Speech Research conference

Hungarian Research Institute for Linguistics

Budapest, 14–15th December 2020

Beszédkutatás konferencia

Nyelvtudományi Intézet

Budapest, 2020. december 14–15.

Organisers/Szervezők:

Gocsál, Ákos
Gósy, Mária
Gráczi, Tekla Etelka
Gyarmathy, Dorottya
Horváth, Viktória
Huszár, Anna
Kohári, Anna
Krepsz, Valéria
Mády, Katalin

DOI: 10.18135/BeszKutKonf.2020

Reviewers/Bírálók

Abari, Kálmán
Audibert, Nicolas
Baditzné Pálvölgyi, Kata
Bakti, Mária
Bárkányi, Zsuzsanna
Bátyi, Szilvia
Bóna, Judit
Csapó, Tamás
Cunha, Conceição
Czap, László
Deme, Andrea
Fehér, Krisztina
G. Kiss, Zoltán
Gósy, Mária
Gosztolya, Gábor
Gráczi, Tekla Etelka
Gyarmathy, Dorottya
Gyuris, Beáta
Hoole, Phil
Horváth, Viktória
Huntley Bahr, Ruth
Imre, Angéla
Jordanidisz, Ágnes
Kallio, Heini
Kas, Bence
Kohári, Anna
Krepsz, Valéria
Kugler, Nóra
Lippus, Pärtel
Ma, Weiyi
Mády, Katalin
Markó, Alexandra
Miranda, Luma
Navracsics, Judit
Németh T., Enikő
Neuberger, Tilda
Reichel, Uwe D.
Schirm, Anita
Sipos, Zsóka
Siptár, Péter
Szekrényes, István
Tar, Éva
Tóth, László
Vainio, Martti
Vakula, Tímea
Wrench, Alan
Zainkó, Csaba
Zajdó, Krisztina

Contents

Tartalomjegyzék

Consistency in perceptual assimilation of foreign sounds as a correlate of language dominance in multilingual speakers

Naiemeh Afshar, Vincent J. van Heuven
University of Pannonia, Hungary

In this paper we examine the potential of the Perceptual Assimilation Method (PAM) as a means of assessing language dominance in bilingual (or multilingual) speakers. Consistency in perceptual labelling has been advanced as a hallmark of familiarity with a language (e.g., [2, 3]). We take this claim one step further and ask whether differences in language dominance in early bilingual speakers of Persian (an Indo-European language with three front and three back vowels: /i, e, æ; u, o, ɑ/, [4]) and Azeri (a Turkic language with three additional central vowels: /y, ɯ, œ/ [6]) are reflected in the consistency with which they assimilate the eleven pure vowels of American English (/iː, ɪ, eː, ɛ, æː, ʌ, ɑː, ɔː, oː, ʊ, uː/ [8]) to either the six vowels of Persian or the nine vowels of Azeri.

Twenty-three adolescent bilingual Azeri-Persian learners of English decided with which vowel in their native languages they identified each of the eleven American-English monophthongs, and how good an exemplar of the chosen category it was (cf. [1]). The vowels, spoken by two male speakers of American English, were presented twice so that the token-to-token consistency could be determined. Participants performed the task once for Azeri and once for Persian, always in that order. For each listener, a consistency index was computed, separately for responses in Azeri and in Persian mode, i.e., the number of pairs in which the first and second token received the same response, divided by the total number of vowel token pairs (= 22).

Consistency proved better in the Persian mode (mean = 78%, range 50 to 91%) than in the Azeri mode (mean = 58%, 27 to 86%), $t(22) = 6.1$, $p < .001$, probably due to the smaller number of response categories in Persian. The difference Δ in consistency between the two language modes varied strongly, between 19 and 50 points in favor of Persian. The poorer the consistency in the Azeri response mode, the larger Δ, $r = .748$, $p < .001$ (Fig. 1). We expect Δ to correlate with other measures of Persian-over-Azeri language dominance.

Participants also filled in the LEAP-Q (Language Experience and Proficiency Questionnaire), a comprehensive survey of (self-rated) familiarity with and proficiency in the languages they know [5]. On the basis of the LEAP-Q scores, the relative strength of the participant's native languages can be estimated. Although all participants considered Azeri their first and earliest spoken language (Table 1), they claimed to be highly fluent also in Persian in all pragmatic domains. They all stated to read better in Persian (the school language), while five students never learned to write in Azeri, and seven never became fluent Azeri readers (Table 1). Large individual differences could be observed in the self-ratings of the spoken-language skills, which tended to be negatively correlated between Azeri and Persian (Table 2, red-shaded cells), i.e., the better the self-rating in Azeri, the poorer it was in Persian (and *vice versa*).

We then computed LEAP-Q-based Persian-over-Azeri dominance measures by subtracting the self-rating for Azeri from the rating for Persian for Listening proficiency. This difference measure was used to predict the Δ in PAM consistency between Persian and Azeri, as defined

above. We computed the standardized residual of the Δ PAM consistency (in Fig. 1), i.e. the prediction error in Fig. 1. This relative dominance measure correlates with the dominance based on Listening proficiency at $r = .585$, $p = .002$ (see Fig 2). Regression analyses with other LEAP-Q based dominance ratios did not reveal any significant contributions.

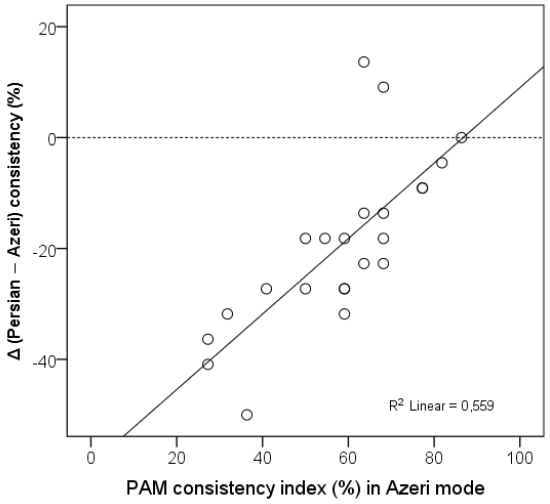


Fig. 1. Difference Δ PAM consistency between Persian and Azeri response mode, as a function of PAM consistency in Azeri mode.

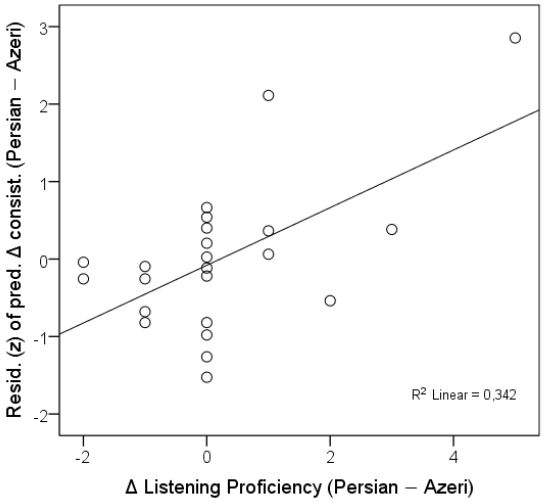


Fig. 2. Standardized residuals (z) of Δ PAM consistency (Persian – Azeri) as a function of Δ Listening proficiency (Persian – Azeri).

Although one of the questionnaire-based dominance measures correlates with the Δ PAM consistency measure we defined, our provisional conclusion is that perceptual labelling consistency is not a convincing measure of language dominance when the labelling task concerns the assimilation of foreign sounds to the native phonologies of bilingual listeners, at least not when the number of response categories differs between the native languages.

Table 1. Results for a selection of the 24 LEAP-Q parameters for 23 Azeri-Persian bilingual adolescents.

	Azeri (A)			Persian (P)			Difference (P – A)			
	Mean	SD	Range	Mean	SD	Range	Δ	t	df	p
Self-reported proficiency[a]										
Speaking	9.61	0.94	7..10	8.39	2.1	2..10	−1.22	−2.4	22	.024
Listening	9.26	1.25	5..10	9.48	0.9	7..10	0.22	.7	22	.504
Reading	6.43	2.97	0..10	9.57	1.1	5..10	3.14	4.9	22	< .001
Age milestones (years of age)										
Started learning	1.30	0.47	1..2	4.04	2.1	1..7	2.74	6.1	22	< .001
Attained fluency	4.57	2.59	2..12	7.83	2.9	4..15	3.26	3.8	22	.001
Started reading	7.67	2.09	5..13	6.65	0.9	5..8	−1.02	1.6	17	.123
Became fluent reader	10.38	2.36	7..16	9.17	1.3	7..12	−1.21	2.0	15	.066
Self-reported foreign accent[b]										
Perceived by self	2.78	2.11	0..6	4.22	2.2	0..9	1.44	2.0	22	.055
Identified by others	4.43	2.86	1..10	5.48	2.9	0..10	1.05	1.0	22	.230

[a] 0 ‘none’ to 10 ‘perfect’; [b] 0 ‘none’ to 10 ‘pervasive’.

Table 2. Correlation coefficients *r* of eight self-rated performance measures (scales from 0 to 10) for 23 bilingual Iranian participants with Azeri as L1 and Persian as L2. ** $p \leq .01$, 1-tailed); * $p \leq .05$ level (1-tailed).

	SpeakA	ListenA	SpeakP	ListenP	AccentA	IdentifA	AccentP
Proficiency Listening in Azeri	**.863****						
Proficiency Speaking in Persian	−.150	−.058					
Proficiency Listening in Persian	−.097	−.037	**.585****				
Non-native accent in Azeri	**−.388***	**−.408***	**.455***	.265			
Identified as non-native in Azeri	**−.509****	**−.440***	.268	.230	**.620****		
Non-native accent in Persian	**.392***	**.389***	−.335	−.276	−.233	−.210	
Identified as non-native in Persian	**.454***	**.502****	**−.430***	**−.356***	**−.487****	**−.513****	**.781****

References

[1] Best, C. T. (1995). 'A direct realist perspective on cross-language speech perception'. In: *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Ed. by Strange, W. Timonium, MD: York Press, pp. 167–200.

[2] Heuven, V. J. van & J. E. van Houten (1989). 'Vowel labelling consistency as a measure of familiarity with the phonetic code of a language or dialect'. In: *New methods in dialectology*. Ed. by Schouten, M. E. H. & P. Th. van Reenen. Dordrecht: Foris, pp. 121–130.

[3] Heuven, V. J. van (2017). 'Perception of English and Dutch checked vowels by early and late bilinguals. Towards a new measure of language dominance'. In: *Future research directions for Applied Linguistics* (Second Language Acquisition 109). Ed. by Pfenninger, S. E. & J. Navracsics. Bristol, Buffalo, Toronto: Multilingual Matters, pp. 73–98. DOI: 10.21832/9781783097135-006

[4] Majidi, M.-R. & E. Ternes (1999). 'Persian (Farsi)'. In: *Handbook of the International Phonetic Association*. Cambridge University Press, pp. 124–125.

[5] Marian, V., H. K. Blumenfeld & M. Kaushanskaya (2007). 'The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals'. *Journal of Speech, Language, and Hearing Research* 50, pp. 940–967. DOI: 10.1044/1qa092-4388(2007/067)

[6] Mokari, P. G. & S. Werner (2016). 'An acoustic description of spectral and temporal characteristics of Azerbaijani vowels'. *Poznan Studies in Contemporary Linguistics* 52, pp. 503–518. DOI: 10.1515/psicl-2016-0019

[7] Tsukada, K, D. Birdsong, D., E. Bialystok, M. Mack, H. Sung & J. Flege (2005). 'A developmental study of English vowel production and perception by native Korean adults and children'. *Journal of Phonetics* 33, pp. 263–290. DOI: 10.1016/j.wocn.2004.10.002

[8] Yavaş M. (2011). *Applied English Phonology*. Chichester: Wiley-Blackwell. DOI: 10.1002/978144439262

The prosodic correlates of stress in European and Argentinian 'Porteño' Spanish

Kata Baditzné Pálvölgyi
Eötvös Loránd University, Budapest

The objective of this research is to study and compare the characteristic prosodic correlates of stress in some main dialects of Spanish. The analyzed corpus come from the 'Map Tasks' of the *Interactive Atlas of Romance Intonation*, by Prieto et al. [8], as well as from spontaneous interviews uploaded to YouTube, and represent the northern and southern dialects of European Spanish (100 utterances, with 297 stressed syllables in the sample), as compared to the Argentinian 'Porteño' variety of Buenos Aires (100 utterances as well, with 279 stressed syllables in the sample). Porteño Spanish was selected in comparison with European variants as it is reported to be melodically different from Peninsular Spanish dialects as far as stress realization is concerned.

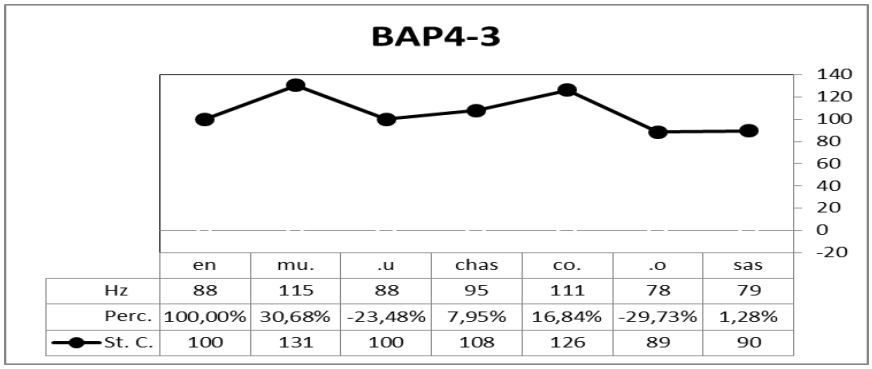
The three prosodic features that can play a prominent role in stress perception are tone, intensity and duration, but until today there has been no complete unanimity in the literature on whether the stressed Spanish syllable is pronounced in a higher tone, with longer duration or with greater intensity as compared to its adjacent context. According to Navarro Tomás [6], the stressed syllable is indicated by greater intensity, according to Llisterri et al. [5], by higher fundamental frequency (f0), and the latter complemented by a longer duration according to Ortega-Llebaria [7].

In this study I am focusing on the comparison of the intonational aspect: I will analyze my corpora to determine what relative tonal values characterize the stressed syllables as compared to the previous and the next syllables in the contrasted dialects. Regarding previous studies on the melodic characteristics of Buenos Aires Spanish, Sosa [9] mentions about the dialect the characteristic tautosyllabic falls from a high syllable, causing the effect of "vowel prolongation". Kaisse [4] also describes Porteño intonation with "long" descending inflections. In this study I try to discover the specific melodic cues that make the listener identify Porteño stressed syllables as high long falling, based on Cantero's Prosodic Speech Analysis (PAS) model [3], in which the fundamental frequency values in case of each syllable are first identified using an acoustic analysis program such as Praat [2] and then are standardized in order to obtain objectively comparable melodic patterns.

