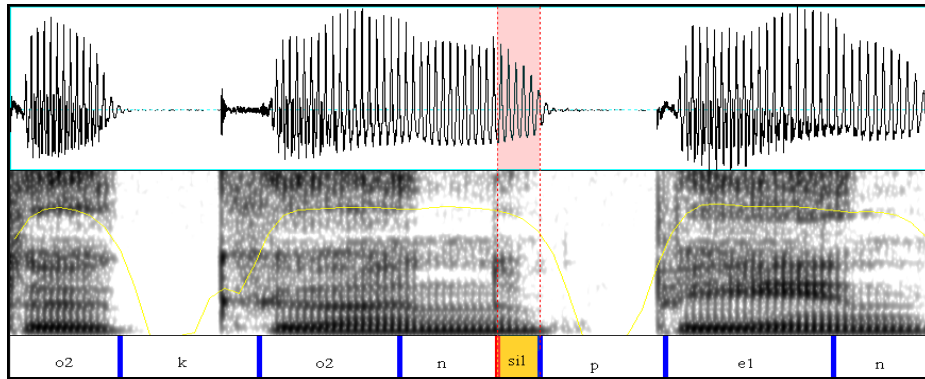


Fig 4. A nazális hang végén ejtett svá elemet a felismerő szünetnek jelölte a példában (szürke sáv), (o2=ö, e1=é)

A (H2) hibatípus a hanghatár eltolódását jelenti a hangon belül, illetve kívül. Ilyen hiba mind zöngés-, illetve zöngétlen hangok kapcsolódásánál, mind pedig zöngétlenek és zöngések találkozásánál előfordul. Nézzünk néhány példát. Az 5. ábra több hanghatár elcsúszását mutatja be zöngés hangok között.

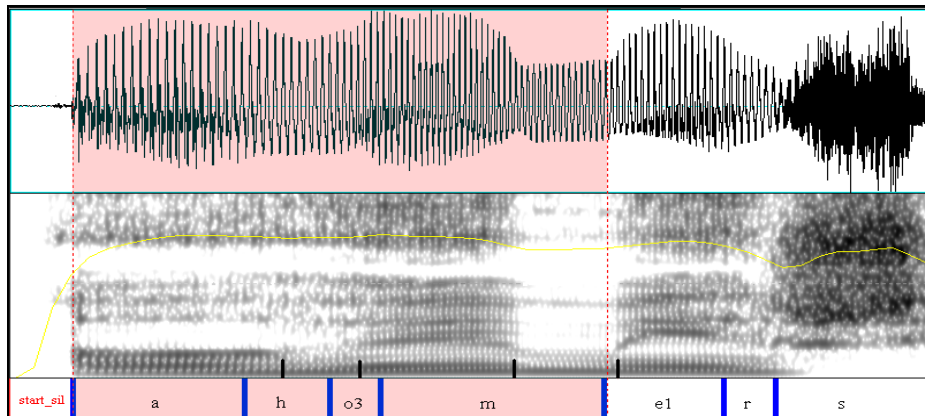


Fig 5. Példa a hanghatárok elcsúszására a *...hőmérséklet...* szó négy első hangjában. A helyes hanghatárokat a fekete jelölésekkel érzékeltetjük (o3=ö, e1=é)

A zöngés hangokban tapasztalt hangelecsúszást algoritmikusan nehéz kijavítani, detektálni azonban egyes esetekben lehet (lásd a példában is, ahol az első hosszú magánhangzó időtartama feltűnően rövid, mindössze 32 ms, így a hangidőtartamok ellenőrzésével felismerhető a hiba). Automatikusan lehet azonban javítani a zöngés-zöngétlen hangok találkozásánál tapasztalt hanghatár elcsúszást. A 6. ábrán egy tipikus példát mutatunk be erre a hibára is.

