

# Gépi beszéd természetességének növelése automatikus, beszédjel alapú hangsúlycímkező algoritmussal

Szaszák György<sup>1</sup>, Beke András<sup>2</sup>, Olasz Gábor<sup>1</sup>, Tóth Bálint Pál<sup>1</sup>

1 Budapesti Műszaki és Gazdaságtudományi Egyetem,  
Távközlési és Médiainformatikai Tanszék  
e-mail:{szaszak,olaszy,toth.b}@tmit.bme.hu  
2 MTA Nyelvtudományi Intézet, Fonetikai Osztály

**Kivonat** A minél természetesebb hangzás elérése a géppel előállított beszédben napjainkban is igen fontos kutatási terület. A hangzás természetességét számos más tényező mellett a prozódia is nagyban befolyásolja, ezért alapvető követelmény egy olyan, precízen annotált korpusz megléte, amely alapján gépi tanulással pontos generatív modelleket állíthatunk elő. A korpusz kézi címkézése költséges és hosszadalmas, még a prozódiai egységekre, hangsúlyokra vonatkozóan is, ráadásul nemzetközi tapasztalatok is igazolják, hogy a szakértő címkézők ítélete is szubjektív, hiszen a különböző szakértők által előállított hangsúlyozásra vonatkozó annotációk közötti átfedés ritkán haladja meg a 80%-ot. A fentiek miatt gyakran használnak automatikus címkéző eljárásokat. A hangsúlycímkezőt leggyakrabban a szöveges átírat alapján végzik el, ami azonban szerényebb pontosságot szolgáltat az emberi annotáláshoz képest. Alternatívaként jelen munkában egy beszédjel alapú hangsúlycímkező algoritmust valósítunk meg. Az így nyert hangsúlycímkezés ellenőrzésére hat (3-3 férfi és női) HMM-TTS rendszert tanítunk, majd szubjektív lehallgatási tesztekkel (CMOS) hasonlítjuk össze a rendszereket.

**Kulcsszavak:** gépi beszédfelismerés, nyelvi elemzés, információkinyerés

## 1. Bevezetés

A gépi beszédelőállítás célját szolgáló beszédkorpuszok tervezése, rögzítése, és különösen precíz címkézése fontos feladat, amely a szöveg-beszéd átalakítás (Text-to-Speech, TTS) minőségét is alapvetően meghatározza. A címkézést kézzel vagy automatikusan végezhetjük. A kézi címkézés általában pontos, de nagyon időigényes, és nem küszöbölhető ki maradéktalanul a szubjektivitás sem. Szakértő címkézők által készített prozódiai annotációban például 70 és 80% között találták az alapfrekvencia-változások jelölésének egyezőségét egy angol nyelvű korpusz ToBI szerinti annotációjában [1]. Saját tapasztalataink is azt támasztják alá, hogy a humán címkéző nem tud a jelentéstől elvonatkoztatni, és lehallgatás alapú címkézés során percepciójában nem tudja például elkülöníteni az akusztikailag (pl. alapfrekvencia-csúcs), illetve a nyelvileg (szintaxis és szemantika) jelölt hangsúlyokat, amelyek az emberben gyakran egységes hangsúlyérzetként jelentkeznek.

Emellett korábbi kísérleti eredmények is arra utalnak, hogy ha a hangsúly a szintaxisból következik, akkor annak az akusztikai megjelölése elmaradhat [2]. A korpuszok címkézésekor jó lenne, ha szelektíven, kizárólag az akusztikai evidencia alapján tudnánk megjelölni, hol található olyan marker, amely a hangsúlyozással kapcsolatba hozható.