The first phase of the PAS analysis guarantees that we get rid of irrelevant micromelodic variations, by the reduction of each syllable to a characteristic tonal value. In case of tonal instability within syllables, the extreme values of f0 are taken. The standardized contour is represented by a line which starts with an arbitrary value of 100% and anchors in each syllable, which is itself characterized by a percentage based on its tonal position as compared to the previous syllable. If the syllable is located lower, it is a negative percentage, and if it is higher than the previous syllable, it is a positive one. Both curves (the absolute one and the standardized copy) are melodically identical, though in order to validate whether the standardized copy sounds the same as the original, it can be synthesized in Praat and submitted to a perceptive test. If correction is needed, it can be realized as a final phase. The standardized curve thus ensures that the described melodies are objectively comparable to each other, regardless of the individual

tonal characteristics of the speakers; what would matter are the proportions of the tonal movements.

According to my results, the main differences between European dialects and 'Porteño' Spanish lie in the fact that 'Porteño' Spanish is characterized by a greater proportion of inner inflections (i.e. tonal movements higher than 10%) within the stressed syllables than the European dialects: 33,33% of the stressed syllables in my Porteño corpus bear an inner inflection as compared to the European proportion (only 4,71%). Stressed syllables in Porteño Spanish are typically realized with a combination of a rise and a fall within the same syllable (cf. Figure 1; as shown in the standardized curve, there are two inner inflections, represented by dots within the syllable).



	en	mu.	.u	chas	co.	.o	sas
Hz	88	115	88	95	111	78	79
Perc.	100,00%	30,68%	-23,48%	7,95%	16,84%	-29,73%	1,28%
St. C.	100	131	100	108	126	89	90

Figure 1: Tautosyllabic high falls on stressed syllables in Porteño Spanish in the utterance *En muchas cosas* 'In many things'

The tonal changes to and from the stressed syllables in Porteño Spanish are accompanied by more intense melodic movements than in European Spanish. The average values of tonal rise to the stressed syllable are higher in the Porteño variant, 19,69% in comparison with the European mean value (6,97%). Stressed syllables in European Spanish are typically followed by a moderate rise (2,47% as a mean value), whereas in Porteño Spanish, they are rather accompanied by a fall (-7,13% as a mean value), cf. Figure 2.

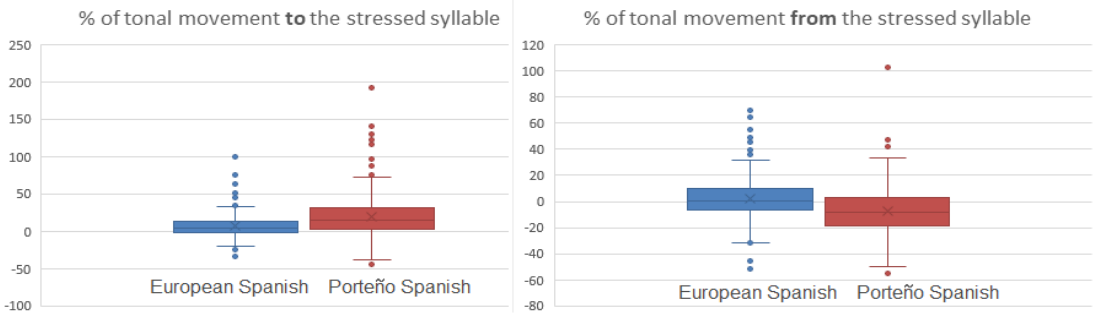


Figure 2: Boxplot diagrams representing the % of the tonal movement to and from the stressed syllables (European Spanish and Porteño variants), generated by Excel (version 2009)

The average proportion of the tonal rise and the fall (in percentages) within the stressed syllables, with the sensation of long tautosyllabic intense falls, shows a plausible Italian influence, also attested by historical reasons: the massive Italian immigration to Buenos Aires from the beginning of the 20th century on (cf. also Baditzné, [1]).

References

[1] Baditzné Pálvölgyi, K. (2020). El español porteño y el italiano meridional: Simetrías en la entonación prelingüística de las oraciones declarativas neutras. *Acta Hispanica (2020) Supplementum; América Latina y el mundo: espacios de encuentro y cooperación : II*. pp. 773-783.

[2] Boersma, P. & Weenink, D. (2020). *Praat: doing phonetics by computer* [Computer program]. Version 6.1.16, https://www.fon.hum.uva.nl/praat/

[3] Cantero Serena, F. J. (2019). Análisis prosódico del habla: más allá de la melodía, In: Álvarez Silva et al. (eds.): *Comunicación Social: Lingüística, Medios Masivos, Arte, Etnología, Folclor y otras ciencias afines. Volumen II.* Santiago de Cuba: Ediciones Centro de Lingüística Aplicada, 485–498.

[4] Kaisse, E. M. (2001). The long fall: An intonational melody of Argentinean Spanish. In: Herschensohn, J. et al. (eds.): *Features and Interfaces in Romance.* Amsterdam: Benjamins, pp. 148-160.

[5] Llisterri, J. et al. (2003). The perception of lexical stress in Spanish. Proceedings of the XV International Congress of Phonetic Sciences, ed. by Solé et al. Barcelona.

[6] Navarro Tomás, T. (1964). *La medida de la intensidad. Boletín del Instituto de Filología de la Universidad de Chile 16*, 231-235.

[7] Ortega-Llebaria, M. (2006). Phonetic Cues to Stress and Accent in Spanish. In: *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, ed. Manuel Díaz-Campos, Somerville, MA: Cascadilla Proceedings Project, 104-118.

[8] Prieto, P. et al. (coords.) (2010-2014). *Interactive Atlas of Romance Intonation.* http://prosodia.upf.edu/iari/

[9] Sosa, J. M. (1999). *La entonación del español. Su estructura fónica, variabilidad y dialectología.* Madrid: Cátedra

Supported by the ÚNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Az ujjbetűzés (daktil) komplex szerepe a siket gyermekek nyelvi fejlődésében és tanulástámogatásában

Bartha Csilla[1,2], Holecz Margit[1] és Tóth Etelka[3]
[1] NYTI Többnyelvűségi Kutatóközpont, Budapest
[2] ELTE BTK Mai Magyar Nyelvi Tanszék, Budapest
[2] KRE TFK Pedagógusképző Intézet, Budapest

A jelnyelvhasználók a (le)betűzést (magyarul gyakran daktil vagy ujjábécé) beszélt nyelvek írásbeli formáinak kódolására használják, melynek során különböző kézkonfigurációk egy-egy betűnek vagy szótagnak feleltethetők meg ([4, 9]). Esősorban olyan helyzetekben alkalmazzák őket, amikor az adott szóra vagy kifejezésre nincs jel, vagy a kommunikációs partner nem ismeri az adott jelet. Az ujjábécék a jelnyelvek alapvető részeként ugyanakkor bonyolult és szisztematikus módon fonódnak össze a nyelvelsajátítással is.

Az ujjbetűzés fontos szerepet játszik a fonológiai tudatosság kiépülésében, az olvasásban, információátadásban, a szülő-gyermek, különösen a halló szülő-siket gyermek közti kommunikációban, a jelnyelv vagy a jelelő számára idegen elemek feldolgozásában, nevek, új fogalmak kivitelezésében, de egyszersmind pragmatikai jelentések létrehozásában is ([1: 359, 3]). Siket családokban szülő-gyermek interakciókban a kisgyermekek a daktilt jóval korábban elkezdik használni, minthogy olvasni és írni tudnának (megjelenése kb. 13 hónapos korban, vö. [6]), és még az előtt, hogy tudatosulna számukra a (le)betűzés és a nyomtatás közötti összefüggés ([5: 46]).

A vizuális jelfonológiát, illetve annak olvasásban betöltött szerepét vizsgáló Petitto és munkatársai ([7]) kutatásai alátámasztották, hogy az olvasáselsajátítás szempontjából nem a hangok és a nyomtatott betűk közti kapocs a létfontosságú. Úgy vélik, hogy ezekben a folyamatokban az agy szegmentációs, kategorizációs és mintázatfelkismerő képessége meghatározó. Ezeket a folyamatokat nemcsak a hangzó nyelvek, de a jelnyelvek vizuális fonológiája is elő tudja segíteni, melybe beletartozik a (le)betűzés és szájkép, valamint azok mintázatai is, az agy ugyanis hangzó és vizuális nyelvi egységek (valamint azok fonológiai egységeinek és mintázatainak) befogadására egyaránt nyitott. A kutatások során olvasási nehézségekkel küzdő halló gyermekek is fejlődést mutattak vizuális jelfonológiára építő vizuális szegmentációs tréning eredményeként.

Az elmúlt évtizedekben vizsgálatok sora igazolta, hogy a jelnyelvhasználat folyékonysága pozitív összefüggést mutat a siketek hangzó nyelvi olvasási készségeivel, arra is rámutatva, hogy a jelnyelvi fluencia mellett a (le)betűzési készség magas szintje az, ami a siket kétnyelvűek olvasási képességeivel pozitív összefüggést mutat ([8]). A kutatók ezeket az eredményeket azokkal a közös mögöttes kognitív képességekkel magyarázták, melyek a szódekódolás pontosságáért és a szófelismerés automatizmusáért felelősek, melynek értelmében a (le)betűzés, a jelnyelv és ortografikai dekódolás közötti kapcsolatok erősítése az olvasáselsajátítás megkönnyítésének működő útja lehet.

A jelnyelvi hatásokon túl a betűzést külön faktorként vizsgálva a tanítás során Haptonstall-Nykaza és Schick ([2]) két különféle módszert hasonlítottak össze. Míg az egyik helyzetben az angol szó mellett annak amerikai jelnyelvi párja jelent meg, addig a másodikban az angol szó mellett nemcsak a jelnyelvi megfelelő szerepelt, de létrehoztak (le)betűzött formákat is. Utóbbi módszer

eredményeként a csak írást és jelet összekapcsoló helyzettel szemben 28%-kal javult a (le)betűzési képesség, 20%-kal a szó leírásának képessége és 10%-kal az írott szó felismerése, melyek alapján feltételezhető, hogy a (le)betűzés jelentheti a fonológiai kapcsolatot az íráshoz.

Több évtized olvasástanítási-módszertani kudarcai után a jelnyelvekre és a kétnyelvűségre nagyban építő fejlesztő programok gyakorlati tapasztalatai, valamint az újabb kutatási eredmények egyre nyilvánvalóbbá teszik, hogy a (le)betűzés jelentheti a fonológiai kapcsolatot az íráshoz, vagyis a daktilhasználat, a jelnyelv és ortográfiai dekódolás közötti kapcsolatok erősítése az olvasáselsajátítás megkönnyítésének működő és hatékony útja lehet. Egyre több ország siket gyermekek adott jelnyelveken történő (magas szintű kétnyelvűség elérését célzó) oktatási programjaiban a produkciós (fingerspelling) és a percepciós (fingerreading) (le)betűzési készségek fejlesztése az óvodától a 12. évfolyam végéig integráns része a tanterveknek.

Magyarországon jelenleg erőteljes a nemzetközi jelek hatása, melynek következtében több intézményben teljesen mellőzik daktilozás során a kétjegyű mássalhangzók használatát is (azok hosszú változatainak megjelenítésére pedig nincsenek általánosan elterjedt módszerek). Ennek eredményeként magyar szavak betűzésénél és szótagolásánál nehézségekbe ütköznek a siket és hallássérült tanulók akkor, amikor ilyen betűkapcsolatokkal találkoznak.

Az elmúlt évtizedekben az olvasással és tanulással kapcsolatos, külön-külön, a saját diszciplináris keretükben születő fontos eredmények a legutóbbi időkig kevéssé voltak átjárhatók egymás számára, s így sokszor részlegesen vagy szervetlenül integrálódhattak a siket gyermekek tanulását és fejlődését leginkább támogató gyakorlatokba. A vizuális tanulási segédeszközöket a siket tanulók egyéni igényeihez régóta megfelelőnek tartják, ma pedig a jobbára online is elérhető multimédiás tartalmakat hang-kép-szöveg, kép-hangzó videó-felirat lehetőségeivel, ritkább esetben jelnyelvi videóval, azokat pedig még ritkábban felirattal is ellátják ([10]). Noha a hangzó nyelv írott szöveggel vagy feliratokkal való megtámogatása növeli a tanulás eredményességét, ám ezek hatékonysága gyermekenként is eltérő lehet. Mind a kutatási, mind pedig a gyakorlati tapasztalatok azt mutatják, hogy a magyarországi siketek jelentős része korlátozott olvasási, szövegértési készségekkel rendelkezik, így számukra a hangzó nyelven folyó oktatás vagy a hangos tananyagok kizárólag feliratokkal való ellátása önmagában nem ad érdemi támogatást. A digitális technológiák széles körű elterjedése és a multimodális tanulási környezet szerepéről szóló vizsgálatok ugyancsak azt támasztják alá, miszerint a jelnyelvi videó és felirat együttese lenne az, ami nagyobb mennyiségű verbális információ átvitelét tenné lehetővé. Ennek hiányában a jó minőségű, megfelelő tipográfiájú és jól látható hangzó nyelv írott (feliratos) formájának relatív hatékonysága tehát továbbra is függ a siket tanuló (jel)nyelvi kompetenciájától, olvasási készségétől, bimodális kétnyelvűségének mértékétől. A megfelelő módszerekkel történő szilárd olvasási alapok megteremtése tehát a formális, informális környezetben történő digitális tanulásnak is megkerülhetetlen feltétele.

A NyelvEsély projekt szótári munkálatai során ezért – a jelnyelvi és hangzó nyelvi fejlesztést együttesen, egymást támogatandó kezelve – a jelkivitelezés összetevőinek megjelenítésén túl hangsúlyozottan vetődött fel a magyar ujjábécé (daktil), illetve az ujjbetűzés (valójában (le)betűzés) kérdése is. A fejlesztések során a munkacsoport többek között bővítette az általánosan használt daktilábécét a magyar ábécé hosszú kétjegyű mássalhangzóival, valamint a dz és dzs betűk daktil megfelelőjével. Ez a hiánypótló bővítés nagy segítséget jelent a magyar írás- és olvasáskészség tanítása, fejlesztése szempontjából, valamint célja, hogy megkönnyítse a

gyermekek számára a magyar nyelv szótagolásának tanulását. A kiegészítés alapvető célja, hogy minden a magyar nyelvben előforduló betű és betűösszetétel teljes körűen hozzáférhető legyen jelnyelven is a siketiskolákban és az integráltan tanulók számára. Az új daktil betűk használata segítséget jelent a magyar nyelvi fonológiai tudatosság megerősítésében is. Az előadás során a (le)betűzés integrálását célzó, a változatos tanulói csoportok igényeihez igazodó szótárfejlesztési munkálatokat ismertetjük.

Irodalom

[1] Bartha, Cs., M. Holecz & P. Z. Romanek (2016a). 'Bimodális kétnyelvűség, nyelvi-szociokulturális változatosság és hozzáférés: A JelEsély modell eredményei és távlatai'. *Általános Nyelvészeti Tanulmányok XXVIII. A többnyelvűség dimenziói: Terek, kontextusok, kutatási távlatok*. Szerk. Bartha, Cs. Budapest: Akadémiai Kiadó. 337–370. old.