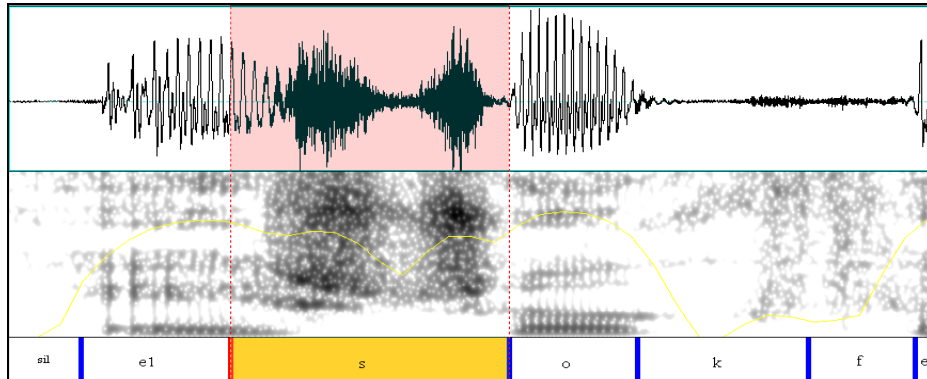


Fig 6. Hibás hanghatár jelölés az ...és sokfelé... hangsor első két hangján

4 A fonetikai javító algoritmus

A javító algoritmus nem vizsgálja a spektrális részleteket, más paraméterek alapján végzi az elemzéseket. Az algoritmus leírásában a következő elnevezéseket használjuk:

- fonemikus hang, vagyis a gépi hangátíró által a szövegből kikövetkeztetett hang;
- jelölt hang, vagyis a beszédfelismerő által meghatározott két hanghatár között elhelyezkedő kikövetkeztetett hang;
- beszédhang, vagyis a hanghullámban lévő, ugyanezen pozíciójú hang.

Hibátlan jelölés esetén a három elemnek szinkronban kell lennie az időtengelyen. Az algoritmusnak három feladata van: (F1) hangazonosítás, (F2) annak eldöntése, hogy a jelölt hanghatár jó helyen van-e, továbbá (F3) annak jelzése, ha egy hang a hangkörnyezetéből fakadóan annyira megváltoztatja a szerkezetét, hogy önálló hangként nem használható fel a szintézis során. Az algoritmus a vizsgálatot külön lépésekben végzi el. Mindhárom esetben a mondat elejétől indulva, a fonemikus hangszimbólumokra támaszkodva megvizsgálja az adott hangkapcsolathoz tartozó jelölt hanghatár helyét és összehasonlítsa a fonetikailag optimálisnak tartott beszédhanggal (a szünet jelzés is hangsorépítő elem, tehát erre is ugyanolyan döntéseket hozunk, mint a beszédhangokra). Az F1 vizsgálatnál döntést hoz arról, hogy egyáltalán az a beszédhang van-e a hangsorban, mint ami a fonetikus átírásban szerepel. Az F2 vizsgálatnál az a kérdés, hogy a hanghatár pozíciója helyes-e. Ha nem tartja helyesnek, akkor döntést hoz arról is, hogy automatikusan javítható-e vagy nem, majd ha javítható, akkor kijavítja. Az F3 vizsgálatnál azt dönti el, hogy a beszédhang felhasználható-e a szintézis során önálló hangként. Az algoritmus a következő döntések egyikét hozza: (D1) a beszédhang és a jelölt hang megegyezik a fonemikus hanggal (jó jelölés); (D2) a beszédhang jelölése rossz, javítása manuális vizsgálatot kíván; (D3) a beszédhang jelölése jó, és a bal oldali hanghatár is jó helyen van; (D4) a beszédhang fonemikus jelölése jó, de a bal oldali hanghatár el van csúszva, javítása automatikusan megtörténhet; (D5) a beszédhang jelölése jó, de a bal oldali hanghatár el van csúszva, javítása csak manuálisan javasolt;

(D6) a beszédhang jelölése jó, de a hangot önálló elemként nem szabad használni (ezért ezt jelölni kell).

A hibajavítási folyamat várható eredménye az, hogy a pontosabb hanghatár eltüntetési a szintetizált hangsorokban korábban észlelt hangzási hibák nagy részét.

Az elemzéshez egyrészt az 1. táblázat paramétereit használjuk, másrésztől figyelembe vesszük a hangokra és a hangkapcsolódásokra jellemző szerkezeti jellemzőket is (a hang várható időtartama, a zárszakasz esetleges rövidülése, svá képződés lehetősége stb.). A vizsgálati paramétereket és a hozzájuk tartozó információkat a 2. táblázatban mutatjuk be. A javító algoritmus először a hangazonosítást (F1) végzi el. A döntéshez az 1. és a 2. táblázat paramétereit használjuk.