A kézi hangsúlycímkézés alternatívája az automatikus módozat, amelyet tipikusan a beszéd szöveges átiratán végzett szövegelemzés alapján végeznek szabály alapon vagy esetleg adatvezérelten. Az automatikus eljárások sem mentesek azonban a hibáktól, ami ismét az akusztikailag és nyelvi jelölt hangsúlyok különbözőségéből, valamint az egyéni variabilitásból, vagy szövegen felüli kommunikációs szándékból fakad. A szabályalapú megközelítések egyelőre elterjedtebbek, pedig az általánosítóképességük korlátai miatt eleve nem hibátlan a szintaktikailag jelzett hangsúlyos pozíciók azonosítása sem. Ez utóbbi kivételkezeléssel javítható, de a szintaktikai és az akusztikai jelzések közötti különbségek ily módon nem kezelhetők.

Cikkünkben egy akusztikai elemzésen alapuló automatikus hangsúlycímkéző eljárást mutatunk be és értékelünk ki. Meglátásunk szerint a gépi szövegfelolvasáshoz az akusztikailag jelzett hangsúlyok jelölése a fontos a tanítókorpuszban, a szövegszinten kikövetkeztethető, de legalábbis percepciósan megjelenő „hangsúlyokat” a természetes beszédben sem jelezzük külön. A nemzetközi irodalomban számos hasonló kísérletről számoltak be [3], de ezek tipikusan a ToBI címkézés automatikus elkészítésére vonatkoztak [4]. Az eljárások közös pontja, hogy szegmentális, legfeljebb szótagszintű elemzésre támaszkodnak, de a szupraszegmentális vetületet korlátozottan képesek figyelembe venni. Bár a hangsúly valóban leginkább a szótaghoz köthető, véleményünk szerint hatékonyabb a szupraszegmentális oldalról, felülről lefelé haladva megközelíteni (vö. napjaink leginkább elfogadott beszédproduktions modelljével [5], amelyben a végső prozódiai struktúra felülről lefelé egyre finomodik a mélyebb szintek hozzáadódó befolyása révén).

A bemutatásra kerülő beszédjel alapú hangsúlycímkéző eljárás fonológiai frázisok automatikus felismerésén alapul [6], ennek háttéréről korábban az MSzNy konferenciákon is részletesen beszámoltunk [7]. Mivel a fonológiai frázis definíció szerint egyetlen hangsúlyos szótagot tartalmaz (magyarban ez az első szótagon kötött hangsúly miatt a fonológiai frázis legelső szótagja), az eljárással automatikus hangsúlycímkézés valósítható meg. A hangsúlycímkézés többszintűvé is tehető, mivel a detektálni kívánt fonológiai frázisok egyes típusai között is éppen a hangsúly jellege, erőssége az egyik elkülönítő kritérium (az intonációs kontúr mellett).

Cikkünk felépítése az alábbiak szerint alakul: elsőként bemutatjuk a szöveg, és a beszéd alapján végzett automatikus hangsúlycímkézési eljárásokat. A címkézés nélküli, valamint a két különféle eljárással címkézett korpuszokon egy-egy TTS rendszert tanítunk férfi és női hangra is, amelyeket szubjektív lehallgatási tesztekkel hasonlítunk össze.

## 2. Automatikus hangsúlycímkézés a szöveg alapján

A szövegből történő hangsúlycímkézés szabályalapon történik, amelyeket kivétel-listák egészítenek ki. A Profivox TTS rendszerben alkalmazott hangsúlycímkézés (és -generálás) teljes körű leírása a [8] irodalomban található, ehelyütt ennek egy rövid áttekintésére szorítkozunk. A szöveg alapú hangsúlycímkézés négy szintet különböztet meg:

- Nagyon erős hangsúly: általában valamilyen kontrasztivitásban, tagadásban jelenik meg, lista alapján határozzuk meg;
- Erős hangsúly: szintén szólista alapján határozza meg az algoritmus;
- Hangsúlyos: szövegszintű szabályok alapján adódik;
- Hangsúlytalan: a fennmaradó, vagy az irtó szabály miatt hangsúlytalanává vált szótagokon.