[2] Haptonstall-Nykaza, T. S. & B. Schick (2007). The Transition From Fingerspelling to English Print: Facilitating English Decoding. Journal of Deaf Studies and Deaf Education 12.2. 172–183. old.

[3] Holecz, M. (2019). *Vizuális fonológia bimodális kétnyelvűségben – Nyelvi gyakorlatok, változatosság és az oktatásban való alkalmazás lehetőségei*. Doktori disszertáció.

[4] Johnston, T. & A. Schembri (2007). *Australian Sign Language (Auslan). An introduction to sign language linguistics*. New York: Cambridge University Press.

[5] Padden, C. A. & B. Le Master (1985). 'An Alphabet on Hand: The Acquisition of Fingerspelling in Deaf Children'. *Sign Language Studies* 47. 161–172. old.

[6] Padden, C. A. 1991. 'The acquisition of fingerspelling in deaf children'. *Theoretical Issues in Sign Language Research, Vol. 2. Psychology*. Szerk. Siple, P. & S. Fischer. Chicago: University of Chicago Press.

[7] Petitto, L-A., C. Langdon, A. Stone, D. Andriola, G. Kartheiser & C. Cochran (2016). Visual sign phonology: Insights into human reading and language from a natural soundless phonology. WIREs Cogn Sci 7. 366–381. old.

[8] Stone, A., G. Kartheiser, P. C. Hauser, L-A. Petitto & T. E. Allen (2015). 'Fingerspelling as a Novel Gateway into Reading Fluency in Deaf Bilinguals'. *PLoS ONE 10.10*. DOI: https://doi.org/10.1371/journal.pone.0139610.

[9] Twilhaar, J. N. & Bogaerde, B.v.d. (2016). *Concise Lexicon for Sign Linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

[10] Yoon, Joong-O., & H. Choi 2011. The effects of captions on deaf students' contents comprehension, cognitive load and motivation in online learning. *American Annals of the Deaf*. 156.3. 283–289.

L1 lexical attrition among post-puberty migrants

Szilvia Bátyi
University of Pannonia

Bilinguals differ from monolinguals in many ways. It has been documented in several studies on language processing that bilinguals cannot deactivate the language not being used [2] which can result in interference or cross-linguistic influence (CLI). CLI has been considered unidirectional for a long time in the bilingual literature, i.e. the language learner's dominant L1 influences the weaker L2. However, as [4] note, with the development of the L2, the first language of the speaker does not remain stable. This effect is called "backward" or "reverse" effect. One of the consequences of this bidirectional influence is that the L1 of bilingual speakers deviates from the monolinguals' [1]. The phenomena of the effects of the L2 on the L1 has been labelled as language attrition [3; 6 etc.] – the weakening of L1 skills. Language attrition can manifest itself at all linguistic level and can be influenced by a multitude of factors. The present study focuses on the influential extralinguistic factors of lexical attrition and speech fluency among Hungarian speakers living in the Netherlands. The target group is composed of 19 participants (M = 41, SD = 8.8) who have moved to the L2 environment at least 8 years prior to data collection. Their data has been compared to an age and education matched control group (N = 19) of Hungarian monolinguals living in Hungary. An extensive social personal questionnaire was used to explore the extralinguistic variables (frequency of use, length of residence, attitude) and elicited spontaneous speech task and semantic verbal fluency (VF) task were administered to measure lexical performance and speech fluency. Lexical attrition was operationalized as productive lexical fluency (VF task) and lexical diversity (D measure) while speech fluency was analysed through temporal measures of speech and the occurrences of disfluency markers (pauses, repetitions and self-corrections). The results show no dramatic decrease in the performance of the attrited group as compared to the control group and length of residence appeared to be the only factor explaining the outcomes. Interestingly, the self-reported L1 proficiency by the participants coincides with these results. These findings confirm what has been found by previous studies, that is, migration after the critical period (post-puberty) does not result in dramatic changes in L1 performance [7; 5].

References

[1] Cook, V. (2003). Introduction: The Changing L1 in the L2 User's Mind. In V. Cook (Ed.), *Effects of the Second Language on the First* (pp. 1–18). Clevedon: Multilingual Matters.

[2] Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? Journal of Memory and Language, 44, 2–31.

[3] Kȍpke, B. & Schmid. M.S. (2004). Language attrition: The next phase. In M.S. Schmid, B. Köpke, M. Keijer, and L. Weilemar (Eds.) *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues* (pp. 1-43). Amsterdam: John Benjamins.

[4] Kroll, J. F., Dussias, P.E., Bogulski, C.A., & Valdes-Kroff, J. (2012). Juggling two languages in one mind. What bilinguals tell us about language processing and its consequences for cognition. *Psychology of Learning and Motivation*, *56*, 229–254.

[5] Navracsics, J. (2015). L1 and dominant language - as reflected in speech disfluencies. In Navracsics, J., & Bátyi, Sz. (szerk.) *Első- és második nyelv: Interdiszciplináris megközelítések. First and second language: Interdisciplinary approaches.* Budapest: Tinta Könyvkiadó. (pp. 61 76).

[6] Schmid, M.S. (2013). First language attrition. *WIRE's Cognitive Science* 4(2), 117-12

[7] Schmid, M.S. (2019). The impact of frequency of use and length of residence in L1 attrition. In M.S. Schmid and B. Köpke (eds.). *The Oxford Handbook of Language Attrition* (pp. 288–297). Oxford University Press.

A természetes fonémaosztályok és a szentimentpolaritás sztochasztikus összefüggései a magyar szókészletben

Benczes Réka, Kovács Gábor
Budapesti Corvinus Egyetem

Több mint egy évszázaddal ezelőtt fektette le Ferdinand de Saussure [1] a nyelvi jel önkényességének doktrínáját – azt az elképzelést, miszerint a jelölő (a nyelvi jel hangalakja) és a jelölt (a nyelvi jel jelentése) közötti kapcsolat többnyire önkényes. Bár a nyelvi jel önkényessége a kortárs nyelvtudomány konstans elemét képezte (és képezi mind a mai napig), számtalan olyan jelenség található a nyelvben, amely nem önkényes, hanem motivált kapcsolatot feltételez a hangalak és a jelentés között [2]. Ezen jelenségek közé tartoznak az úgynevezett fonesztémák, azaz a nyelv olyan szubmorfémikus hangkombinációi, amelyek valamely expresszív jelentéssel asszociálhatók [3]. A magyar szakirodalomban a fonesztémák korpuszalapú vizsgálata viszonylag szegényesnek mondható. A jelenség elsősorban a hangutánzó vagy hang(ulat)festő szavak elemzése kapcsán merül fel [4, 5], és gyakran stilisztikai keretek között kerül tárgyalásra (ld. pl. [6]). Ahogyan arra Székely [3, 17. old.] is rámutat, széles körben elfogadott nézet, hogy a magyar nyelvben a fonesztémáknak „elsősorban a költői nyelvben van jelentősége” és jelentésük többnyire a szövegkörnyezettől függ [7].

Előadásunk merőben új nézőpontba kívánja helyezni a magyar fonesztémákat (és azok kutatását). Szakítva az eddigi stilisztikai, illetve erősen korlátozott korpuszú nyelvészeti kutatásokkal, vizsgálatunk tárgyát a *Magyar Szentimentszótár* ([8]) 5938 pozitív és 1748 negatív polaritású szóalakja képezte. Arra voltunk kíváncsiak, hogy létezik-e összefüggés a magyar szóalakok fonológiai összetétele és a szentiment között. A szavak hangtani összetételét tizenkilenc független változóval jellemeztük, melyek mindegyike egy-egy fonémaosztály relatív gyakorisága (a szóalakban szereplő fonémák, magánhangzók, illetve mássalhangzók számához viszonyítva).

A Bonferroni-korrekcióval elvégzett Mann–Whitney-féle U-próbák eredménye szerint a tizenkilenc hangtani jellemző közül tizenhárom szignifikánsan összefügg a szentimentpolaritással: a pozitív polaritású szavakban gyakoribbak az elől képzett magánhangzók, a zengőhangok és a huzamos mássalhangzók, míg a negatív polaritású szavakban magasabb a zárhangok és foghangok részaránya. A hangtani jellemzők együttes hatását egy bináris logisztikus regressziós modellel vizsgáltuk. A modell paramétereinek becsléséhez a mintában szereplő szóalakok véletlenszerűen kiválasztott 70%-át használtuk fel, és a modell prediktív erejét az adatsor maradék 30%-án teszteltük. Az ROC-görbe alatti terület értéke azt mutatja, hogy egy véletlenszerűen kiválasztott pozitív–negatív polaritású szópárról a modell – pusztán a hangtani jellemzők alapján – az esetek 63,3%-ában helyesen állapítja meg, melyik szó melyik szentimentkategóriába tartozik.

Kutatásunk tehát egyértelmű bizonyítékot szolgáltat a hangtani összetétel és a szentimentpolaritás közötti sztochasztikus összefüggésre, és a fonesztémák szövegkörnyezettől független, vélhetően motivált jelentésére.

Irodalom

[1] Saussure, F. de. (1915/1959). *Course in General Linguistics*. Szerk. Bally, E., A. Sechehaye & A. Riedlinger. Ford. W. Baskin. New York: McGraw Hill.

[2] Benczes, R. (2019). *Rhyme over Reason: Phonological Motivation in English*. Cambridge: Cambridge University Press.

[3] Székely, Zs. (2015.) ‘A motiváció kérdése a nyelvészetben’. *Motiváltság és nyelvi ikonicitás*. Szerk. Kádár E. & Szilágyi N.S. Kolozsvár: Erdélyi Múzeum Egyesület, 11–22. old.

[4] Benő, A. & Szilágyi, N. S. (2015). ‘Hangzásséma és motiváltság a hangutánzó és hangulatfestő igéink körében’. *Motiváltság és nyelvi ikonicitás*. Szerk. Kádár E. & Szilágyi N. S. Kolozsvár: Erdélyi Múzeum Egyesület, 43–57. old.

[5] Dimény, H. (2018). ‘Sound symbolism and meaning patterns: The case of Hungarian verbs’. *Roczniki Humanistyczne* 66.11, 47–57. old.

[6] Fónagy, I. (1961). ‘Communication in poetry’. *Word* 17, 194–218. old.

[7] Boda, I. K. & Porkoláb, J. (2013). ‘Hang- és színszimbolika a poétikai kommunikációban’. *Alkalmazott Nyelvészeti Közlemények* 8.2, 87–96. old.

[8] Szabó, M. K. (2015). ‘Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai és dilemmái’. *Nyelv, kultúra, társadalom* [Segédkönyvek a nyelvészet tanulmányozásához 177]. Szerk. Gecső T. & Sárdi Cs. Budapest: Tinta, 278–285. old.

Hibázások és hibajavítások jellemzőinek változása 4. és 5. osztályosok hangos olvasásában – egy longitudinális vizsgálat eredményei

Bóna Judit[1] és Steklács János[2]
[1] ELTE Eötvös Loránd Tudományegyetem
[2] Pécsi Tudományegyetem

A hangos olvasás vizsgálatával az olvasási képességet, a szövegértés képességének számos elemét tudjuk megvizsgálni, amelyek fontos szerepet játszanak az olvasási problémák diagnosztikájában, illetve a fejlesztésben [1]. Az olvasás fluenciájának mérése tulajdonképpen a dekódolási folyamat automatizálódottságáról ad képet. Ez azért fontos, mert a dekódolási folyamatok automatizáltságának megléte teszi lehetővé hogy az olvasó a mentális folyamatait a megértésre és az olvasás monitorozására tudja koncentrálni [3]. A hangos olvasás során a beszédbeli jellemzők közül a tempó, a szünetek és a tévesztések vizsgálata ad képet a fluenciáról. Emellett árulkodóak lehetnek a szemmozgások is, ezek elemzése azonban specifikus eszközt és módszereket igényel [4, 2].

A jelen vizsgálatban iskolás gyermekek hangos olvasásának változását elemezzük a 4. és az 5. osztály között. Ez a két osztályfok azért érdekes, mert a szakirodalom szerint 5. osztályra válik az olvasásértés olyan szintűvé, mint a hallás utáni szövegértés, illetve a szakirodalom szerint sok probléma mutatkozik az alsóról a felső tagozatra váltás miatt a gyermekek teljesítményeiben. Az sem elhanyagolható szempont, hogy míg a tízéves tanulók a PIRLS vizsgálatokon nemzetközi viszonylatban is jól szerepelnek, a felső tagozaton és a középiskolában már sokkal rosszabbak az eredményeik a TIMSS és PISA adatok alapján.

Saját korábbi kutatásunkban már vizsgáltuk ugyanezen gyermekek felvételein az olvasási időnek, a hibázások számának és javítási gyakoriságának, illetve a szemmozgások gyakoriságának és átlagos időtartamának az összefüggéseit. A jelen vizsgálatban arra keressük a választ, hogy milyen szemmozgások történnek az egyes hibajavítások és a hezitációs funkciójú megakadások szóbeli produkciója közben. Megvizsgáljuk, hogy változtak-e a gyermekek hibajavítási stratégiái, illetve az azokhoz kapcsolódó szemmozgások. Ez nóvum nemcsak a hazai, hanem a nemzetközi szakirodalomban is.

A vizsgálatban egy átlagos fővárosi általános iskola átlagos képességű tanulói vettek részt. A vizsgálat két alkalommal történt, egy év különbséggel: a gyermekek az első mérés időpontjában 4. osztályba, a második méréskor 5. osztályba jártak. Az előadásban 10 olyan gyermek (5 fiú és 5 lány) hangos olvasásának az eredményeit mutatjuk be, akiknek esetében a szemkamerás felvételeken az adatvesztés 10 % alatti volt.

A gyermekek feladata mindkét esetben az volt, hogy olvassák fel ugyanazon szöveget a monitorról. A felvételeket és elemzésüket mindkét esetben ugyanazzal a műszerrel (Tobii X120) és szoftverrel végeztük. A beszédelemzés és a szemmozgások kiértékelése a Praat és az ELAN szoftverrel történt.

Az előadásban bemutatjuk a hibázások és a megakadások típusait, a hibajavításokat, illetve a hozzájuk kapcsolódó szemmozgásokat. Összevetjük a két mérési időpontban kapott adatokat, és következtetéseket vonunk le az olvasás fejlődéséről, fejlesztési lehetőségeiről.

Irodalom

[1] Alt, S. J. & S. J. Samuels (2011). 'Reading fluency: What is it and how should it be measured?' *Handbook of reading disability research*. Szerk. McGill-Franzen, A. & R. L. Allington. New York, London: Routledge.

[2] Duchowski, A. T. (2007). *Eye tracking methodology*. London: Springer.

[3] Kamil, M. L., P. D. Pearson, E. B. Moje & P. P. Afflerbach (eds.) (2011). *Handbook of reading research, Volume IV*. New York: Routledge.

[4] Rayner, K., A. Kennedy & R. Radach (2004). *Eye movements and information processing during reading*. Hove, New York: Psychology Press.

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal K-120234 pályázata támogatta.

