A HEZITÁCIÓS JELENSÉGEK GÉPI OSZTÁLYOZÁSA A SPONTÁN BESZÉDBEN

BEKE ANDRÁS – HORVÁTH VIKTÓRIA

1. Bevezetés

A spontán beszéd számos megakadásjelenséget tartalmaz, amelyek közül az egyik leggyakoribb a hezitálás, vagy más néven kitöltött szünet (a köznyelvben *öözés*nek is nevezik hangzása miatt, mert leggyakrabban az /ö/-re emlékeztető semleges magánhangzóként realizálódik). A hezitálás a spontán beszéd természetes jelensége, számos funkciót tölt be a beszédben. Egyrészt időt biztosít a válogatáshoz a gondolat nyelvi formájának tervezése során, egyúttal reflektál a keresési folyamatra (Beattie–Butterworth 1979). A kitöltött szünet megjelenhet az egyes beszédtervezési szinteken fellépő hiba kísérőjelenségeként is, egyúttal időt biztosít a hibajavítási folyamatokhoz (Levelt 1989). A kitöltött szüneteknek a társalgásban is számos funkciójuk van (beszédszándék jelzése, beszélőváltás kivitelezése).

A beszélők között nagy különbségek vannak abban a tekintetben, hogy milyen gyakran produkálnak kitöltött szünetet (Goldman-Eisler 1968, Markó 2004, Fehringer–Fry 2007, Horváth 2010), de ez nem független a beszéd műfajától sem. A spontán narratívákban és a képleírásban (Markó 2004) nagyobb a hezitálás aránya, mint a dialógusokban (Horváth 2004). A feladat vagy a beszédműfaj mellett a közlés témája is hatással van a hezitálás gyakoriságára. Ha a beszélőnek olyan témában kell megnyilatkoznia, amelyet kevéssé ismer, növekszik a kitöltött szünetek aránya (Bortfeld et al. 2001, Merlo–Mansur 2004). Új információt vagy témát nagyobb arányban előz meg hezitálás a társalgás során, mint egy már ismert információt hordozó nyelvi egységet (Arnold et al. 2000). A közlés hossza is előjelezheti a hezitálások gyakoriságát: hosszabb közlések előtt gyakrabban jelennek meg (Shriberg 1996). Minél hosszabb és komplexebb ugyanis maga a megnyilatkozás, annál nagyobb az esélye valamilyen diszharmónia megjelenésének a beszédtervezés során. A hezitálás tehát azt jelzi a hallgató számára, hogy a következő közlés relatíve hosszú és komplex lesz (Watanabe et al. 2008).

A kitöltött szünetek elemzése nemcsak a pszicholingvisztika vagy a fonetika számára fontos; a beszédtechnológiai alkalmazások is fontos tényezőként kezelik. A hezitálások ugyanis sok esetben rontják a beszédfelismerés eredményét, azon belül a szóbeszúrások és -törlések, illetve a téves elutasítások számát növelik (Kaushik et al. 2010). A beszédfelismerők egy részében ezért megtalálhatók a hezitálásokra vonatkozó modellek (Ward 1991, Nakagawa–Kobayashi 1995, Kai–Nakagawa 1995). Az egyik HMM-es beszédfelismerőben például (Nakagawa–Kobayashi 1995) a gyakran előforduló kitöltött szüneteket hozzáadták a rendszer szókészletéhez, míg egy másik alkalmazásban (Kai–Nakagawa 1995) a kitöltött szüneteket mint a szótáron kívüli szót