Ezen belül a szabályok elsősorban a hangsúlyos szótagok meghatározásában működnek közre. A főbb szabályok az alábbiak:

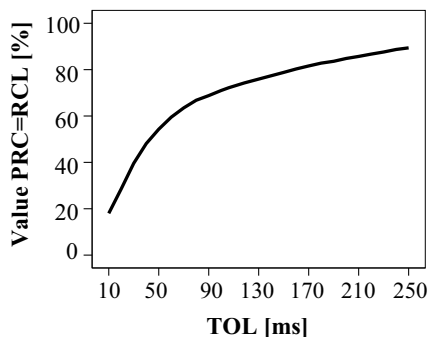
- A mondatkezdő szavak hangsúlyosak;
- Névelő és az *és* kötőszó után álló szavak hangsúlyosak;
- Vessző után hangsúlyos szó következik (figyelembe véve egy erre a célra kialakított kivétel-listát);
- A mondat utolsó szava sosem hangsúlyos;
- Névelők és erősen hangsúlyos szó után álló szavak sosem hangsúlyosak.

A Profivox TTS applikáció jelenleg használt változatában háromszintű hangsúlymodellezés van: erősen hangsúlyos (nagyon erős és erős hangsúly összevontan), hangsúlyos és hangsúlytalan szótagcímkéket használunk. Cikkünk hátralévő részében a szöveg alapú hangsúlycímkézésre angol elnevezése után a **TBSM** (Text Based Stress Modelling) rövidítéssel utalunk.

## 3. Automatikus hangsúlycímkézés a beszédjel alapján

Az automatikus hangsúlycímkézés fonológiai frázisok detektálásán alapul. A fonológiai frázisokat prozódiai jellemzők alapján Viterbi-algoritmussal illesztünk a beszédjelre. A fonológiai frázis [9] egyetlen hangsúlyos pozícióval rendelkezik, ez magyar nyelv esetén az első szótagon kötött hangsúlyozás miatt a frázis első szótagja. A szótagláncot és a szótagok kezdő- és végidőpontját ismerjük a korpuszból, így a fonológiai frázishatárok ismeretében már csak a hangsúlyos szótagok azonosítása van hátra közvetlenül a fonológiai frázishatár utáni szótagon.

A fonológiai frázisok detektálását végző algoritmust részletesen bemutatunk a [6] irodalomban, illetve korábban az MSzNy konferencián [7], így ehelyütt részleteiben nem ismertetjük, csak az algoritmusban a [6] forrásban dokumentálthoz képest végzett változtatásokat emeljük ki: az alapfrekvencia-követőt lecseréltük a Kaldi toolkit *compute-kaldi-pitch* eszközére, amely zöngétlen keretekre is szolgáltat értéket (a pontos algoritmust lásd: [10]). Ez az alapfrekvencia-követő nagyon



1. ábra. A fonológiai frázisszegmentáló pontossága (és hatékonysága) a  $TOL$  toleranciaérték függvényében,  $PRC = RCL$  munkapontokra.

kedvező viselkedésű, a Viterbi-algortmuson és néhány paraméterezzhető költségfüggvényen keresztül könnyen elérhető, hogy a szolgáltatott alapfrekvencia-kontúr oktávugrásoktól lényegében mentes, konzisztens, simított görbe legyen, amely további utófeldolgozást már nem igényel. Használatával jelentős pontosságnövekedést értünk el.

### 3.1. A fonológiai frázisszegmentáló kiértékelése

A [6] irodalomban megadott tanítókorpuszon (BABEL) és feltételekkel, de a Kaldi alapfrekvencia-követőjével kinyert jellemzőkön tanítottuk a fonológiai frázisszegmentáláshoz használt modelleket. A tanított HMM/GMM modelleket tízszeres keresztvalidációban ki is értékeltük, kézi fráziscímkézést használva referenciaként. Egy frázis detektálását akkor tekintettük helyesnek, ha a két frázishatár közötti eltérés egy toleranciaértéken ( $TOL$ ) belüli volt. A detektált frázishatárookra ezután hatékonyság (recall,  $RCL$ ), pontosság (precision,  $PRC$ ) és átlagos eltérés (average time deviation,  $ATD$ ) értékeket számítottunk. Az 1. ábrán látható a frázisszegmentáló frázishatár-detektálásra vonatkozó hatékonysága és pontossága  $TOL$  függvényében azokra a munkapontokra, ahol  $RCL = PRC$ . Ha  $TOL = 100ms$ , akkor ez a munkapont  $PRC = RCL = 71,0\%$ , ahol  $ATD = 31,9ms$ .  $TOL = 200ms$  toleranciaértékre  $PRC = RCL = 84,8\%$ ,  $ATD = 54,3ms$ .