A fáradtság hatása zöngétlen explozívák zöngekezdési időire sclerosis multiplexben

Bóna Judit[1], Svindt Veronika[2] és Hoffmann Ildikó[2,3]
[1] ELTE Eötvös Loránd Tudományegyetem, Budapest
[2] Nyelvtudományi Intézet, Budapest
[3] SZTE Szegedi Tudományegyetem, Szeged

A screlosis multiplex (SM) a központi idegrendszer gyulladásos megbetegedése, egyike a leggyakoribb neurodegeneratív betegségeknek. A tünetei rendkívül változatosak. A jelen előadásban ezek közül a fáradtság és a beszédbeli tünetek egymásra hatását vizsgáljuk.

Az SM betegek közel 2/3-a számol be valamilyen nyelvi vagy beszédbeli tünetről, ezek között a leggyakoribbak a dysarthria, a szótalálási nehézségek, a romló verbális fluencia, a mondatismétlés nehézségei, a magasabb szintű nyelvi folyamatok korlátozódása [6], illetve a csökkent kommunikációs késztetés [3]. A betegek gyakran számolnak be arról, hogy beszédbeli nehézségeik a fáradtság hatására jelentősen megnövekednek [1]. Az SM betegek beszédbeli tünetei közül a szakirodalomban a leggyakrabban a tempó megváltozását vizsgálják, kevés adatunk van arról, hogy különbözik-e, illetve miben különbözik a betegek artikulációja a kontroll csoport artikulációjától. A témával kapcsolatos kevés kutatás szerint az SM betegek nagy százalékára jellemző dysarthria az artikuláció pontatlanná válásával, a beszéd érthetőségének csökkenésével, így a VOT megváltozásával is együtt járhat [4]. Ezt a változást fokozhatja a fáradtság, amely több beszédjellemző megváltozását okozhatja [5]. Kutatásunk kiindulásaként feltételeztük, hogy a mind a betegség állapota, mind a fáradtság befolyásolják az artikuláció folyamatát. A jelen előadásban akusztikai mérések segítségével elemezzük azt, hogy milyen hatással van a betegség és a fáradtság a zöngétlen explozívák kiejtésére.

A vizsgálatban 6 SM beteg és 6 életkorban és nemben illesztett kontroll beszélő vett részt. Mindegyikükkel kétszer készítettünk hangfelvételt: egyet a reggeli órákban, mielőtt munkába indultak volna, vagy elkezdték volna a napi teendők intézését, egyet pedig a késő délutáni/esti órákban, amikor a betegek jellemzően már a fáradtság tüneteit érezték. A jelen vizsgálatban elemzett hangfelvételek egy hosszabb, több feladatból álló protokoll utolsó feladataként lettek rögzítve. A feladat során az adatközlőknek azonos hordozómondatban szereplő álszavakat kellett felolvasniuk. Az álszavakban a *p, t, k* mássalhangzók szerepeltek VCV környezetben, az *í, á, ú* magánhangzók előtt. Minden hang 36-szor szerepelt a felvételen, beszélőnként összesen 108 explozívát elemeztünk.

A méréseket a Praat szoftverrel végeztük. Megmértük a VOT-k időtartamát, illetve a szótagidőtartamot is, amelyben az explozíva szerepelt (ezt a felpattanástól a követő magánhangzó végéig mértük, vö. [2]). Ezután kiszámítottuk a VOT időtartamának az arányát a szótagban, így kiküszöbölve az artikulációs tempó eltéréseiből fakadó beszélők közötti különbségeket. Végül összevetettük az adatokat a beszélőkön belül és a beszélők között is a reggeli és az esti felvételek függvényében.

Eredményeink szerint mind az SM, mind a fáradtság befolyásolják a VOT-t, azaz a zöngétlen explozívák artikulációját. Különbséget találtunk a betegek és a kontrollok között a betegség

előrehaladottságának függvényében mind a nyers időtartamokban, mind a VOT arányban; a különbség az előbbi érték esetén volt nagyobb.

Eredményeink rávilágítanak az SM betegek beszédbeli nehézségeinek egy újabb aspektusára, illetve új szempontokat nyújtanak a logopédiai terápiához.

Irodalom

[1] Blaney, B. E. & A. Lowe-Strong (2009). 'The impact of fatigue on communication in multiple sclerosis. The insider's perspective'. *Disability and Rehabilitation* 31.3, 170–180. old. DOI: https://doi.org/10.1080/09638280701869629

[2] Fischer, E. & A. M. Goberman (2010). 'Voice onset time in Parkinson disease'. *Journal of Communication Disorders* 43.1, 21–34. old. DOI: https://doi.org/10.1016/j.jcomdis.2009.07.004

[3] Gerald, F. J. F., B. E. Murdoch & H. J. Chenery (1987). 'Multiple sclerosis: Associated speech and language disorders'. *Australian Journal of Human Communication Disorders* 15.2, 15–35. old. DOI: https://doi.org/10.3109/asl2.1987.15.issue-2.02

[4] Kisomi, K. F., M. Soltani, M. Dastoorpoor, N. Madjdinasab & N. Moradi (2020). 'Comparison of Voice Onset Time in People with Spastic Dysarthria and Healthy Group'. *Shiraz E-Medical Journal* 21.5, e94573. DOI: 10.5812/semj.94573

[5] Krajewski, J., R. Wieland & A. Batliner (2008, July). 'An acoustic framework for detecting fatigue in speech based Human-Computer-Interaction'. *International Conference on Computers for Handicapped Persons*. Berlin, Heidelberg: Springer. 54–61. old. DOI: https://doi.org/10.1007/978-3-540-70540-6_7

[6] Laakso, K., K. Brunnegård, L. Hartelius & E. Ahlsén (2000). 'Assessing high-level language in individuals with multiple sclerosis: a pilot study'. *Clinical Linguistics and Phonetics* 14.5, 329–349. old. DOI: https://doi.org/10.1080/02699200050051065

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal K-132460 pályázata támogatta.

Stress and phrasal prosody in Udmurt: initial results

Lena Borise and Ekaterina Georgieva
Research Institute for Linguistics, Budapest

In this paper, we investigate the prosodic realization of stress in Udmurt (Uralic, Permic), in the contexts of (i) indicative verbs (PRS.3SG) and (ii) imperative verbs (IMP.2SG/PL). We also show that these results align with our earlier results for (iii) negated indicative verbs (PRS/PST.2/3PL). The accepted view is that stress in Udmurt is fixed on the final syllable (see [3]; [2]), which applies to contexts (i); in contexts (ii) and (iii) stress is, exceptionally, on the initial syllable. Based on novel experimental evidence, we show that (a) the main acoustic cue that the realization of stress in Udmurt relies on is vowel duration, and (b) instrumental evidence supports the traditional descriptions of the stress properties of contexts (i), (ii), and (iii). Additional support for the current analysis comes from the interaction of stress and the spread of high tone associated with the interrogative particle *a* (pre-theoretically, we will refer to it as H).

Background. According to the only instrumental investigation of Udmurt stress [1], di- and trisyllables with final stress are marked by greater duration of the final vowel and lower f_0 values, as compared to the vowel in the penultimate syllable. Notably, the test words in [1] were uttered in isolation, which alone may explain the obtained results. In minimal pairs formed by verbs of (i) and (ii) types, the stressed vowels (either initial or final) had greater duration than their unstressed counterparts within the minimal pair; the f_0 results were not consistent. No statistical analysis was offered.

Materials and methods. The current experiment targeted minimal pairs formed by verbs of (i) and (ii) types (di- and trisyllables). The items were controlled for syllable shape (CV), vowel height (low vs. mid/high), and information structure (backgrounded vs. focused). All items were embedded in carrier phrases. Six native speakers of Udmurt (5F, 1M, age range 20-40) participated in the experiment, which took place in June 2020. The recordings were made in a quiet room with a head-worn microphone. The results reported here are based on the data from Speaker 1 (F); other speakers' data are currently being processed.

Results. Our results show that, within verbs of both (i) and (ii) types, the final vowel has greater duration than the preceding one(s). The difference between the verbs of (i) and (ii) types lies in the comparison of vowel duration across verb types. Stressed initial vowels (i.e., those found in imperatives) are significantly longer than their unstressed counterparts (i.e., those found in indicatives). Similarly, stressed final vowels in indicatives are longer than their unstressed counterparts in imperatives, though the significance results are less consistent. cf. Table 1. Note that an unpaired t-test was used on the initial results due to small sample size; on the full data set, a linear mixed effects model is going to be used.

With respect to f_0 in verbs of (i) and (ii) type, there is preliminary evidence that verbs of (i) type carry a flat f_0 contour, which drops sharply on the final (stressed) syllable, while initial stress in verbs of (ii) type is cued by a high f_0 target, followed by a more gradual drop in f_0 throughout the rest of the word.

		Focused		Backgrounded	
		Indicatives (i)	**Imperatives (ii)**	**Indicatives (i)**	**Imperatives (ii)**
Disyllabic	initial	90.84 (n=16)	150.21*** (n=16)	94.66 (n=15)	150.69*** (n=13)
	final	189.44	174.3	207.96	173.53**
Trisyllabic	initial	74.79 (n=14)	125.03*** (n=10)	73.7 (n=15)	129.54*** (n=12)
	final	178.04	158.72	173.51	153.43

Table 1: Average durations of vowels in initial and final syllables (ms); asterisks mark the values in the imperative verbs that are significantly different from their counterparts in the indicative verbs, based on an unpaired t-test.

References

[1] Denisov, V. N. (1980). 'Foneticheskaja xarakteristika udarenija v sovremennom udmurtskom jazyke [A phonetic characteristic of stress in Udmurt]'. PhD thesis. Leningrad University.

[2] GSUJa I (1962). *Grammatika sovremennogo udmurtskogo yazyka. I. Fonetika i morfologija*. Ed. by Perevoshchikov Petr, I. Izhevsk: Udmurtia.

[3] Yemelyanov, A. I. (1927). *Grammatika votyackogo jazyka*. Leningrad: Izdanie Leningradskogo vostochnogo instituta im. A. S. Enukidze.

Artikuláció-beszéd szintézis MRI alapon

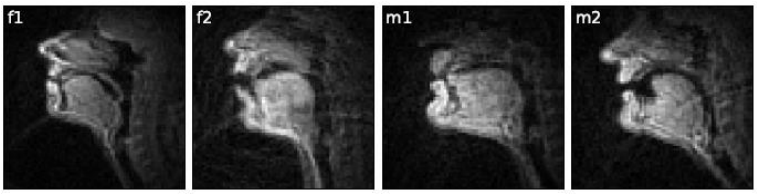
Csapó Tamás Gábor [1,2]
[1] Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
[2] MTA-ELTE „Lendület” Lingvális Artikuláció Kutatócsoport

Az artikuláció alapú beszédszintézishez az elmúlt években többféle artikulációs mozgást rögzítő technológiát alkalmaztak: 2D nyelvultrahang [3,4], elektromágneses artikulográf [1], permanens mágneses artikulográf [6], felszíni elektromiográfia (sEMG) [7], ajakvideó [9], valamint multimodális megoldások [5]. A szakirodalmi áttekintés során [2] nem találtunk olyan módszert, amely valós idejű mágneses rezonanciás képalkotást (real-time Magnetic Resonance Imaging, rtMRI) bemenetként használva a legmodernebb mély neuronhálós megoldásokat (deep neural network, DNN) alkalmazná; így jelen kutatási témánk erre koncentrál.

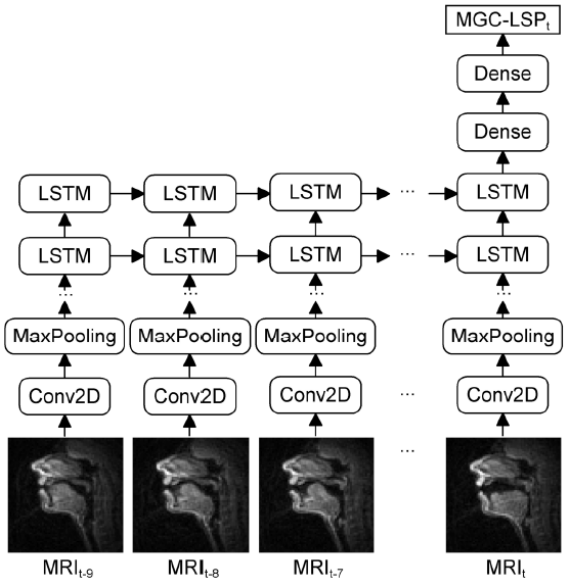
A kutatáshoz az USC-TIMIT MRI adatbázis [8] két férfi (’m1’ és ’m2’) valamint két női (’f1’ és ’f2’) beszélőjét alkalmaztuk. Ez az adathalmaz szinkronizált beszéd és rtMRI adatot tartalmaz amerikai angol beszélőktől. Az rtMRI képeken a teljes artikulációs csatorna látható (ld. 1. ábra), 68x68 pixel térbeli és 23.18 kép/másodperc időbeli felbontással. Az MRI előnye a hasonló képalkotó artikulációs technológiákhoz (pl. nyelvultrahang, ajakvideó) képest, hogy a teljes vokális traktus (lingvális, labiális, és állkapocs mozgás, valamint lágy szájpad és faringális terület) látható. Az MRI hátránya viszont, hogy az időbeli felbontás viszonylag alacsony, és így a beszéd gyorsan változó részei (pl. zárfelpattanások) nem követhetőek; és az is akadályt jelenthet, hogy felvétel közben jelentős a háttérzaj.

A kutatási projekt során az rtMRI bemenetből szintetizáltunk beszédet, a korábban ultrahangos kísérletekhez használt vokóder [3] segítségével. Az MRI-ből származó nyers adatokat közvetlenül képként kezeltük további előfeldolgozás nélkül. A kísérletek során mély neurális hálón alapuló gépi tanulást alkalmaztunk (előrecsatolt, konvolúciós és rekurrens hálózatokkal, ld. 2. ábra) az artikuláció-beszéd szintézishez, melynek bemenete a szürkeárnyalatos MRI kép volt, kimenete pedig a beszéd spektrális paraméterei, ún. „mel-általánosított kepsztrum” reprezentációban. A kutatás eredményeként szintetizált mintákat objektív mérőszámokkal (mel-kepsztrális távolság) és meghallgatásos tesztekkel is vizsgáltuk, melyek szerint a kísérleti alanyok érthetőnek tartották az MRI alapú akusztikum-artikuláció becslés során generált beszédet. A MUSHRA jellegű teszt eredményeit a 3. ábra mutatja. A szintetizált minták meghallgathatóak a http://smartlab.tmit.bme.hu/interspeech2020_mri2speech honlapon.