vették figyelembe és a szóalatti dekóderrel azonosították ismeretlen szóként. Ezek a beszédfelismerő rendszerek azonban nem tudták a kitöltött szüneteken belüli variációkat megkülönböztetni, sem a funkciójukat azonosítani. Masataka és munkatársai (2000) olyan rendszert építettek, amely a hezitálásokat és a szóvégi nyújtásokat detektálja a japán spontán beszédben; az osztályozáshoz az alaphangmagasságot és a spektrális jellemzőket használták fel. Ezzel a metódussal a kitöltött szünetek és nyújtások 84,9%-át tudták helyesen osztályozni. Audhkhasi és munkatársai (2006) formánsalapú kitöltöttszünet-osztályozót hoztak létre. A hipotézisük az volt, hogy a hezitálások realizálódásakor a vokális csatorna relatíve állandó, így a kiejtett hang formánsmenete is közel állandó lesz, amit a formánsok szórásával jellemeznek. E mellett spektrális jellemzőket (MFCC: Mel Frequency Cepstral Coefficients) és az alaphangmagasságon alapuló jellemzőket is használtak. Az eredmények azt mutatták, hogy a formánsalapú osztályozó teljesített a legjobban. Wu és Yan (2002) huszonhat jellemzőt alkalmaztak a kitöltött szünetek osztályozásához. A jellemzőket Karhunen-Loéve transzformációval (KLT) és lineáris diszkriminancia analízissel (LDA) szűrték; osztályozó algoritmusként pedig kevert Gauss-modellt alkalmaztak. Az LDA-val szűrt jellemzőkkel 86,8%-os eredményt tudtak elérni, míg a KLT-vel 84,4%-osat. Magyar nyelvre célirányzottan még nem történt meg a hezitálások automatikus osztályozása spontán beszédben. Mivel a hezitálás multifunkcionális jelenség, ezért szükséges egy automatikus osztályozó kialakítása, amely képes a beszédben automatikusan bejelölni, hogy az adott hang hezitáció-e, és ha igen, akkor milyen típusú (svá vagy nazális hang, esetleg hangkapcsolat). A jelen kutatás fő kérdése az volt, hogy a hezitálások osztályozhatók-e automatikusan a spontán beszédben.

2. Anyag, módszer, kísérleti személyek

A kutatáshoz a BEA spontánbeszéd-adatbázisból (Gósy 2008) 10 interjút használtunk fel: a kísérletvezető az adatközlők munkájáról, családjáról, hobbijáról tesz fel kérdéseket (azzal az előzetes instrukcióval, hogy lehetőleg minél hosszabban beszéljen az adott témáról). A beszélők fele nő, fele férfi, mindannyian budapestiek, egynyelvűek, életkoruk 22 és 35 közötti. A korpusz összesen 57 perc időtartamú (adatközlőnként 3–8 perc), amelyet a Praat 5.1 programban (Boersma–Weenink 2009) annotáltunk.

A kitöltött szünetek és a többi beszédhang osztályozásához akusztikai jellemzőként az emberi hallást is modellező MFC együtthatókat (Mel Frequency Cepstral Coefficients) és azok első két deriváltját (+delták, delta-delták) használtuk, míg osztályozó algoritmusként rejtett Markov-modellt (HMM: hidden Markov-model) (Young 2005). A HMM építésénél 3 állapotú balról jobbra modelleket alkalmaztunk, illetve a modellkomplexitást legfeljebb 16 komponenst tartalmazó Gauss-keverék (GMM) sűrűségfüggvényig növeltük. A különböző hezitálásokat szupport vektor gépekkel (SVM: Support Vektor Machine) osztályoztuk, ahol az akusztikai jellemzőket az emberi hallásérzeten alapuló PLP együtthatók (Perceptual Linear Predition) és azok első két deriváltját (+delták, delta-delták) adták. Az SVM-hez radiális bázis (RBF – Radial Basis Function) kernel függvényt alkalmaztuk (OSU SVM függvénykészletet használtuk MATLAB-ban).

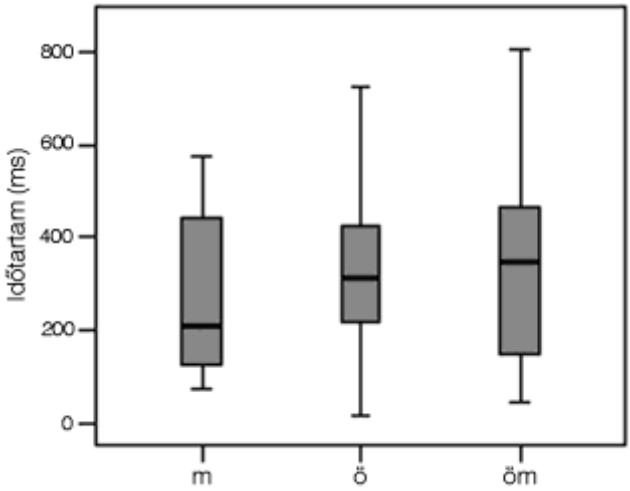
A kitöltött szüneteket az időtartamukkal is jellemeztük. A statisztikai elemzésekhez az SPSS 13.0 programot használtuk.