### 3.2. Hangsúlyok szótagra illesztése

A beszédjel alapú hangsúlycímkézés is háromszintű, az egyes szinteket a fonológiai frázis típusa alapján különítjük el. Mivel a fonológiai frázisok típusainak elkülönítésében éppen a hangsúly erőssége az egyik alkalmazott kritérium, ez nem okoz különösebb nehézséget (lásd az 1. táblázatot). A fonológiai frázisok (FF) hangsúlyának erősségét az intonációs frázison (IF) belüli pozíció (IF kezdetre eső FF erősen hangsúlyos), illetve a szintaktikai, szemantikai és pragmatikai

viszonyok alakítják (pl. a mondathangsúlyt tartalmazó FF is erősen hangsúlyos lesz).

1. táblázat. A fonológiai frázisokhoz tartozó szótaghangsúly erőssége (az első szótagon)

FF típusa	Hangsúly	Jellemzés
me	erős	Intonációs frázis kezdete
fe	erős	Erősen hangsúlyos FF
fs	normál	Normál FF
mv	normál	IF végén ereszkedő kontúrú
fv	normál	IF végén emelkedő kontúrú
s	nincs	Hangsúlytalan(ná vált) FF
sil	nincs	Csend

Az így kapott hangsúlycímkézésre angol elnevezése után (Audio Based Stress Modelling) **ABSM** rövidítéssel hivatkozunk a továbbiakban.

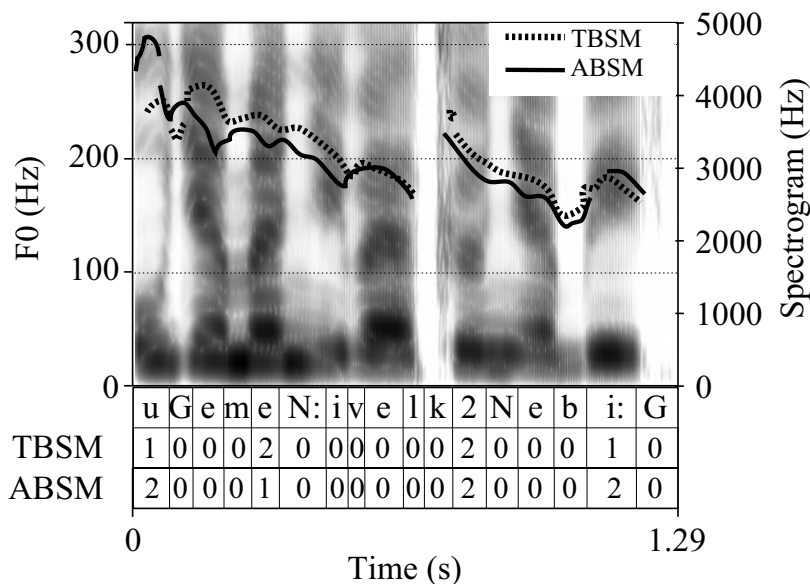
#### 4. A gépi szövegfelolvasó tanítókorpusza

A TTS betanításához használt beszédkorpusz a Magyar Párhuzamos Precíziós Beszédadatbázis, amely 1984 mondatot tartalmaz 14 beszélő felolvasásában [11]. A precíziós címkézés a fonetikai átiratra és a beszédhangszintű címkézésre utal, a kézi hangsúlycímkézés egyelőre még hiányzik az adatbázisból.