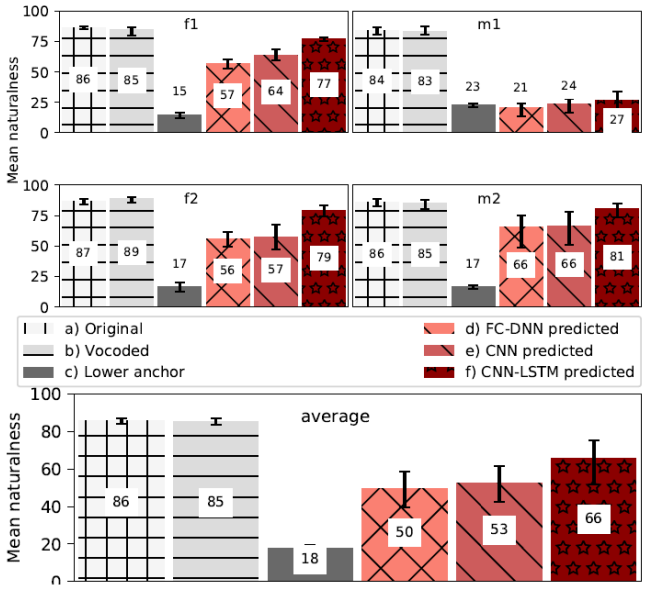
Az MRI alapú artikuláció-beszéd szintézis eredményeink hozzájárulhatnak nyelvultrahang és ajakvideó alapú „némabeszéd-interfész” rendszerek kifejlesztéséhez [3,4], bár a technológia nem alkalmazható valós időben a mágneses rezonanciás képalkotás korlátai miatt.



1. ábra. Példa a négy beszélő MRI felvételére.



2. ábra. A CNN-LSTM hálózat blokkdiagramja.



3. ábra. A szubjektív meghallgatásos teszt eredménye, beszélőnként (felül) és összesítve (alul).

Irodalom

[1] Cao, B., M. Kim, J. R. Wang, J. Van Santen, T. Mau, & J. Wang (2018). "Articulation-to-Speech Synthesis Using Articulatory Flesh Point Sensors' Orientation Information". In Proc. Interspeech 2018, pp. 3152–3156. Hyderabad, India.

[2] Csapó, T. G. (2020). "Speaker dependent articulatory-to-acoustic mapping using real-time MRI of the vocal tract," In Proc. Interspeech 2020, 2722-2726, DOI: 10.21437/Interspeech.2020-0015.

[3] Csapó, T. G., Grósz T., Tóth L. & Markó A. (2017). "Beszédszintézis ultrahangos artikulációs felvételekből mély neuronhálók segítségével," in MSZNY 2017, pp. 181–192.

[4] Csapó T. G., Gosztolya G., Grósz T., Tóth L., Markó A., "Némabeszéd-interfész nyelvultrahanggal (Beszédgenerálás artikulációs mozgás alapján)," in Beszédkutatás 2018, kivonat.

[5] Freitas, J., A. J. Ferreira, M. A. T. Figueiredo, A. J. S. Teixeira & M. S. Dias (2014). "Enhancing multimodal silent speech interfaces with feature selection," in Proc. Interspeech 2014, pp. 1169–1173.

[6] Gonzalez, J. A., R. K. Moore, J. M. Gilbert, L. A. Cheah, S. Ell & J. Bai (2016). "A silent speech system based on permanent magnet articulography and direct synthesis," Comput. Speech Lang., vol. 39, pp. 67–87.

[7] Janke, M. & L. Diener (2017). "EMG-to-Speech: Direct Generation of Speech From Facial Electromyographic Signals," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 25, no. 12, pp. 2375–2385.

[8] Narayanan, S., A. Toutios, V. Ramanarayanan, A. Lammert, J. Kim, S. Lee, … M. Proctor (2014). "Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC)," The Journal of the Acoustical Society of America, 136(3), 1307–1311.

[9] Rácz B., Csapó T. G. (2020). "Ajakvideó alapú beszédszintézis konvolúciós és rekurrens mély neurális hálózatokkal," in Beszédtudomány - Speech Science, elfogadva.

Investigation of F0 Estimation Algorithms in Ultrasound-to-Speech Synthesis

Pengyu Dai[1], Mohammed Salah Al-Radhi[1], Tamás Gábor Csapó[1,2]
[1]Department of Telecommunications and Media Informatics
Budapest University of Technology and Economics, Budapest, Hungary
[2]MTA-ELTE Lendület Lingual Articulation Research Group, Budapest, Hungary
pengyudaii@gmail.com, {malradhi,csapot}@tmit.bme.hu

Abstract

This paper shows recent progresses on our Silent Speech Interface (SSI) that translates tongue and lip motions, into audible speech. In our previous studies, the prediction of fundamental frequency (F0) from Ultrasound Tongue Images (UTI) was achieved using articulatory-to-acoustic mapping based on deep learning. We investigated here several discontinuous F0 detection algorithms in this UTI-based SSI system. Besides, our statistical parameters (F0, Maxi-mum Voiced Frequency and Mel-Generalized Cepstrum) are predicted using recent advances in convolutional neural networks, with UTI as input. We found that during the articulatory-to-acoustic mapping experiments, those discontinuous F0 algorithms are predicted with lower error, and they are slightly more natural synthesized speech than the baseline continuous F0 algorithm. Experimental results confirmed that discontinuous algorithms (e.g. Yaapt and Yin) are closest to original speech in objective metrics and subjective listening test.

Introduction

Over the last decade, there has been a significant interest in the articulatory-to-acoustic conversion research field, which is often referred to as Silent Speech Interfaces (SSI) [1]. The SSI system will be highly usefully in some particular situations; for example, in an extremely noisy environment that regular speech is not feasible, or some people have a laryngectomy, or in military usage. A few studies attempted to predict the voicing feature and the F0 curve using articulatory data as input. Nakamura et al. utilized electromyography (EMG) data, while the EMG-to-F0 estimation achieved a correlation of 0.5 and the voiced/unvoiced (V/U) decision accuracy was 84% [2]. Hueber et al. experimented with predicting the V/U parameter along with the spectral features of a vocoder, using ultrasound and lip video as input articulatory data. They applied a feed-forward deep neural network (DNN) for the V/U prediction and attained an accuracy score of 82%, which is very similar to the results of Nakamura et al [3].

Although there has been some research on articulatory-to-F0 prediction, only one deep learning experiments for estimating the F0 curve from ultrasound tongue images alone are proposed. In a previous study, we presented our results for DNN-based F0 estimation from ultrasound images and we attained a correlation rate of 0.74 between the original and the predicted F0 curve [4]. However, in the previous experiments only one F0 estimation algorithm based on Swipe was implemented. Here, we extended our study by investigating different robust F0 techniques: Yaapt, Rapt, DIO, Yin and PnYin. In contrast with our recent work where Swipe worked as a continuous pitch algorithm that implemented with a continuous vocoder, the

new four algorithms are discontinuous and implemented with a discontinuous vocoder. We discovered that in our experiments that all discontinuous algorithms got better values than Swipe in objective and subjective measurements.

Methods

The articulatory input of the system is synchronized 2D ultrasound tongue images and speech signals, whose recording environment was a 'Micro' ultrasound system (Articulate Instruments Ltd.). We applied DNNs to estimate Mel-Generalized Cepstrum-based Line Spectral Pair (MGC-LSP) coefficients, which then served as input to a standard pulse-noise vocoder for speech synthesis. For the analysis and synthesis of speech, a standard vocoder was used from the speech signal processing toolkit (SPTK). To estimate F0 curve, DNNs were used in two major machine learning components, one dedicated to making the voiced/unvoiced decision, while the role of the second was to estimate the actual F0 value for voiced frames. The first tasks, since V/U decision for each frame has a binary output, we treated it as a classification task. While working on the same input images, the second DNN seeks to learn the F0 curve. The outputs of the two DNNs were merged during the evaluation (synthesis) step. For Idiap (baseline), this is achieved by taking the output value of the F0 predictor network where the voicing network decided in favor of voicing, and returning a constant value for frames judged to be unvoiced. For Yaapt and another three algorithms, only those predicted F0 values from voiced frames are used.

We trained DNNs with 5 hidden layers of 1000 ReLU neurons. The F0 parameter was predicted together with the gain and the 12 LSP parameters. This DNN contained 14 linear neurons in its output layer. The network trained for the binary V/U decision task had the same structure, but with a binary classification output layer.

Results

In our experiments, a female Hungarian speaker's data was used. 200 sentences were used for training and validation, and 9 for testing. Performance of F0 detection algorithms are evaluated by comparing their synthesized speech and original speech in objective and subjective measurements. In objective measurements, 5 metrics are used: IS (Itakura–Saito) [5], LLR (log likelihood ratio) [5], CEP (cepstrum distance measures) [6], fwSNRseg (frequency-weighted segmental SNR) [7] and ESTOI (Extended ShortTime Objective Intelligibility) [8]. These metrics show how close they are to original speech in various aspects. As shown in Table 1(note that our goal is to minimize IS, LLR and CEP, while maximize fwSNRseg and ESTOI), Yaapt has the best performance followed by PnYin, and all of the 5 discontinuous algorithms have better performance than baseline. In the subjective listening test, the listeners had to rate the naturalness of each speech in a randomized order relative to the reference (which was the natural sentence), from 0 (very unnatural) to 100 (very natural). As shown in Figure 1, after gathering 20 samples, the result shows that Yin and DIO have the best performance, and all 5 discontinuous algorithms are better than the continuous baseline algorithm. The results of objective and subjective evaluation demonstrated that F0 predicted by discontinuous algorithms synthesized sentences are outperformed the one based on continuous F0 (baseline).

Table 1: Objective Metrics

Method	IS	LLR	CEP	fwSNRseg	ESTOI
Baseline	4.4821	0.6078	4.5801	5.7718	0.3645
RAPT	1.1673	0.5014	3.9928	6.9196	0.3897
Yaapt	**0.5664**	**0.4772**	**3.8166**	**7.1242**	**0.4134**
DIO	1.4039	0.5103	3.9604	7.0647	0.3881
Yin	3.0025	0.5397	4.071	6.8494	0.3754
PnYin	1.3579	0.4831	3.8808	4.969	0.3927

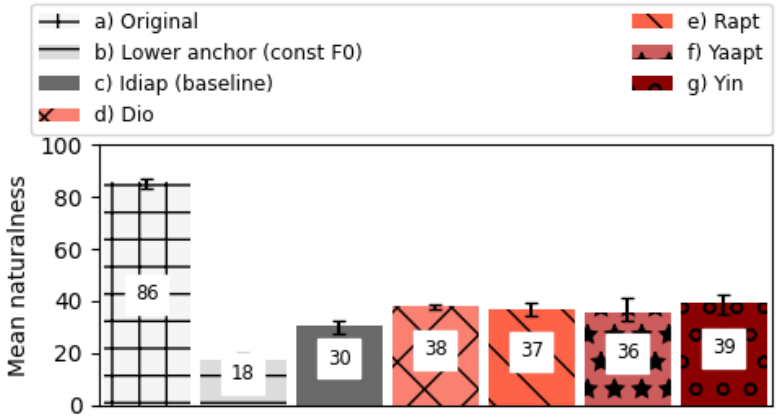


Figure 1: Subjective Listening Test Results

References

[1] Bruce Denby, Tanja Schultz, Kiyoshi Honda, Thomas Hueber, James M. Gilbert, and Jonathan S. Brumberg (2010). 'Silent speech interfaces'. *Speech Communication*, vol. 52, no. 4, pp. 270–287.

[2] Keigo Nakamura, Matthias Janke, Michael Wand, and Tanja Schultz (2011). 'Estimation of fundamental frequency from surface electromyographic data: EMG-to-F0'. In: *Proc. ICASSP*, Prague, Czech Republic, pp. 573–576.

[3] Thomas Hueber, Elie-laurent Benaroya, Bruce Denby, and Gérard Chollet (2011). 'Statistical Mapping Between Articulatory and Acoustic Data for an Ultrasound-Based Silent Speech Interface'. In: *Proc. Interspeech*, Florence, Italy, pp. 593–596.

[4] Tamás Grósz, Gábor Gosztolya, Lászlo Tóth, Tamás Gábor Csapó, and Alexandra Markó (2018). 'F0 Estimation for DNN-Based Ultrasound Silent Speech Interfaces'. In: *Proc ICASSP*, pp. 291-295.

[5] Schuyler R. Quackenbush, Thomas Pinkney Barnwell and Mark A. Clements (1988). Objective Measures of Speech Quality, Prentice-Hall, Englewood Cliffs, NJ, USA.

[6] Kitawaki Nobuhiko, Nagabuchi Hiromi, and Itoh Kenzo. (1988). 'Objective quality evaluation for low bitrate speech coding systems'. *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 262–273.

[7] J. M. Tribolet, P. Noll, B. J. McDermott and R. E. Crochiere (1978). 'A study of complexity and quality of speech waveform coders'. In: *Proc. ICASSP*, Oklahoma, USA, pp.586-590.

[8] Jesper Jensen and Cees H. Taal (2016). 'An Algorithm for Predicting the Intelligibility of Speech Masked by Modulated Noise Maskers'. *IEEE/ACM Trans. Audio Speech Lang. Process*, vol. 24, no. 11, pp. 2009–2022.

Magánhangzók akusztikai és artikulációs változatossága az előrefelé ható és a hátrafelé ható magánhangzók közti koartikulációban – a minőségbeli változás és a variáció mértéke

Deme Andrea[1,2], Bartók Márton[1,2], Csapó Tamás Gábor[2,3], Gráczi Tekla Etelka[2,4] és Markó Alexandra[1,2]
[1]Eötvös Loránd Tudományegyetem, Budapest
[2]MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport
[3]Budapest Műszaki és Gazdaságtudományi Egyetem, Budapest
[4] Nyelvtudományi Intézet, Budapest

Köztudomású, hogy a beszédfolyamban ejtett beszédhangok hatást gyakorolnak egymás ejtésére, és ez akár egymástól távolabb eső szegmentumok, például két, egymástól egy mássalhangzóval elválaszott magánhangzó esetében is megfigyelhető [5]. A szakirodalom szerint erre a magánhangzók (a továbbiakban: V) közötti koartikulációra számos tényező hat, ilyen például a kérdéses V minősége, a koartikuláció iránya, illetve az a tény, hogy a cél- vagy a kiváltó V hangsúlyos helyzetben áll-e. Korábbi eredmények szerint egyes V-k ellenállóbbak (rezisztensebbek) másoknál a koartikuláció módosító hatásaival szemben (pl. [4], ahol a zártabb V-k ellenállóbbnak bizonyultak, és ez a rezisztencia a nyíltsági fok növekedésével csökkent). Továbbá kimutatták, hogy a hangsúly ellenállóbá teszi a V-kat [3, 1, 2], és fokozza a kiváltó V koartikulációs erejét (agressziját, pl. [7]). Végül pedig arra is találni eredményeket, hogy az előrefelé ható (carryover) koartikuláció hatása jelentősebb, mint a hátrafelé hatóé (anticipatory) (pl. [1, 6]). A kísérletes eredmények egymással azonban nehezen olvashatók össze, ugyanis a legtöbb kísérlet a fentiek közül csak egy(-két) tényező szerepét vizsgálta egyszerre, miközben az egyes kísérletek merőben eltérő módszertanokat alkalmaztak, például az egyes vizsgálatokban szinte kivétel nélkül vagy csak a beszédhangok akusztikai szerkezetét, vagy csak az ejtés közben megfigyelhető artikulációs működéseket elemezték a szerzők.