3. Eredmények

Az 57 perces korpuszban 1326 darab néma szünetet és 260 kitöltött szünetet adatoltunk. A néma szünetek percenkénti előfordulása átlagosan 23 darab volt, átlagos időtartamuk pedig 510 ms (23–3036 ms). A kitöltött szünetek átlagosan 4,5-szer fordultak elő percenként; legnagyobb arányban semleges magánhangzóként realizálódtak (84%, 219 darab). Az *öm* aránya már csak 10% körüli (33 db), nazálisként realizálódó kitöltött szünetre összesen 8 példát találtunk.

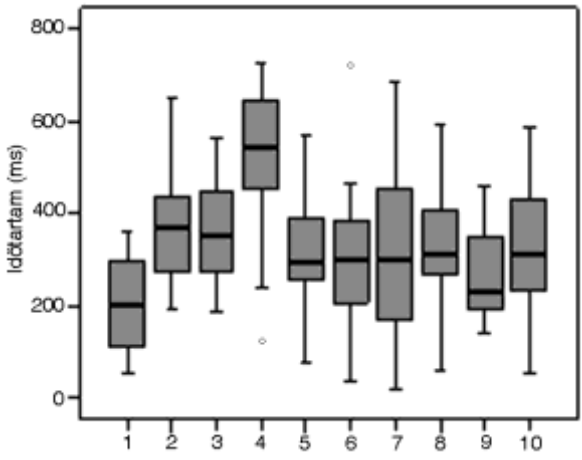
A hezitálás számos funkcióban jelent meg a korpuszban. Voltak olyan beszélők, akiknek a gondolat nyelvi formába öntése bizonyos esetekben gondot okozott, emiatt sokszor produkáltak kitöltött szünetet egy adott közlésrészen belül is, például: *öö □ közben öö elvégeztem az □ ELTE □ öö bölcsészkarán a magyar szakot illetve most öö fejezem be*. A hezitálás megjelenhet akkor, amikor a beszélő még abban sem biztos, hogy milyen gondolatot akar közölni (Levelt 1989); de már hezitál azért, hogy beszédszándékát jelezze. A közlés elején ejtett hezitálás tehát az esetek nagy részében arra szolgál, hogy a beszélő kiválogassa a közlésre szánt gondolato(ka)t; például: *…öö hát most egyszerre elég sok mindent csinálok* (egy kérdésre adott válasz indult ilyen formában). Előfordul az is, hogy a mentális lexikon pillanatnyi aktiválási nehézsége miatt, időnyerési célból hezitál a beszélő: *főleg így □ hát ilyen öö racionális megfontolásból*. A példában feltehetően a *racionális* szó előhívása okozott nehézséget, ezt a keresési időt jelzi a hezitálás mellett ejtett többi bizonytalansági megakadás is. A hezitálás tényleges kivitelezési hibák környezetében is előfordul, mintegy előre jelzi a téves kezdést: *beszédszintetizátorral □ öö □ me m létre akarunk hozni*.