A korpuszt a bemutatott két eljárással (TBSM és ABSM) is felcímkéztük hangsúlyokra, majd a címkézést összevetettük hasonlóságuk tekintetében, illetve TTS rendszerekben is.

##### 4.1. A szöveg és a beszédjel alapú hangsúlycímkézés összevetése

A 2. ábrán látható egy rövid példamondatra vonatkozóan a kétféle eljárással generált hangsúlycímkesor. Általánosan elmondható, hogy mind a 14 beszélőt figyelembe véve, ABSM módszerrel az összes szó 48,4%-a, TBSM módszerrel pedig 33,1%-a kapott valamilyen hangsúlyt, tehát a beszédjel alapján másfélszer gyakrabban ítéltünk valamely szótagot hangsúlyosnak. A két módszer közötti fedést vizsgálva meglepő jelenséget tapasztaltunk (lásd 3. ábra): csak hangsúlyos és hangsúlytalan szótagokat megkülönböztetve a két eljárás legalább valamelyike által hangsúlyosnak címkézett szavakra a szavak kevesebb mint 1/3-át jelöli mindkét módszer egységesen hangsúlyosnak. Ennek a viszonylag gyenge átfedésnek a mélyebb vizsgálata kívül esik a cikk jelenlegi témáján, így csak annyit jegyzünk meg, hogy ebben egyrészt vélhetően a TBSM módszer heurisztikus jellege, általánosítóképességének korlátai játszatnak közre, másrészt befolyásolhatja az eredményt az is, hogy a szintaktikailag kikövetkeztethető hangsúly nem feltétlenül realizálódik akusztikailag is (vö. [12]), de ezt a jelenséget magyar nyelvre tudtunkkal még nem vizsgálták, jóllehet részben [2] eredményei is ebbe az irányba is engednek következtetni.

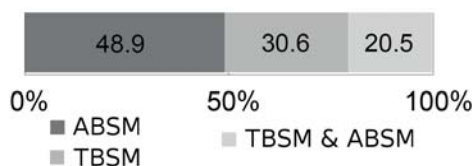


2. ábra. Az „Ugye, mennyivel könnyebb így.” mondat ABSM és TBSM címkéi. A beszédhangokat SAMPA kódjukkal adtuk meg, a szótagok hangsúly szerinti címkézésében 0=hangsúlytalan, 1=hangsúlyos, 2=erősen hangsúlyos.

#### 4.2. Kísérleti TTS mintarendszerek

A hullámforma alapján készített hangsúlymodell hatásait magyar nyelvű rejtett Markov-modell alapú szövegfelolvasó rendszerben (Hidden Markov Model based Text-to-Speech, HMM-TTS) [13] vizsgáltuk meg. A HMM-TTS tanítókorpuszaként a magyar nyelvű, párhuzamos, precíziós beszédadatbázis egy női és egy férfi beszédhangját használtuk. A tanító adatbázis mindkét beszélő esetén a teljes, 1984 mondatból álló halmazt tartalmazta. A mondatok 44 kHz-en, 16 biten lettek rögzítve. A döntési fák építéséhez az MDL (Minimum Description Length) kritériumot használtuk. Mind a női, mind pedig a férfi beszélő esetén három-három különböző szövegfelolvasó rendszert készítettünk el az alábbiak szerint:

- Az első rendszer döntési fája nem tartalmaztak hangsúllyal kapcsolatos jellemzőket, tehát a tanítás során explicit módon nem adtunk meg hangsúlyozásra vonatkozó információt. Ezt úgy értük el, hogy a tanítás során a döntési fák építéséhez szükség összes hangsúllyal kapcsolatos kérdést eltávolítottuk korábbi szövegfelolvasó rendszerünkől [13]. A továbbiakban erre a rendszerre **NOSM** rövidítéssel (NO Stress Model) hivatkozunk.
- A második rendszer minden hangsúllyal kapcsolatos kérdést tartalmazott, továbbá a tanító adatbázisban a hangsúlyos szótagokat szabály alapon becsültük. Ez a rendszer megegyezik a korábban bemutatott HMM-TTS rend-