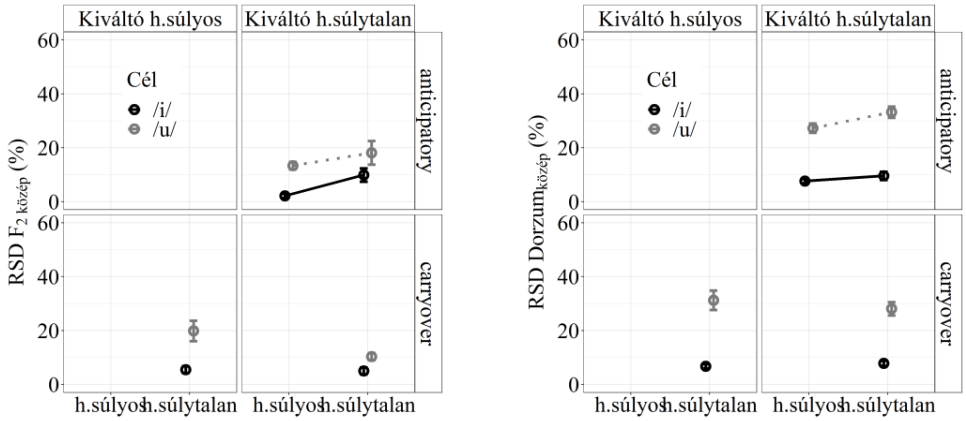
A jelen kísérletben a fenti tényezők közül kettőt, a (mondat)hangsúly hatását (mind a kiváltó, mind a célmagánhangzón) és a koartikuláció irányát mint a V-k közti koartikulációt befolyásoló tényezőket együttesen vizsgáltuk meg a produkció mindkét vetületében, az akusztikumban és az artikulációban egyaránt. Utóbbi elemzéséhez elektromágneses artikulográfiát (EMA) alkalmaztunk, mely eljárás az artikulációs szervekre helyezett tekercsek segítségével teszi mérhetővé az adott szervek elmozdulását az EMA által generált elektromágneses térben. A vizsgálatban az /u/ és /i/ ejtését elemeztük 9 felnőtt magyar anyanyelvű kísérleti személy ejtésében, álszavakban. Mivel ezek a V-k elsősorban a nyelv vízszintes helyzete, másodsorban pedig az ajakműködés szerint térnek el, az akusztikai szerkezetben az F_2 frekvenciaértékét, az artikulációban pedig a nyelv vízszintes elmozdulását vizsgáltuk. A fenti módszertani újítások (az említett szempontok együttes vizsgálata és a produkció mindkét vetületének tekintetbe vétele) mellett kísérletünk további újdonsága, hogy a koartikulációs hatásokat kétféleképpen is számszerűsítettük. Egyfelől a korábbi szakirodalmi megoldások példáját követve leképeztük a V-k minőségbeli változását a szegmentumnak a szomszédos V-hoz közelebbi határánál mért adatok alapján, tehát a minőségében eltérő V-szomszéddal bíró, koartikuláló (pl. *pup<u>u</u>pipi* az *i* vizsgálatára hangsúlytalan helyzetben, ahol a cél V-t félkövér, a kiváltó V-t pedig aláhúzás jelöli) és a neutrális környezetben álló (pl. *pipipipi*), fonológiailag azonos minőségű V-k térbeli és akusztikai távolságát (a továbbiakban *távolság*). Ezekben az adatokban a centralizációt az /i/ esetében a negatív, az /u/ esetében a pozitív előjelű értékek jelölik. Másfelől pedig számszerűsítettük az adott V artikuláci-

ós/akusztikai céljának kontextusok közötti változatosságát (tehát pl. a *pupupipi* és a *pipipipi* célok közös változatosságát) is az egyes paramétereknek a V közepén mért relatív szórása segítségével (a továbbiakban *szóródás*), ahol a nagyobb érték nagyobb változatosságra utal.

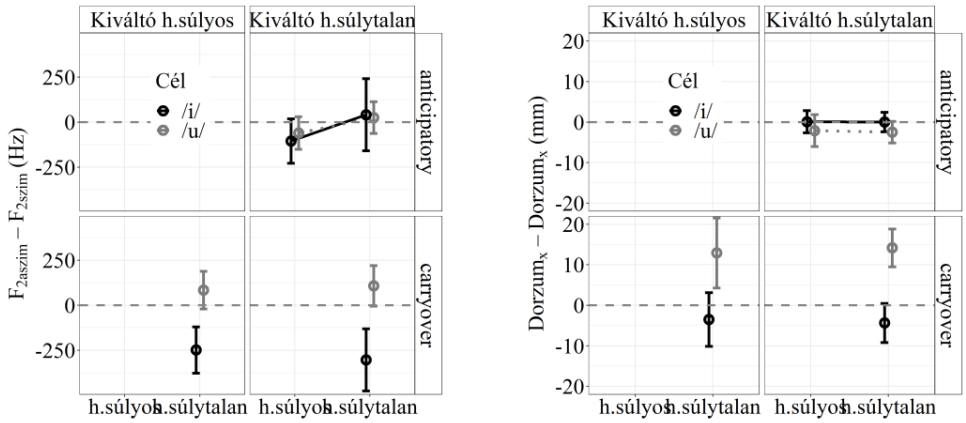
A szóródás és távolság egymástól eltérő képet mutatott a vizsgált hatások tekintetében a következők szerint. A szóródásadatok szerint az /u/ általánosságban változatosabb volt, mint az /i/ (1. ábra), és az /u/-ban nagyobb változatosság volt tapasztalható az artikulációban, mint az akusztikumban, míg az /i/-ben hasonló mértékű volt a szóródás a két doménben. A szóródásadatok a hangsúly tekintetében részben megerősítették a korábban dokumentált tendenciákat. A várt módon kisebb mértékű változatosságot találtunk a hangsúlyos szótagi magánhangzós célban (1. ábra, 2. képnegyed) – de ez az eltérés csak az artikulációban és csak az /u/ esetében volt szignifikáns. Szintén a várt módon nagyobb volt a változatosság akkor, ha a kiváltó V hangsúlyos volt, de ennek az eltérésnek a jelentőségét is csak az /u/ magánhangzóban és csak az artikulációban támasztotta alá a statisztikai elemzés (1. ábra, 3. és 4., képnegyed). A koartikuláció irányát illetően azonban a várakozásokkal szemben a hátrafelé ható (anticipatory) koartikuláció okozott nagyobb változatosságot, tehát ez bizonyult „erősebbnek", bár az eltérés itt sem volt szignifikáns a csoportok között az akusztikumban, csak az artikulációban, és kizárólag az /u/ esetében (1. ábra, 2. és 4. képnegyed).

A távolságadatok szerint akusztikailag az /i/, artikulációsan az /u/ centralizálódott nagyobb mértékben a koartikuláció hatására (2. ábra). Emellett ezek az adatok a szóródással szemben nem mutatták a hangsúly hatását (sem a cél- sem a kiváltó V-n) egyik doménben és egyik magánhangzóban sem. Végül ebben a mérőszámban a koartikuláció irányát illetően a szóródásadatokban látottakkal ellentétes trendek rajzolódtak ki, itt ugyanis a várakozásoknak megfelelően az előrefelé ható (carryover) koartikuláció bizonyult nagyobb hatásúnak (2. ábra, 2. és 4. képnegyed).

Mindezek alapján felmerül a kérdés, hogy a szóródás vagy a távolság paraméterek közül melyik tekinthető a koartikulációs változatosság adekvátabb leképezésének, és mely eredményeket kellene a koartikulációs változatosság irányára, illetve a hangsúly szerepére vonatkoztatnunk.



1. ábra. A magánhangzó közepén, az F_2 adatok alapján megállapított akusztikai változatosság (szóródás) (balra) és a a dorzum vízszintes elmozdulása alapján megállapított artikulációs változatosság (szóródás) (jobbra)



2. ábra. A koartikuláló és neutrális helyzetű tokenek akusztikai (balra) és artikulációs (jobbra) távolsága

Irodalom

[1] Cho, T. (2004). 'Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English'. *Journal of Phonetics* 32, 141–176. old.

[2] Deme, A., M. Bartók, T. E. Gráczi, T. G. Csapó, A. Markó (2019). 'V-to-V coarticulation induced acoustic and articulatory variability of vowels: The effect of pitch-accent'. In: *Interspeech* 2019, 3317–3321. old.

[3] Fowler, C. S. (1981). 'Production and perception of coarticulation among stressed and unstressed vowels'. *Journal of Speech and Hearing Research* 24, 127–139. old.

[4] Mok, P. K. (2011) 'Effects of vowel duration and vowel quality on vowel-to-vowel coarticulation'. *Language and Speech* 54, 527–544. old.

[5] Öhman, S. (1966). 'Coarticulation in VCV utterances: Spectrographic measurements'. *Journal of the Acoustical Society of America* 39, 151–168. old.

[6] Daniel R. (1984). 'Vowel-to-vowel coarticulation in Catalan VCV sequences'. *Journal of the Acoustical Society of America* 76, 1624–1635. old.

[7] Farnetani, E., & D. Recasens (1993). 'Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian'. *Language and Speech* 36, 279–302. old.

Köszönetnyilvánítás

A Bolyai János Kutatási Ösztöndíj, a Tématerületi Kiválósági Program és az Innovációs és Technológiai Minisztérium ÚNKP-20-5 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

Perceptual assimilation and identification of American-English monophthongs by Palestinian-Arabic learners of English

Bashar M. Farran[1]
[1] Pannonia University, Veszprém, Hungary

This research focuses on English as a Foreign Language (EFL) spoken by Palestinian Arabic (PA) learners. We aim to identify (and predict) areas of difficulty in the perception and production of the sounds of American English (AE) so that teaching (materials) can address these rather than spend time on sounds that do not constitute a problem.

We first examined how the eleven AE monophthongs [6] map onto the six vowels of PA [5]. Perceptual Assimilation Model (PAM) predicts learning problems when two L2 phonemes are perceived as good tokens of a Single Category in the L1 (SC scenario) [1]. SC contrasts will yield an incorrect perceptual representation of the AE vowel system in the mind of the EFL learner, with insufficient spectral or temporal separation of categories (compared to native AE listeners). Forty (20 female), adolescent PA high-school learners of EFL listened to the monophthongs of AE (four tokens of each, in different random orders per participant) spoken in /hVd/ words, and classified these as one of the six PA vowels /i, iː, a, aː, u, uː/ while rating them on a 5-point goodness scale. Stimulus presentation and data collection were done by a Praat MFC script [2]. Seven SC contrasts were identified in the results, i.e., *heed-hayed* /iː-eː/, *hid-head* /ɪ-ɛ/, *hud-hood* /ʌ-ʊ/, *hod-hawed* /ɑː-ɔː/, *hawed-hoed* /ɔː-oː/, *hawed-who'd* /ɔː-uː/, and *hoed-who'd* /oː-uː/. Moreover, a Category Goodness (CG, intermediate difficulty predicted) problem was identified for the *had-hod* /æː-ɑː/ contrast. Contrasts that rely on a difference in vowel length did not cause any problems.

The same 40 listeners were asked then to identify all of the 43 artificial vowel sounds (7 degrees of height, 9 degrees of backness/rounding) sampled with perceptually equal steps along the F1 and F2 dimensions of the vowel space, excluding 20 impossible combinations [3], in /mVf/ nonwords with vowel durations of 200 or 300 ms (86 tokens in all). Listeners identified each token as one of the eleven AE monophthongs, while rating them on a 3-point goodness scale. The experiment was repeated with a control group of 20 native AE listeners (10 female). The results show that the PA learners' conception of the AE vowel system is incorrect in several important respects. The EFL learners rely almost exclusively on vowel duration to differentiate spectrally adjacent vowels (in *feel-fill* or *fool-full*), while native listeners of AE rely on vowel quality rather than length, confirming [4]. Also, the EFL learners accept monophthongal /eː/ and /oː/ (as in *sale* and *whole*), which are rejected by the native listeners because of insufficient diphthongization. The vowels in *fill-tell* are not differentiated, and most mid-low vowel sounds are incorrectly identified as /ʌ/ (as in *null*).

Keywords: Palestinian Arabic, EFL, Vowel Perception, PAM-test.

References

[1] Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (ed.), Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research (pp. 167–200). Timonium, MD: York Press.

[2] Boersma, P. & Weenink, D. (2020). Praat, a system for doing phonetics by computer. <www.praat.org>.

[3] Heuven, V. J. van (2017). Perception of English and Dutch checked vowels by early and late bilinguals. Towards a new measure of language dominance. In S. E. Pfenninger & J. Navracsics (eds.) Future research directions for Applied Linguistics (Second Language Acquisition 109). Bristol, Buffalo, Toronto: Multilingual Matters, 73–98. DOI: 10.21832/9781783097135-006.

[4] Hillenbrand, J. M., Clark, M. J. & Houde, R. A. (2000). Some effects of duration on vowel recognition. Journal of the Acoustical Society of America, 108, 3013–3022. DOI: 10.1121/1.1323463.

[5] Thelwall, R. (1990). Illustrations of the IPA: Arabic. Journal of the International Phonetic Association, 20, 37–41. DOI:10.1017/S0025100300004266.

[6] Yavaş, M. (2011) Applied English Phonology. Chichester: Wiley-Blackwell.

Glottal period differences in the vowel of VC sequences with regard to obstruent voicing. Preliminary study on Hungarian

Gráczi, Tekla Etelka[1,5], Csapó, Tamás Gábor[2,5], Deme, Andrea[3,5], Juhász, Kornélia[4,5], Markó, Alexandra[3,5]
[1]Hungarian Research Institute for Linguistics,
[2]Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics,
[3]Eötvös Loránd University, Department of Applied Linguistics and Phonetics,
[4]Eötvös Loránd University, Doctoral School of Lingustics
[5]MTA–ELTE "Lendület" Lingual Articulation Research Group

Introduction

The patterns of the cessation of vocal fold vibration in voiced osbtruents varies across languages and speakers [7]. The vocal fold motion changes while the transglottal pressure drop lowers: The changes are not simultaneous in the upper and lower part of the vocal chords [1]. While [1] describe the ongoing process detailed, their model did not reach the entire cessation of the vibration, it only slowered. However, voiced fricatives may reach entire devoicing, as well. The glottal period patterns in vowels preceding stops showed difference across languages with regard the voicing of the stop consonants [3], but were not described in vowels preceding fricatives – according to the authors' knowledge. As the vocal fold abduction is targeted in voiceless fricatives, but not in voiced once, we can hypothesize that this target difference may appear already in the preceding vowel's duration reaching into the fricative due to coarticulation, as this could be detected in stops [3] The question may arise if there is a difference on the glottal periods of vowels preceding fricatives based on the consonant's phonological and phonetic voicing. Our hypothesis was that the vocal fold vibration will show earlier opening in vowels preceding voiceless fricatives than in ones preceding voiced fricatives. We also hypothesized that the vocal fold vibration patterns will show larger variability in vowels preceding voiced fricatives as speakers can be divided into groups who tend to maintain vocal fold vibrations and into devoicing speakers.