A vizsgálati lépések az (F1) típusú elemzéshez és javításhoz:

1. A hang hangsorkezdő (hangsor belseji) pozícióban van.
2. A hang fonemikus azonosítása és besorolása az 1. táblázat 11 osztályába, illetve a szünet osztályba.
3. Hosszú magánhangzó esetén a jelölt időtartam mérése.
4. Összevetés a jellemző időtartammal (2. táblázat 14. paraméter)
5. Ha az időtartam nincs a jellemző tartományban, a D2 hibajelzés elhelyezése manuális vizsgálathoz.
6. Egyezés esetén a további paraméterek vizsgálata (1. táblázat függőleges oszlop)
7. Eltérés esetén hibajelzés elhelyezése.
8. Egyezés esetén tovább lépés a következő hangra.
9. Ez szünet, vagy hangsorvégi szünet?
10. Ha nem a 2. ponttól új vizsgálat indul.

Vegyük példának az 1. ábrán látható hibás hangjelölés vizsgálatát. A hangjelölés szünetet (sil) állapít meg, ezért az algoritmusunk a 2. táblázat 11, 12, 13 paraméterét vizsgálja. Az időtartam 110 ms, ez megfelelő, beleesik a megadott sávba, akár levegővételi szünet is lehetne. A 12. paraméter vizsgálatánál kiderül, hogy ebben a szünetben az intenzitás szint sokkal magasabb, mint a megadott jellemző szint, ez azt sugallja, hogy a szünet jelölés mögötti akusztikai tartalom nem szünet, hanem valamilyen beszédhang. A 13. paraméter vizsgálata ezt még jobban alátámasztja, mivel a vizsgált szünetben találunk zöngésségi periódusjelzéseket is, ez is hiba. Ennek megfelelően a D2-es hibajelést kell elhelyezni. A 2. ábrán bemutatott szünet elem vizsgálata már a 11. paraméterrel való összehasonlításnál sejteti, hogy nem valódi szünetről van szó, mivel a jelölt határok közötti időtartam mindössze 33 ms. Ez alatta van a megadott minimális értéknek. A 3. ábrán látható szünet jelzés ugyanezzel a méréssel mondható hibásnak, de a 12. és 13. paraméter vizsgálati eredménye is hibára utal. Ebben az esetben tehát háromszoros megerősítéssel mondhatjuk ki, hogy a jelzés hibás.

2. Táblázat: a beszédhangok és a jelölt hanghatárok vizsgálatánál használt paraméterek és azok jellemző adatai

	PARAMÉTER	JELLEMZÉS
1.	zöngés a hang	a Praat program zöngeszinkron jeleket helyez el a hangsor zöngés részein
2.	zöngétlen a hang	a Praat program nem helyez el zöngeszinkron jeleket a zöngétlen részekben
3.	vegyes gerjesztés	a fonemikus hangjelből döntjük el. (z, zs, dz, dzs)
4.	nazális	a fonemikus hangjelből döntjük el. (m, n, ny)
5.	Orális	a fonemikus hangjelből döntjük el. minden hang, kivéve a 4. sort
6.	intenzív	a hangintenzitás mérésével döntjük el. Ez a maximális szint
7.	közepesen intenzív	a hangintenzitás mérésével döntjük el. Jellemző: 6-10 dB-lel alacsonyabb, mint a 6. sor
8.	gyenge	a hangintenzitás mérésével döntjük el. Jellemző: 12-20 dB-lel alacsonyabb, mint a 6. sor
9.	összetett szerkezetű	a fonemikus hangjelből döntjük el. (p, t, ty, k, c, cs)
10.	egyszerű szerkezetű	a fonemikus hangjelből döntjük el. Minden hang, kivéve a 9. sort
11.	Szünet paraméter: időtartam	rövid szünet, felesleges szünet: min. 40 ms, max. 60 ms levegővételi szünet: a rövid szünetnél hosszabb, max. 200 ms
12.	Szünet paraméter: intenzitás szint	a szünet elvárható intenzitás szintje a 6. sor maximális értékétől - 20 db-nyi a jellemző. Amennyiben nem így van, akkor a jelölt szünet intenzitás szerkezetét vizsgálni kell. Hibának számít, ha a szünetben erős intenzitás ingadozás van (glottalizációra utal), illetve ha magas az intenzitás.
13.	Szünet paraméter: zöngétlen gerjesztés	a szünetben a gerjesztés nem lehet zöngés
14.	jellemző hangidőtartam	hosszú mgh. esetén hangsor belsejében az elfogadható érték min. 70 ms, max 120 ms között van
15.	rövidre módosult zárszakasz a zárhangban	a hangkapcsolat: mb, md, nc, ncs, nd, ng, nk a zárhang jellemző hossza 20-30 ms.
16.	koartikulációs néma fázis	CC kapcsolatoknál, a hangkapcsolat: s,sz,c,cs + m,n,ny A hangkapcsolat első hangjának végén jön létre a koartikulációs néma fázis

A vizsgálat második szakaszában az (F2) javító algoritmus lép működésbe és a hangok bal oldali hanghatárának esetleges elcsúszását vizsgálja. Ezek a hibák többnyire automatikusan javíthatók is. Ennél a lépésnél feltételezzük, hogy az (F1) elemzéssel megtalált hibákat már kijavítottuk, minden hang jelölése megfelel a hangsor adott hangjának.