Elemeztük a kitöltött szünetek időtartamát. A hezitálások átlagos időtartama 324 ms (átl. elt.: 162,47); a legrövidebb 21 ms-os, a leghosszabb pedig 804 ms-os időtartamban realizálódott. Az időtartamokban különbséget találtunk attól függően, hogy a kitöltött szünet milyen fonetikai formában realizálódott, ugyanakkor mindhárom típusnál óriási az adatok szóródása. A nazális hezitációk átlagosan a legrövidebbek (276 ms, átl. elt.: 195,58), ez feltehetően az artikulációs kivitelezés sajátosságaiból adódik. A svá-hezitálások átlagosan 50 ms-mal hosszabbak (323 ms, átl. elt.: 153,73). Átlagosan a leghosszabbak természetesen a hangkapcsolatként realizálódó kitöltött szünetek (338 ms, átl. elt.: 208,55). Noha a típustól függő időtartamokban tendenciaszerűen látszik a különbség, az elemszámok nagymértékű eltérése és az adatok nagy szóródása miatt a különbség statisztikailag nem szignifikáns (1. ábra).



1. ábra. A hezitálástípusok időtartama

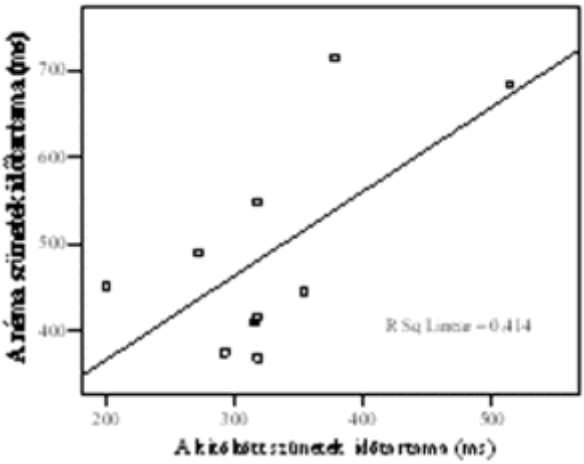
Elemeztük a beszélők közötti különbségeket. Az elemszámok nagy különbsége miatt az elemzéshez csak a svá-hezitálásokat használtuk, ezek átlagos időtartama 323 ms. A legrövidebb svá 20 ms-os időtartamban valósult meg, a leghosszabb pedig 720 ms volt. Az egytényezős varianciaanalízis különbséget mutatott a beszélők között a hezitálások időtartamában ($F(9, 218)=6{,}704$; $p<0{,}001$), de a post-hoc teszt szerint ez a különbség csak egy adatközlő (4. számú) és az összes többi beszélő között volt valóban szignifikáns.



2. ábra. A svá-hezitálások időtartama

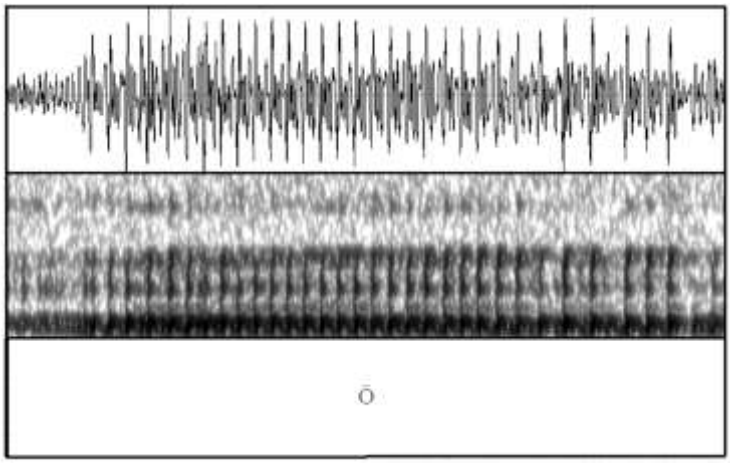
Elemeztük a korpuszban a néma és a kitöltött szünetek időtartamának összefüggését. A korrelációelemzés eredményei szerint a szünettartás egyéni jellegzetességeket mu-

tat; ha egy beszélő beszédére hosszabb néma szünetek jellemzők, akkor feltételezhetően a kitöltött szünetek is hosszabb időtartamban valósulnak meg (3. ábra).