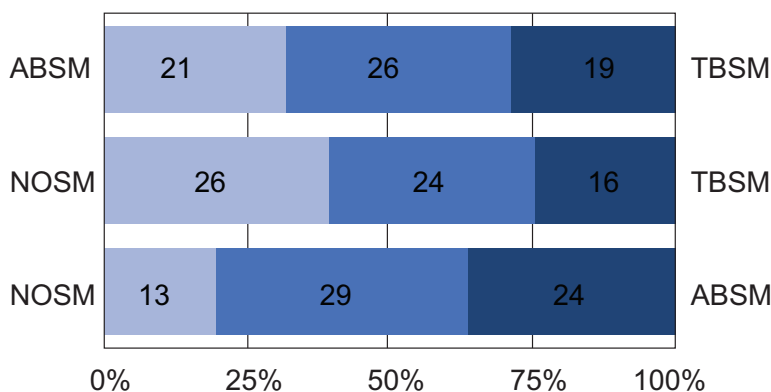
3. ábra. Az ABSM és TBSM hangsúlycímkézések hasonlósága (fedése).

szerünkkel [13]. A cikkben **TBSM** rövidítéssel hivatkozunk erre a megoldásra (Text Based Stress Model).

- A harmadik rendszer szintén minden hangsúllyal kapcsolatos kérdést tartalmazott. Ez esetben azonban a tanító adatbázisban a hangsúlyos szótagokat a jelen cikkben ismertetett módon, statisztikai módszerrel, pusztán a hullámforma alapján határoztuk meg. Szintézis során ez esetben is szabály alapon becsültük a hangsúlyokat. Továbbra is **ABSM** (Audio Based Stress Model) rövidítéssel jelöljük ezen rendszerünket.

## 5. Kiértékelés

A jelen cikkünkben bemutatott módszer érzeti hatásait szövegfelolvasó rendszerekben párösszehasonlításos meghallgatásos teszttel (Comparison Mean Opinion Score, CMOS) értékeltük ki. A teszt során egymástól függetlenül vizsgáltuk meg a férfi és női beszélőket. A meghallgatásos tesztben a korábban bemutatott három-három rendszer vett részt: NOSM, TBSM és az ABSM. A tesztalanyoknak az egyes rendszerek által generált mondatokat páronként kellett összehasonlítaniuk, aszerint, hogy mennyire találják természetesnek azok prozódiaját. Három lehetőség közül lehetett választani: (1) az első mondat természetesebb hangzású; (2) azonos a két mondat hangzása; (3) a második mondat természetesebb hangzású. Minden mondatpárban a két mondat két különböző rendszerrel lett elkészítve (NOSM vs. TBSM, NOSM vs. ABSM és TBSM vs. ABSM). Egy tesztalany összesen 18 mintapárt hasonlított össze. A mintapárok sorrendjét, és a mintán belül a rendszerek sorrendjét álvéletlen módon alakítottuk ki az esetleges memóriahatások elkerülése céljából. Összesen 21 alany (9 férfi, 12 nő) vett részt a meghallgatásos tesztben, akik összesen 378 mintapárt értékelték. Minden alany magyar anyanyelvű volt. A legfiatalabb tesztelő 22, a legidősebb 70 éves volt. A tesztalanyok átlagéletkora 34 év volt. A meghallgatásos tesztet az interneten keresztül lehetett kitölteni. A meghallgatásos teszt eredményeit a 4. és az 5. ábra mutatja be. Az eredményeket megvizsgálva a hangsúly-információt nem tartalmazó rendszer (NOSM) mindkét beszélő esetében jobban teljesített, mint a fonetikus átírat alapú hangsúlymodell (TBSM). Bár elsőre meglepő ez az eredmény, a 3. ábrán látottak fényében egybecseng korábbi megállapításainkkal, hogy a beszédkorpuszban ténylegesen megjelenő hangsúlyok és a szöveg alapján becsült hangsúlyok között kevés átfedés lehet. A beszédjel alapú hangsúlymodell (ABSM) férfi beszélő esetén több szavazatot kapott, mint a NOSM, valamint