Methods

12 female, monolingual native speakers of Hungarian read aloud 5 sentences starting with /izi nɛ/ and 5 with /isi nɛ/ (as a part of the distractors of a larger experiment). In the present study, the /i/ in /iz/ and /is/ was the target sound. The glottal periods were recorded by a D200 EGG. The speech signal was recorded with a head-mounted microphone. The /i/s, the fricatives and the fricatives' voiced part ratio (VPR) were labelled in Praat [2].

The glottal vibration patterns were described by the amplitude change within and among the glottal periods based on [4] and [5]. The data were collected by modifying [4]'s script. The amplitude of the glottal periods were collected automatically at 40 points per period. Then the amplitude values were normalized between 0 and 1, and the time point along the period was also normalized between 0 and 1. The periods' location is also collected, and normalized to the vowel's duration, between 0 and 1. In case of no detectable period, the values were set to NA. This appeared only rarely and only at the start of V duration due to the phrase initial position.

Generalised additive mixed models (GAMM) were used (R [6]: mgcv [8], itsadug [9]). A model with consonant contrast and another without were compared: the latter described the data better: (Chi-Square test on the ML scores: $\chi^2(8.00)=240.625$, $p<2e\text{-}16$). The final model tested if the glottal periods' amplitude in /i/ showed difference between the /z/ and /s/ contexts (i) in average, (ii) along the normalised period or (iii) vowel duration, (iv) or if the pattern of the periods changed along the V duration. First-order autoregressive model and random smooth by speakers were used.

Results

The amplitude patterns in the time course of /i/ showed difference between /iz/ and /is/ (explained deviance: 92.8%). Overall difference was found between the two conditions ($t(2, 36679) = 4.025$, $p < 0.001$), and also both smooth terms ($F(1, 36679) = 28.114$, $p < 0.001$, $F(1, 36679) = 22.335$, $p < 0.001$) and the tensor ($F(1, 36679) = 6.002$, $p < 0.001$) showed significant difference between them. The point to highlight is that this means that the change of the amplitude within the glottal periods changed differently across the vowel duration between the two conditions. The global pattern of the periods became flatter (Fig. 1). The changes across the V duration were more pronounced in /i/ realisations in /is/ than in /iz/. The opening started earlier within the periods in /i/ realisations preceding /s/ already at the 0.8 point of the vowel. This difference increased towards the VC boundary. The flatter pattern also meant that the open phase showed more abrupt opening in /i/ realisations preceding /z/ in the last 20% of the periods' duration after the 0.8 point of the vowels' duration.

The difference between /i/ realisations in /iz/ and /is/ increased in each speaker regardless of the VPR in /z/s (at least 25% of the /z/s were devoiced in 0-1 /z/s in 6 speakers, in 2-3 /z/s in 3 speakers and in 4-5 /z/s in 3 speakers). One speaker's all /z/s became devoiced already in the first 10% of the duration, while the others' only after the 25%.

Discussion & conclustions

The target of the fricative voicing had an effect on the glottal periods shape in the preceding /i/ realisations suggesting earlier glottal opening start before the voiceless fricative. Though the random effects of the speakers were significant, the tendencies were very similar across speakers regardless of the /z/s voicing profile. As the present preliminary study only analysed one sequence pair and as the maintenance of voicing depends on various articulatory gestures, a larger study including more sequences, more types of obstruents and supraglottal articulatory analyses is needed to build a clear picture.

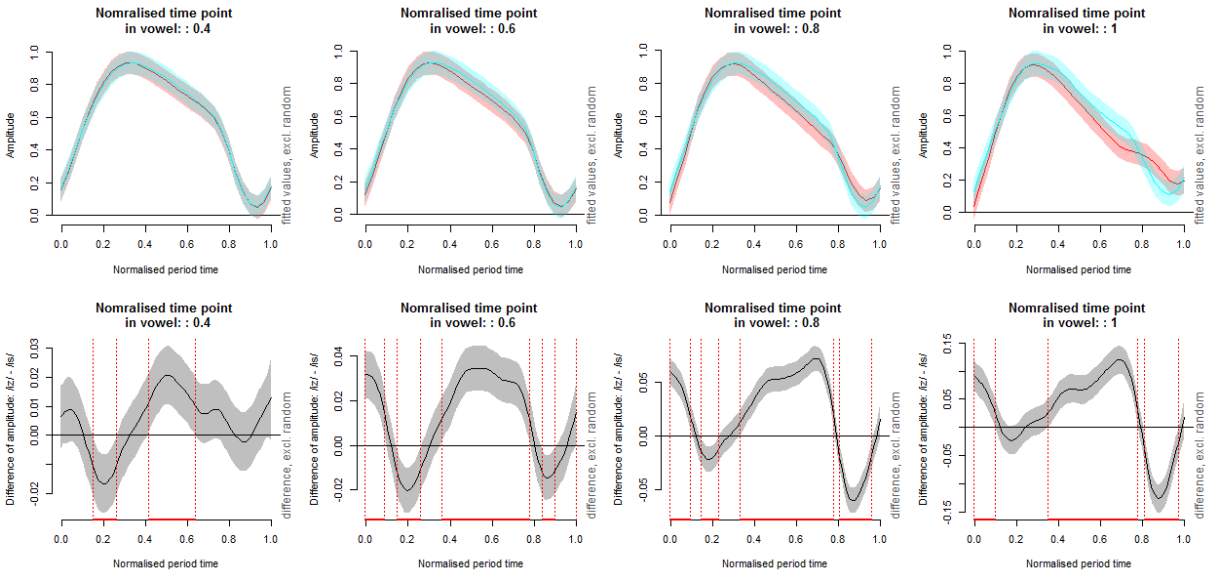


Figure 1: The model fit at 4 normalised time points within /i/ realisations preceding /z/ or /s/
Top panel: The estimated smooth and its 95% confidence intervals. Bottom panel: The estimated difference of the glottal periods (/iz/ - /is/). Red lines indicate the significant differences.

References

[1] Bickley, C. A. & Stevens, K. N. (1986). 'Effects of a vocal-tract constriction on the glottal source: experimental and modelling studies'. *J. Phonetics,* 14, 373–382.

[2] Boersma, P., Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1., http://www.praat.org/ http://www.fon.hum.uva.nl/praat/download_win.html.

[3] Coretta, S. (2017). *coretta2017egg. R data package of Pilot study on EGG data analysis*. https://github.com/stefanocoretta/coretta2017egg

[4] Coretta, S. (2019). *Modelling electroglottographic data with wavegrams and generalised additive mixed models.* https://doi.org/10.31219/osf.io/m623d. https://osf.io/xq8k3/

[5] Herbst, C. T., Fitch, W. T. S., Švec, J. G. (2010). 'Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively'. *The Journal of the Acoustical Society of America,* 128, 3070 (2010); https://doi.org/10.1121/1.3493423

[6] R Core Team (2019). *The R Project for Statistical Computing.* https://www.R-project.org/.

[7] Shih, C., Möbius, B. & Narasimhan, B. (1999). 'Contextual effects on consonantal voicing profiles: A cross-linguistic study'. In: *Proc. 14th ICPhS,* 2, 989-992.

[8] Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R (2nd edition).* Chapman and Hall/CRC

[9] van Rij, J. et al. (2017). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.* R package version 2.3.

Acknowledgments
The paper was supported by the Bolyai János Research Scholarschip of the Hungarian Academy of Sciences, the ÚNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund, and the Thematic Excellence Program of the Ministry for Innovation and Technology.

Production of Brazilian Portuguese lateral sounds by Hungarian learners of L2 Portuguese

Tekla Etelka Gráczi[1,6], Luma Miranda[2], Tamás Gábor Csapó[3,6], Kornélia Juhász[4,6], Alexandra Markó[5,6]
[1] Department of Phonetics, Hungarian Research Institute for Linguistics
[2] Department of Portuguese Language and Literature, Eötvös Loránd University
[3] Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics
[4] Doctoral School of Linguistics, Eötvös Loránd University
[5] Department of Applied Linguistics and Phonetics, Eötvös Loránd University,
[6] MTA–ELTE „Lendület" Lingual Articulation Research Group

Introduction

The second language acquisition research holds that the sound system of a L1 interacts with the sound system of a L2 [8]. The present study explores the Brazilian Portuguese lateral sounds produced by Hungarian speakers with the aim of analysing the influence of their L1 onto the L2 sound system. Brazilian Portuguese phonological system differentiates between the alveolar lateral /l/ and the palatal lateral /ʎ/, while in the standard variety of Hungarian this contrast does not exist. In Hungarian, only the alveolar lateral /l/ is produced, and, at the palatal region, the central approximant /j/ can be found. According to the markedness differential hypothesis (MDH), when a sound of the target language, which differs from the native language, is marked, it causes learning problems [7]. The lateral palatal /ʎ/ is considered marked since this sound is absent in most of the language inventories [7]. Hence, Hungarian speakers might have difficulties in the pronunciation of /ʎ/.

Methods

The aim of this study is to reveal whether L2 Portuguese learners with Hungarian as L1 show the difference between the production of the alveolar lateral /l/ and the palatal lateral /ʎ/ in Brazilian Portuguese minimal pairs. The difference only occurs in the intervocalic position of the words (e.g., "tela": te[l]a 'screen' vs. "telha": te[ʎ]a 'roof tile').

According to the literature [3, 4, 5, 10], the main articulatory differences between the Brazilian Portuguese /l/ and /ʎ/ are shown in (i) the interdental channel (which is typically longer in the palatal laterals than in the alveolar ones); (ii) the position of the tongue body (which is raised in palatal lateral and lowered in the alveolar one); (iii) the tongue root (which is advanced in /ʎ/ and retracted in /l/); and (iv) the supralingual cavity (which is generally longer in the palatal laterals than in the alveolar ones). In the acoustic domain, the alveolar lateral is characterized by a low F2, while the palatal lateral shows a high F2. (Additionally, the palatal lateral differs from the palatal glide primarily in terms of F2 frequency: F2 is higher in the glide).

Our research question is twofold. First, we plan to analyse whether the Hungarian learners produce the lateral sounds /l/ and /ʎ/ in Portuguese, and the abovementioned differences can be found (compared also to the native control group). On the basis of MDH [7], we assume that Hungarian speakers can produce the alveolar lateral /l/ but not the lateral palatal /ʎ/. Second, we plan to reveal the strategies that Hungarian students employ in the production of the lateral palatal in Portuguese. With respect to this, we hypothesize that Hungarians produce the glide /j/,

which is part of their L1 system, instead of the lateral palatal. Some students (with higher proficiency level) may employ the production of /lj/, which is an allophone of the lateral palatal produced by Brazilians, trying to approximate their production to the /ʎ/.

In this study, 4 native speakers of Brazilian Portuguese (all females) and 10 Hungarian learners of Portuguese as L2 (8 females) participated. Minimal pairs embedded in the identically structured sentence were applied as linguistic material. The sentences were presented in a computer screen, and 5 repetitions of target words per subject were recorded in a random order. We recorded ultrasound and acoustic signal synchronously. The tongue contours were recorded in midsagittal orientation using the "Micro" ultrasound system [1]. The speech signal was recorded with a head-mounted microphone (time aligned to the ultrasound recording automatically by the recording software). The segmentation was carried out by forced alignment and manually corrected in Praat [2]. The tongue contours of the target sounds were traced manually in [1] at the closest frame at the middle of the target consonant. Formant values were obtained at 11 equidistant points in the consonant duration. F2 values were z-score normalized per speaker. Generalised additive mixed models were used (GAMM and polar GAMMs: R [9]: mgcv [12], itsadug [11] and rticulate [6]).

Results

The F2 data were analysed by building one basic and two contrastive models. The first one included only the smooth of F2 on the normalized time points of the liquids. The second one included the contrast of the consonants as a parametric term and as smoothed over the normalised time. The third model included the interaction of the speaker group contrast and the consonant contrast, smooth terms for both contrasts. The models were compared if the added terms increased the model fit. The Chi-Square test on the ML scores and AIC scores indicated that adding the abovementioned terms into the model increased the models significantly. Thus, the random smooths for speakers, the autocorrelation treatment and the basis dimension checking were carried out on the model including both contrasts. The difference between the two liquids in the two groups shows that all parametric terms had significant differences. This means that the interaction of the speaker groups (native vs. L2 learners) and the consonants (/l/ and /ʎ/) was significant ($t(4, 15004) = 4.102$, $p < 0.001$). The smooths on the two liquids showed larger distance in the L2 learners' group (Fig. 1).

Discussion

The F2 data showed that L2 learners make larger difference between the target liquids (/l/ and /ʎ/) than the native speakers. This result suggest that Hungarian L2 learners of Brazilian Portuguese pronounce [j]-like consonant instead of the target /ʎ/. The outcome of this study can be applied in the design of pronunciation training of Portuguese as L2.

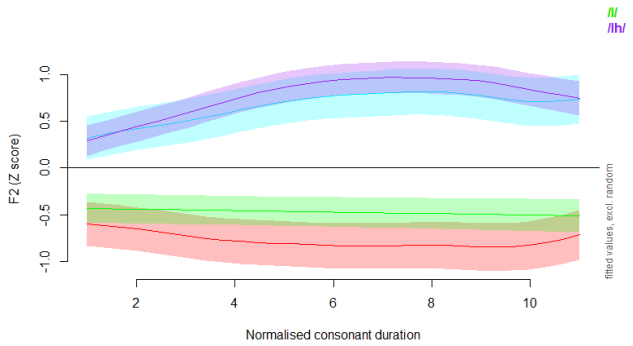


Figure 1: The estimated F2-values (in Z-scores) and their 95% confidence intervals for /l/ and /ʎ/ (noted as /lh/) along the 11 equidistant points in the consonants' duration. In the native group, the red line indicates the results for /l/ and the blue line shows the results for /ʎ/; in the L2 learner group, the green line shows the results for /l/ and the purple line indicates the results for /ʎ/.

References

[1] Articulate Instruments Ltd: AAA software.
[2] Boersma, P.; Weenink, D. (2019). *Praat: doing phonetics by computer [Computer program].* http://www.fon.hum.uva.nl/praat.
[3] Casero, K. T. B.; Brum-De-Paula, M. R.; Ferreira Gonçalves, G. (2016). 'A consoante lateral palatal: análise acústica e articulatória à luz da Fonologia Gestual'. *ReVEL,* v. 14, n. 27, [www.revel.inf.br].
[4] Charles, S.; Lulich, S. M. (2018). 'Case Study of Brazilian Portuguese Laterals using a Novel Articulatory-Acoustic Methodology with 3D/4D Ultrasound'. *Speech Communication,* 103, 37–48.
[5] Charles, S.; Lulich, S. M. (2019). 'Articulatory-acoustic relations in the production of alveolar and palatal lateral sounds in Brazilian Portuguese'. *Journal of the Acoustical Society of America,* 145(16), 3269–3288.
[6] Coretta, S. (2019). *rticulate: Ultrasound Tongue Imaging in R. R package version 1.6.* https://github.com/stefanocoretta/rticulate
[7] Eckman, F. (1977). 'Markedness and the contrastive analysis hypothesis'. *Language Learning* 27, 315–330.
[8] Flege, J. E. (1995). 'Second Language Speech Learning: Theory, Findings, and Problems'. In: Strange, W. (Ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press, pp. 233–277.
[9] R Core Team. (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing,* Vienna, Austria. URL https://www.R-project.org/
[10] Silva, A. H. P. (1996). *Para a descrição fonético-acústica das líquidas no português brasileiro: dados de um informante paulistano.* Master's Thesis, Universidade Estadual de Campinas.
[11] van Rij, J. et al. (2017). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.* R package version 2.3.
[12] Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R (2nd edition).* Chapman and Hall/CRC

A néma szünetek és a hallható levegővétel összefüggései L1 és L2 spontán beszédben

Gyarmathy Dorottya
Nyelvtudományi Intézet

A néma szünet a spontán beszéd leggyakoribb jelensége [8]; gyakoriságát és időtartamát befolyásolja a beszélő személye, fizikai állapota, életkora, neme, beszédben való jártassága, a beszédhelyzet [7], a beszédkörnyezet, a beszédtípus [16]. Meghatározó továbbá a nyelv [11], a szintaktikai tényezők [6], illetve a közlésben elfoglalt hely. A szupraszegmentális szerkezet nyelvspecifikus [9], az erre irányuló tervezési folyamatok kevésbé tudatosak, mint a szegmentális szerkezetre vonatkozóak, így a nyelvtanulóknak több nehézséget okozhat az adott nyelvre jellemző szupraszegmentumok produkciója, mint a beszédhangok, a szókincs, a nyelvtani szabályok elsajátítása és alkalmazása [10].