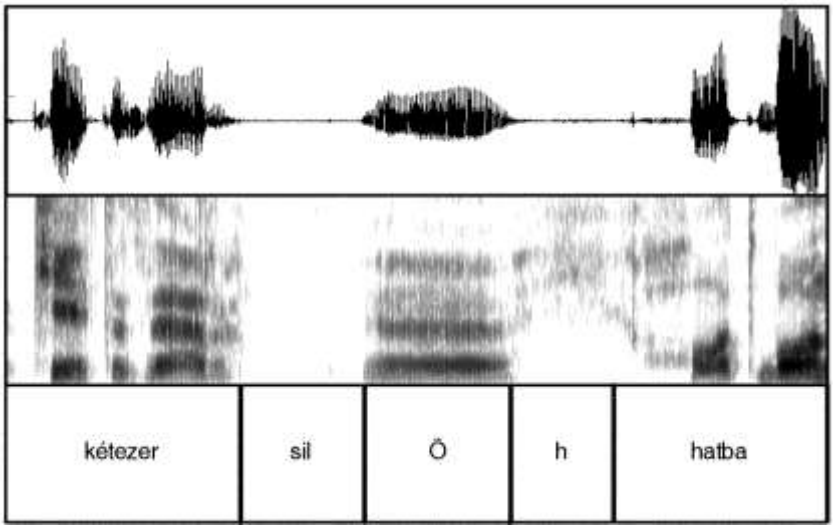
3. ábra. A néma és a kitöltött szünetek összefüggései

A zöngeképzés, a fonáció során a hangszalagok általában közel periodikusan (kváziperiodikusan) rezegnek. Ilyenkor a hangszalagok nagyjából állandó időközönként összecsapódnak – a rezgés reguláris. Rövidebb-hosszabb ideig azonban ez a rezgés lehet irreguláris. Ekkor a hangszalagok összecsapódásai között eltelt idő széles határok között ingadozik, és általában jóval hosszabb, egyes periódusok „kimaradoznak". Irregulárisnak tekinthető a rezgés akkor is, ha az alapfrekvencia hirtelen a beszélő hangterjedelme alá csökken, amelyet a hallgató érdes, rekedtes hangként azonosít (Markó 2005, Bőhm–Ujváry 2008). A jelen korpuszban a kitöltött szünetek 28%-a így valósult meg (4. ábra).



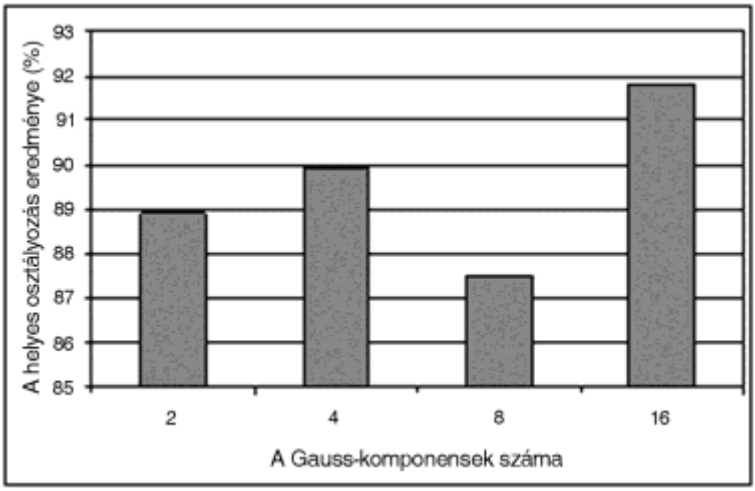
4. ábra. Irreguláris fonációjú svá-hezitálás rezgés- és hangszínképe

A beszélők a spontán beszéd során mintegy 10–15%-ban a szón belül tartanak szünetet, amely szintén lehet néma vagy jellel kitöltött (Gósy 2005, Horváth 2009, Bóna 2010). Az igekötőket vagy összetett szavak első tagját követő szón belüli szünet a lexikális előhívás problémájára utal, míg a toldalékot megelőző szünet a grammatikai tervezés zavarát jelzi. A jelen korpuszban is előfordult, hogy a beszélő egy szón belül tartott kitöltött szünetet (5. ábra), feltehetően azért, mert gondot okozott neki a pontos év felidézése.