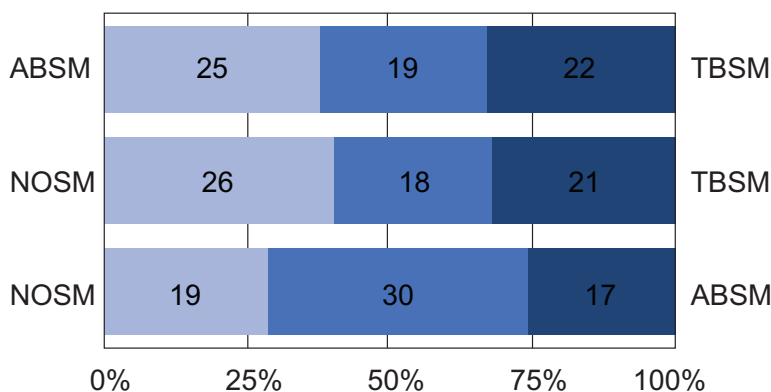
4. ábra. A meghallgatásos teszt eredményei férfi beszélő esetén.

mindkét beszélő esetén jobban teljesített, mint a TBSM. A szignifikanciát egy-mintás t-tesztel vizsgáltuk  $\alpha = 0,05$  mellett. Szignifikáns eltérést találtunk a férfi beszélő esetén a NOSM (hangsúly-információ nélküli) és az ABSM (beszédjel alapú hangsúlymodell) rendszer összehasonlítása során, az utóbbi javára. A női beszélőnél nem sikerült szignifikáns eltérést igazolni, de a szavazatok megoszlásából látható, hogy a hangsúlymodell nélküli rendszer és a beszédjel alapú hangsúlymodell szinte egyenlő szavazatokat kapott a két rendszer prózodiáját azonosnak értékelő hallgatók magas aránya mellett.

## 6. Összegzés

Cikkünkben automatikus hangsúlycímkézést, illetve hangsúlymodellezést vizsgáltunk a szöveg, valamint a beszédjel alapján magyar nyelvű HMM-TTS rendszerben. A két eljárást az explicit hangsúlyjelölés nélküli esettel és egymással is összehasonlítottuk, páronkénti szubjektív meghallgatásos teszttel. A hangsúlymodellezés hatása csak a tanult HMM-TTS modelleken keresztül érvényesülhet, szintézisidőben ugyanis mindig a szöveg alapján becsültük a hangsúlyokat. Az eredményekből fontos következtetéseket vonhatunk le: a korpuszon végzett, fonetikus átírat alapú hangsúlycímkézésnél előnyösebb, a meghallgatásos teszt alapján történő hangsúlymodellezés nélküli eset, hiszen jobb eredményt ad. A beszédjel alapú hangsúlycímkézés, illetve az ezen a címkézésen végzett modellezés a férfi beszélő esetén szignifikáns javulást eredményezett a beszéd természetességének szubjektív megítélésében, míg a női beszélőnél nem volt szignifikáns különbség a hangsúlymodellezés nélküli esethez képest ( $\alpha = 0,05$  mellett). Fontos megjegyezni, hogy a tesztalanyoknak kizárólag a prózodia természetességének megítélése volt a feladatuk, de eközben elkerülhetetlenül befolyásolta döntésüket az érzeti általános beszédminőség is. Az eredmények, beleértve a szöveg és a beszédjel alapján generált hangsúlyok közötti csekélynek mondható átlapolást





5. ábra. A meghallgatásos teszt eredményei női beszélő esetén.

is, felvetik annak a lehetőségét, hogy az emberi percepció a hangsúlyozásban nem a prozódia szintaxist megerősítő szerepét várja, hanem bizonyos túrérháttárral „megengedi” a hangsúlyos helyek váltakozását ugyanazon közlésben, és a hangsúlyra járulékos információforrásként tekint. Ezt a felvetést jelen munkában azonban nem vizsgáltuk, a jelentésbeli percepció eltérések és a hangsúlyozás kapcsolatáról tehát nem tudunk ennél biztosabb következtetést levonni a rendelkezésünkre álló adatokból. Eredményeink alapján fontosnak találjuk a téma további vizsgálatát, a beszédjel és a hangsúlyok kapcsolatának egzaktabb meghatározását, és a hullámformán alapuló, pontosabb hangsúlymodell gépi beszéd természetességére gyakorolt hatásának elemzését.