Az algoritmus vizsgálati lépései a következők:

1. A hang hangsorkezdő (hangsor belseji) pozícióban van.
2. A jelölt hang fonemikus azonosítása és besorolása az 1. táblázat 12 osztályába, illetve a szünet osztályba.
3. A jelölt hangot balról határoló szomszédos hang azonosítása és besorolása az 1. táblázat 12 osztályába (bal szomszéd). Ilyen elemnek tekintjük a szünetet is.
4. A jelölt hang bal oldali hanghatárának a vizsgálata és hiba esetén javítása
 - 4.1 Ha zöngés beszédhangról van szó, akkor összevetjük a hanghatár helyét a periódusjelzőkkel (2. táblázat 1. paraméter), valamint a bal szomszéd gerjesztési besorolásával.
 - 4.1.1 Ha a bal szomszéd zöngés hang, akkor a hanghatárt elfogadjuk jónak (D3), tovább lépünk jobbra a következő hangra. A vizsgálat újból kezdődik a 2. ponttól.
 - 4.1.2 Ha a bal szomszéd zöngétlen gerjesztésű és a vizsgált hanghatár zöngétlennek jelzett hangrészben van, akkor ez azt jelenti, hogy a zöngés beszédhang bal oldali jelzett hanghatára kicsúszott a hangból és tartalmaz egy részt az előző zöngétlen hangból (esetleg szünetből) is.

JAVÍTÁS- A hanghatárt jobbra mozgatjuk addig, amíg zöngés periódusjelzést nem találunk. A hibát ezzel kijavítottuk (D4), léphetünk a következő hangra.
 - 4.1.3 Ha a bal szomszéd zöngétlen gerjesztésű és a vizsgált hanghatár balról az első zöngésnek jelzett hangperióduson, vagy annak 10ms-os körzetében van, akkor a jelölés nem hibás (D3), léphetünk a következő hangra.
 - 4.1.4 Ha a bal szomszéd zöngétlen gerjesztésű és a vizsgált hanghatár balról nem az első zöngésnek jelzett hangperióduson van, hanem attól jobbra, akkor a hibás a jelölés, mivel a hanghatár nem a zöngés hang kezdetén van, hanem annak a belsejében.

JAVÍTÁS- A hanghatárt balra mozgatjuk az első zöngés periódus kezdetéig (D4). A hibát kijavítottuk, léphetünk a következő hangra.
 - 4.2 Ha zöngétlen beszédhangról van szó (2. táblázat 2. paraméter), akkor összevetjük a jelölt hanghatár helyét az időtengelyen a periódusjelzőkkel, valamint a bal szomszéd gerjesztési besorolásával.
 - 4.2.1 Ha a bal szomszéd zöngétlen beszédhang és a vizsgált hanghatár zöngétlen részen van, akkor tovább lépünk jobbra a következő hangra, a hanghatárt elfogadjuk jónak (D3).
 - 4.2.2 Ha bal szomszéd zöngés hang és a vizsgált hanghatár zöngés területre esik a hanghatár, akkor ez azt jelenti, hogy a zöngétlen beszédhang bal oldali jelzett hanghatára kicsúszott a hangból és tartalmaz egy részt az előző zöngés hangból.

JAVÍTÁS- A hanghatárt jobbra mozgatjuk, annyira, hogy a zöngés-zöngétlen váltóponttól balra 10ms-ra kerüljön (azért nem tesszük pontosan a váltópontra, mert a zöngésből zöngétlenbe való váltás során a gerjesztések mintegy 10ms-nyi részen átfedik egymást, tehát az előző zöngés hang egyre csökkenő nagyságú periódusaira szuperponálódik már az induló zöreje). Ezzel a hibát kijavítottuk (D4). Léphetünk a következő hangra.
 - 4.2.3 Ha bal szomszéd zöngés hang és a vizsgált hanghatár zöngétlen területre esik, de 10ms-nál nem messzebb a bal oldali utolsó zöngés periódus jelzőtől, akkor a hanghatár jó (D3), léphetünk a következő hangra.