5. ábra. A kitöltött szünetet tartalmazó kétezerhatba szó rezgés- és hangszínképe

A kitöltött szünetek és a többi beszédhang osztályozásához használt 3 állapotú HMM-eket használtunk. Az egyes állapotokban a kibocsátási eloszlásokat Gauss-függvények súlyozott összegével szokás leírni. 2, 4, 8 és 16 komponenssel vizsgáltuk a kitöltött szünetek osztályozási eredményességét. A legjobb eredmény a 16 Gauss-komponensű modell adta, amelynek eredménye 91,8% volt (6. ábra).



6. ábra. Az osztályozás eredménye a Gauss-komponensek számának függvényében

A 16 Gauss-komponensű modellel a hezitálások 98,33%-a osztályozható helyesen. Az algoritmus 1,67%-ban keveri össze a kitöltött szünetet valamilyen magánhangzóval, de mássalhangzóval nem.

	Magánhangzó	**Mássalhangzó**	**Hezitálás**
Magánhangzó	92,33%	0%	7,67%
Mássalhangzó	0,13%	89,31%	10,56%
Hezitálás	1,67%	0%	98,33%

1. táblázat. A HMM 16 gaussos modell osztályozási mátrixa

A hezitáláson belül az egyes típusokat SVM-mel modelleztük. Alapvetően két modellt építettünk: Ö, azaz svá-modellt; és az Öm, azaz hangkapcsolat-modellt. Az osztályozás eredménye 59,25%-os volt. Ebben a svá-hezitálások 64,28%-ban osztályozhatók helyesen, míg a nazális hangkapcsolatban lévő svát csupán 53,84%-ban kategorizálta helyesen az algoritmus.

	Ö	Öm
Ö	**64,28%**	**35,71%**
Öm	46,15%	53,84%

2. táblázat. Az SVM osztályozási eredménye

4. Következtetések

A hezitálás a spontán beszéd gyakori jelensége, amely számos funkciót tölt be a tervezési és önellenőrzési folyamatokban, de fontos szerepe van a társalgásban is a beszédszándék vagy a beszélőváltás jelzésére. A jelen kutatás kérdése az volt, hogy a hezitálások hogyan valósulnak meg a spontán beszédben és osztályozhatók-e automatikusan.

A jelen korpuszban a hezitálások többféle formában, számos funkcióban és nagyon változatos időtartamban realizálódtak. A kitöltött szünetek 28%-a irreguláris fonációval valósult meg.

A hezitálások MFC együtthatókkal előfeldogozva 3 állapotú HMM-ekkel 98,33%-os eredménnyel osztályozhatók a spontán beszédben. A hezitálástípusok azonban csak 59%-os eredménnyel osztályozhatók automatikusan SVM-mel.

A kutatás során kialakított gépi osztályozó eredménye kiválónak számít a nemzetközi irodalomban leírtak tükrében is.

Eredményeink felhasználhatók a beszédfelismerés eredményének javítására, illetve a beszédkutatás számos területén.

Irodalom

Arnold, J. E. – Wasow, T. – Losongco, A. – Ginstrom, R. 2000. Heaviness vs. Newness: the effects of structural complexity and discourse status on constituent ordering. *Language*. 76/1: 28–55.

Audhkhasi K. – Kandhway, K. – Deshmukh, O. D. – Verma, A. 2009. Formant-based technique for automatic filled pause detection in spontaneous spoken English. *IEEE International Conference on Acoustics, Speech and Signal Processing*. Taipei. 4857–4860.

Beattie, G. W. – Butterwoth, B. L. 1979. Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*. 22: 201–211.

Boersma, P. – Weenink, D. 2009. Praat: doing phonetics by computer (version 5.1). http://www.praat.org

Bóna Judit 2010. Bizonytalansági megakadások idősek és fiatalok spontán beszédében. *Beszédkutatás 2010*. 125–138.

Bortfeld, H. – Leon, S. D. – Bloom, J. E. – Schober, M. F. – Brennan, S. E. 2001. Disfluency Rates in Conversations: Effects of Age, Relationship, Topic, Role and Gender. *Language and Speech*. 44/2: 123–147.