## Köszönetnyilvánítás

A szerzők köszönetüket fejezik ki Bartalis István Mátyásnak, a meghallgatásos teszt megtervezésében és kialakításában nyújtott segítségével; a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatalnak, amely a PD-112598 projekt keretében a kutatást támogatta; a Swiss National Science Foundationnak (Svájci Államszövetség), amely az „SP2: SCOPES project on speech prosody” (SNSF N<sup>o</sup> IZ73Z0-152495/1) számú projekt keretében a kutatásunkat támogatta.

## Hivatkozások

1. Pitrelli, J.F., Beckman, M.E., Hirschberg, J.: Evaluation of prosodic transcription labeling reliability in the ToBI framework. In: Proceedings of the 1994 International Conference on Spoken Language Processing. Volume 1. (1994) 123–126
2. Beke, A., Szaszák, Gy.: Combining NLP techniques and acoustic analysis for semantic focus detection in speech. In: Proceedings of the 5th IEEE International Conference on Cognitive Infocommunications. (2012) 493–497

3. Heggveit, P.O., Natvig, J.E.: Automatic prosody labelling of read Norwegian. In: Proceedings of Interspeech. (2004) 2741–2744
4. Wightman, C., Syrdal, A., Stemmer, G., Conkie, A., Beutnagel, M.: Perceptually based automatic prosody labeling and prosodically enriched unit selection improve concatenative speech synthesis. In: Proceedings of International Conference on Spoken Language Processing. Volume 2. (2000) 71–74
5. Levelt, W.J.M.: Speaking: From Intention to Articulation. MIT Press, Cambridge (1989)
6. Szaszák, Gy., Beke, A.: Exploiting prosody for syntactic analysis in automatic speech understanding. *Journal of Language Modelling* **0**(1) (2012) 143–172
7. Vicsi, K., Szaszák, Gy.: Folyamatos beszéd szó- és frázisszintű automatikus szegmentálása szupraszegmentális jegyek alapján: II. rész: Statisztikai eljárás, finn-magyar nyelvű összehasonlító vizsgálat. In: III. Magyar Számítógépes Nyelvészeti Konferencia. (2005) 360–370
8. Olaszy, G., Németh, G., Olaszi, P., Kiss, G., Zainkó, Cs., Gordos, G.: Profivox – a Hungarian TTS system for telecommunications applications. *International Journal of Speech Technology* **3-4** (2000) 201–215
9. Selkirk, E.: The syntax-phonology interface. In: *International Encyclopaedia of the Social and Behavioural Sciences*. Oxford: Pergamon (2001) 15407–15412
10. Ghahremani, P., BabaAli, B., Povey, D., Riedhammer, K., Trmal, J., Khudanpur, S.: A pitch extraction algorithm tuned for automatic speech recognition. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. (2014) 2494–2498
11. Olaszy, G.: Precíziós, párhuzamos magyar beszédatadabázis fejlesztése és szolgáltatásai. *Beszédkutatás* (2013) 261–270
12. Ananthakrishnan, S., Narayanan, S.: Automatic prosodic event detection using acoustic, lexical, and syntactic evidence. *IEEE Transactions on Audio Speech and Language Processing* **16**(1) (2008) 216–228
13. Tóth, B., Németh, G.: Improvements of Hungarian Hidden Markov Model-based Text-to-Speech Synthesis. *Acta Cybernetica* **19**(4) (2010) 715–31