4.2.4 Ha bal szomszéd zöngés hang és a vizsgált hanghatár zöngétlen területre esik, de 10ms-nál messzebb (jobbra) a bal oldali utolsó zöngés periódus jelzőtől, akkor a hanghatár jelölése hibás, az a zöngétlen hang belsejében van.

JAVÍTÁS- A hanghatárt balra kell mozgatni az bal oldali zöngés hang utolsó zöngés periódus jelzésére (D4). Ezzel helyére toltuk a hanghatárt, léphetünk a következő hangra.

A hanghatárt vizsgáló és hibajavító algoritmus esetenként az 1. táblázatban 11 osztályának más paramétereit is figyelembe veszi, azonban erre itt részleteiben nem tudunk kitérni. A fenti algoritmus működését egy példán is bemutatjuk. A 6. ábrán látható hangsor részben a réshang vizsgálatakor a következőket állapítja meg az algoritmus. A hang zöngétlen gerjesztésű (4.2), a bal szomszédja zöngés (3). A zöngétlen hang bal oldali hanghatára a 4.2.2 pont szerint zöngés területre esik, ez hibás, javítást kell végezni.

Az (F3) jelzésű vizsgálat során a módosult szerkezetű hangokat keressük és jelöljük meg, hogy a későbbi szintézis során a válogató modul ezeket ne használja fel önálló hangként. Ezzel a hibás hangzásokat kerüljük el. Ennél az elemzésnél feltételezzük, hogy az (F1) és (F2) elemzések hibáit már kijavítottuk. Az elemzéshez a 2. táblázat 15. és 16. sorát is felhasználjuk. Amennyiben a vizsgált hang és baloldali szomszédja a 15. sorban definiált hangkapcsolatokba tartozik, akkor a vizsgált hangra az „önállóan ne használj” jelölést tesszük. Amennyiben a vizsgált hang és baloldali szomszédja a 16. sorban definiált hangkapcsolatokba tartozik, akkor a bal szomszédra az „önállóan ne használj” jelölést tesszük.

5 Összefoglalás

A nagyméretű beszédatbázisok címkézése csak gépi megoldásokkal végezhető el, ami sok esetben téves. A bemutatott hibadetektáló és javító algoritmus fonetikai és akusztikai jellemzők ötvözéséből kialakított kritérium rendszert használ fel a hangsorba már gépi úton bejelölt hangok és hanghatárok ellenőrzésére és javítására. A manuális vizsgálati eredmények szerint a gépi hang- és hanghatár-jelölés 95%-ban ad korrekt eredményt. A meghallgatásos vizsgálatok szerint ez a magasnak tűnő szám nem elégséges, mert a szintézis során számos esetben a hibás hangazonosítás, illetve a hibás hanghatár miatt torzulások kerülnek a szintetizált hullámformába, ami rontja a hangminőséget. Reméljük, hogy az ismertetett algoritmussal gépi úton lehet majd az eddigi 95%-os hatásfokot egy-két százalékkal emelni. A kombinált módszert más, a jövőben készítenő ilyen beszédatbázisokra is lehet alkalmazni. További eredmény az is, hogy a helyes hanghatárok megnyitják az utat másfajta mérések (hangidőtartamok, időszerkezeti elemek, szóhosszúságok stb.) tömeges vizsgálatához is.

Bibliográfia

1. Boersma P., Weenink D.: Doing Phonetics by Computer. [Computer software], www.praat.org
2. Fék, M., Pesti, P., Németh, G., Zainkó, Cs., Olaszy, G.: Corpus-Based Unit Selection TTS for Hungarian. In: Sojka P., Kopcek I., Pala K. (eds.) Text, Speech and Dialogue TDS 2006, Springer, Brno (2006). 367–374
3. Mihajlik, P., Révész, T., Tatai, P.: Phonetic Transcription in Automatic Speech Recognition, *Acta Linguistica Hungarica*, Vol. 49 (3-4), (2002) 407–425
4. Olaszy Gábor: Hangidőtartamok és időszerkezeti elemek a magyar beszédben. *Nyelvtudományi Értekezések 155*. Akadémiai Kiadó. (2006)
5. Olaszy Gábor: Mássalhangzó-kapcsolódások a magyar beszédben. Tinta Könyvkiadó. (2007)

Ezt a kutatást az NKFP 2. programja (szerződés szám: 2/034/2004) támogatta.