Bőhm T. – Ujváry I. 2008. Az irreguláris fonáció mint egyéni hangjellemző a magyar beszédben. In Gósy Mária (ed.): *Beszédkutatás 2008.* Budapest: MTA Nyelvtudományi Intézet. 108–120.

Fehringer, C. – Fry, C. 2007. Hesitation phenomena in the language production of bilingual speakers. *Folia Linguistica*. *Acta Societatis Linguisticae Europae.* 41: 38–72.

Goldman-Eisler, F. 1968. *Psycholinguistics. Experiments in Spontaneous Speech*. London–New York: Academic Press.
Gósy Mária 2005. *Pszicholingvisztika.* Budapest: Osiris Kiadó.
Gósy Mária 2008. Magyar spontánbeszéd-adatbázis – BEA. *Beszédkutatás 2008*. 194–207.
Horváth Viktória 2004. Megakadásjelenségek a párbeszédekben. *Beszédkutatás 2004*. 187–199.
Horváth Viktória 2010. Funkció és kivitelezés a hezitációs jelenségekben. In Navracsics Judit (szerk.): *Nyelv, beszéd, írás. Pszicholingvisztikai tanulmányok I*. Budapest: Tinta Könyvkiadó. 65–73.
Kai, A. – Nakagawa, S. 1995. Investigation on unknown word processing and strategies for spontaneous speech understanding. In Pardo, J. M. (ed.): *Proceedings of Eurospeech '95*. Madrid: Universidad Politecnica. 2095–2098.
Kaushik, M. – Trinkle, M. – Hashemi-Sakhtsari, A. 2010. Automatic detection and removal of disfluencies from spontaneous speech. *Proceedings 13th Australasian International Conference on Speech Science and Technology Melbourne.* 98–101.
Levelt, W. J. M. 1989. *Speaking. From Intention to Articulation*. Cambridge: The MIT Press.
Markó Alexandra 2004. Megakadások vizsgálata különféle monologikus szövegekben. *Beszédkutatás 2004*. 209–222.
Markó Alexandra 2005. A spontán beszéd néhány szupraszegmentális jellegzetessége. Monologikus és dialogikus szövegek összevetése, valamint a hümmögés vizsgálata. PhD-disszertáció. Budapest: ELTE.
Masataka, G. – Katsunobu, I. – Satoru, H. 2000. A Real-Time System Detecting Filled Pauses for Spontaneous Speech. IEICE Transactions on Information and Systems, Pt. 2, Vol. J83-D-2; No. 11. Japan. 2330–2340.
Merlo, S. – Mansur L. L. 2004. Descriptive discourse: topic familiarity and disfluencies. *Journal of Communication Disorders*. 37: 489–503.
Nakagawa, S. – Kobayashi, S. 1995. Phenomena and acoustic variation on interjections, pauses and repairs in spontaneous speech (in Japanese). *Journal of the Acoustical Society of Japan.* 51/3: 202–210.
O'Shaughnessy, D. 1999. Detecting filled pauses in spontaneous speech. *Journal of the Acoustical Society of America*. 106/4: 2181–2182.
O'Shaughnessy, D. – Clark, Z. L. – Hesham, T. – Rachid, E. M. – Weiying, L. – Zhong-Hua, W. 1998. Detecting hesitations in the automatic recognition of spontaneous speech. *Journal of the Acoustical Society of America*. 104/3: 1805.
Shriberg, E. 1996. Disfluencies in Switchboard. *Proceedings, International Conference on Spoken Language Processing (1996)*. Philadelphia: Addendum. 11–14.
Ward, W. 1991. Understanding spontaneous speech: The Phoenix system. *International Conference on Acoustics, Speech and Signal Processing*. Toronto, Canada. 365–367.
Watanabe, M. – Hirose, K. – Den, Y. – Minematsu, N. 2008. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*. 50: 81–94.
Wu, C. H. – Yan, G. L. 2004. Acoustic Feature Analysis and Discriminative Modeling of Filled Pauses for Spontaneous Speech Recognition. *Journal of VLSI Signal Processing Systems*. 36: 91–104.