A néma szünet közlésbeli funkciói sokrétűek: fiziológiai szükséglet, értelmi tagolás, gondolkodási/hatásszünet, új információ jelzése, diskurzusszervezői szerep [2]. A korai kutatások megkülönböztették a tervezési nehézségből adódó néma szünetet, illetve a szintaktikai szerkezet határán létrejövő junktúrát [1]; a közelmúltban pedig magyar és idegennyelvi spontán beszédre is igazolták a néma szünetek funkció- és pozíciófüggő realizációit [4].

A hallható levegővétel és a néma szünet összefüggéseit vizsgáló korai kutatások megállapították, hogy hallható levegővétel általában a prozódiai szerkezet határán tartott szünetekben fordul elő, és a levegővételt tartalmazó szünetek általában hosszabbak [3, 5].

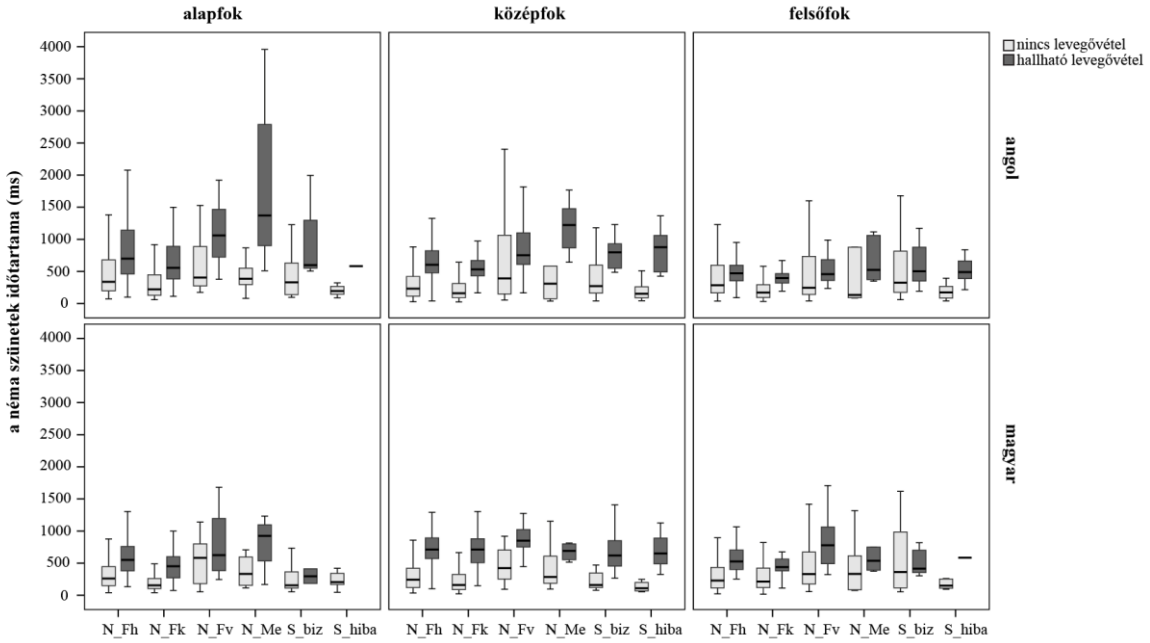
Magyarra még nem született olyan vizsgálat, amely a néma szünet funkcióinak és a hallható levegővételnek az összefüggéseit elemezné anya- és idegennyelvi spontán beszédben. Kutatásunk magyar anyanyelvűek L1 és L2 (angol) spontán beszédében elemzi a néma szünetek előfordulási és temporális mintázatait a hallható levegővétel függvényében. Célunk L2-re is igazolni, hogy a légzés fiziológiai szükséglete a tervezési folyamatoknak van alárendelve, továbbá feltárni a nyelvek közti összefüggéseket és különbségeket a nyelvtudási szintek alapján.

A kutatásban a Nyelvtudományi Intézet Fonetikai Osztályán készülő HunEng Adatbázis 12 beszélőjének L1 és L2 felvételét elemeztük a nyelvtudási szintenként (B1, B2, C1). A hanganyagok hossza mintegy 118 perc, ebből az L1 beszédanyag 55, az L2 63 percet tett ki. A felvételekben összesen 3477 db néma szünetet adatoltunk, az anyanyelvi beszédrészben 1368 db-ot, az idegen nyelviben 2109 db-ot; a néma szünetekben előforduló levegővételt csak abban az esetben vontuk be az elemzésbe, amikor az tisztán meghatározható (hallható és a hangszínképen látható) volt. A szüneteket az általunk korábban kidolgozott kategóriarendszer alapján osztályoztuk (vö. [4]). Elsődlegesen aszerint különítettük el őket, hogy 1. azok megakadásjelenségekhez köthetők-e, azaz szerkesztési szakaszként funkcionálnak-e (hiba detektálására és/vagy javítására biztosítanak időt), vagy 2. az értelmi tagolást szolgálják-e. Az értelmi tagolást segítő néma szüneteket N, a szerkesztési szakaszokat S betűvel jelöltük. Mindkét fő csoporton belül elkülöníthetők további alcsoportok. A tagolási pozícióban megjelenő néma szüneteket (N) a közlésben elfoglalt helyük szerint különítettük el egymástól; így megkülönböztettük a megnyilatkozás eleji (N_Me) néma szüneteket, a frázishatáron lévő (N_Fh) néma szüneteket, a frázisközi (N_Fk) szüneteket és a frázisvégi (N_Fv) szüneteket. A

szerkesztési szakaszokat (S) aszerint kategorizáltuk, hogy hiba típusú (S_hiba), vagy a beszélő bizonytalanságából fakadó (S_biz) megakadásjelenségekhez kapcsolódnak. Az annotálást a Praat, a statisztikai elemzéseket az SPSS szoftverrel végeztük.

Feltételeztük, hogy 1. a néma szünetek típusai elkülöníthetők gyakoriságuk és időtartamuk szerint mindkét nyelvben; 2. minél magasabb a nyelvtudási szint, az idegennyelvi megnyilatkozások temporális paraméterei annál hasonlóbbak az anyanyelvi stratégiákhoz; és 3. a szünettípusok gyakorisága és/vagy időtartama mindkét nyelven eltérően alakul a hallható levegővétel függvényében.

Az eredmények igazolták, hogy a szünet típusa meghatározza az előfordulási gyakoriságot és azidőtartamot, valamint a hallható levegővétel megjelenését. A szünettartási stratégiák a nyelvtudás szintje szerint eltérően alakultak. A statisztikai elemzés bebizonyította, hogy a néma szünetek időtartamát elsődleges a szünettípus határozza meg; a nyelv és a nyelvtudási szint csak gyenge hatásként érvényesült. A hallható levegővételt tartalmazó szünetek nyelvtől, nyelvtudási szinttől és szünettípustól függetlenül hosszabb időtartamban realizálódtak, az előfordulásuk azonban az egyes szünetkategóriákat eltérően érinti. A grammatikai-szintaktikai szerepet betöltő szünetek és a szerkesztési szakaszként realizálódók élesen elkülönültek; a hallható levegővétel főként a tagoló szüneteket jellemezte, ami igazolja, hogy a légzés fiziológiai szükséglete a beszédtervezési folyamatoknak van alárendelve. Eredményeink arra is rámutattak, hogy a nyelvtudási szint függvényében eltérően alakulnak a szünettartási stratégiák. A három csoport között (alap-, közép- és felsőfokú) olyan eltéréseket találtunk az egyes szünettípusok időtartamában, amelyek megmutatják, hogy a beszélők az adott szinten milyen problémákkal küzdenek. Általánosságban minden csoportra jellemző, hogy L2 megnyilatkozásaikban több és hosszabb szünetet tartottak. A leginkább széttagolt az alapfokúak L2 beszéde volt; a magyarhoz képest a grammatikai struktúrát megtörő szünetek előfordulása nőtt leginkább, ami jelzi nyelvi nehézségeiket. A középfokú beszélőknél egyfajta stratégiaváltás látszik; az angol rész szünetidőtartamai tendálnak a magyarhoz. A gyakorisági mutatók és az átlagidőtartamok legkevésbé a felsőfokúak csoportjában különböztek nyelvenként. Elmondható tehát, hogy minél magasabb a nyelvtudási szint, az idegen nyelvi szünettartási stratégiák annál inkább idomulnak az anyanyelvihez (1. ábra).



1. ábra. A néma szünetek időtartama L1 és L2 spontán beszédben a nyelvtudás szintje és a hallható levegővétel függvényében

Irodalom

[1] Boomer, D. S. (1965). 'Hesitation and grammatical encoding'. *Language and Speech* 8. 148–158.

[2] Esposito, A., V. Stejskal, Z. Smékal & N. Bourbakis (2007). *The significance of empty speech pauses: Cognitive and algorithmic issues. Advances in Brain, Vision, and Artificial Intelligence*. Berlin–Heidelberg: Springer. 542–554.

[3] Grosjean, F. & M. Collins (1979). 'Breathing, pausing and reading'. *Phonetica 36*. 98–114.

[4] Gyarmathy, D. (2017). 'A néma szünetek funkciói a spontán beszédben'. *Beszédkutatás 2017*. 67–93.

[5] Gyarmathy, D. (2019). 'A néma szünetek és a hallható levegővétel viszonya a spontán beszédben'. *Beszédkutatás 2019*. 154–186.

[6] Krivokapic, J. (2007). 'Prosodic planning: Effects of phrasal lenght and complexity on pause duration'. *Journal of Phonetics* 35/2. 162–179.

[7] Markó, A. (2005). 'A temporális szerkezet jellegzetességei eltérő kommunikációs helyzetekben'. *Beszédkutatás 2005*. 63–77.

[8] Misono, Y. & S. Kiritani (1990). 'The distribution pattern of pauses in lecture-style speech'. *Logopedics and Phoniatrics* 2. 110–113.

[9] Szaszák, Gy. (2008). *A szupraszegmentális jellemzők szerepe és felhasználása a beszédfelismerésben*. PhD értekezés. Budapest. BME.

[10] Trofimovich, P. & W. Baker (2006) 'Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech'. *Studies in second language acquisition* 28/1. 1–30.

[11] Trouvain, J., C. Fauth & B. Möbius (2016). 'Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 Read Speech'. *Proceedings of Speech Prosody 2016*. Boston. 31–35.

A kutatást a Bolyai János Kutatási Ösztöndíj támogatta.

A beszélőváltások dinamikus jellemzői II.: Pragmatikai/társalgáselemzési aspektusok
Lexikális elemek a fordulók végén és kezdetén a magyar társalgásokban: a *'tehát'*, az *'úgyhogy'* és a *'hát'*

Hámori Ágnes[1], Dér Csilla Ilona[2]
[1]Nyelvtudományi Intézet, Budapest
[2]KRE Bölcsészettudományi Kar, Budapest

Az előadás egy tágabb kutatás részeként a fordulók elején és végén megjelenő tipikus kifejezések néhány fő típusát tárgyalja, elsősorban a fordulóváltások rugalmas és gördülékeny szerveződése, valamint a társalgásbeli közös jelentéslétrehozás szempontjából. A konverzációelemzésben kiemelt kérdéskörnek számít a beszélőváltást segítő mechanizmusok és nyelvi jelenségek (szómegtartó, -átadó vagy -átvevő jelzések) köre, melyek közé lexikális elemek, kifejezések is tartoznak [3], az utóbbiakat mégis viszonylag kevés önálló kutatás vizsgálta ilyen keretben (például [4, 5]). Ugyanakkor más nyelvészeti megközelítésben több munka is foglalkozik a fordulók elején és végén álló tipikus elemekkel, különösen a korpusznyelvészet vagy a diskurzusjelölő-kutatás körében [1, 2]. Mindezek a kutatások rámutatnak a fenti jelenségkör fontosságára, ugyanakkor néhány lényeges problémára is (pl. a funkciók meghatározásának nehézsége).

A jelen kutatásban a BEA adatbázis 10 magyar nyelvű, háromszereplős társalgását (összesen 131 percnyi beszélgetést) elemeztünk a fordulók végén és kezdetén álló tipikus lexikai elemek szempontjából, ezen belül néhány kiemelten gyakori elem − a fordulóvégi *'tehát', 'hát'* és *'úgyhogy'* és a fordulókezdő *'hát'* – társalgásbeli (ezen belül főleg a beszélőváltással kapcsolatos) jellegzetességét és szerepét vizsgálva. Ennek során arra is keressük a választ, hogy a fenti elemek gyakorisága ebben a pozícióban hogyan függ össze diskurzusbeli funkciójukkal. A kiindulópontként használt mennyiségi elemzés után elsősorban a tipikus példák kvalitatív, a lokálisan érvényes jelentéseket és funkciókat feltáró elemzésére fókuszáltunk, a társalgáselemzés és a társas-kognitív pragmatika funkcionális megközelítésének összekapcsolásával.

Az eredmények a fordulóvégi elemekre vonatkozóan azt mutatták, hogy a vizsgált társalgásokban a *'tehát', 'hát', 'úgyhogy'*, *'hogy*', *'így'* és *'ilyesmi'* szerepelt leggyakoribb tipikus fordulózáró elemként (*tehát/dehát/hát*: 21, *úgyhogy:* 12, *hogy:* 5, *így:* 4, *ilyesmi*: 3 alkalommal). A beszélgetésekben ezek szóátadó és szómegtartó („turn-yielding" és „turn-holding") funkcióban is állhattak. Ez a kettősség nem esetleges jellegű, hanem egy speciális („indirekt") fordulózárási módhoz kapcsolódik („trail-off", [5, 6]), mely egy önálló fordulókonstrukció-típust is alkot: az elemzések alapján bemutatjuk, hogy ez itt egy olyan fordulóváltási technika eszköze, mely a szemantikai bizonytalanság pragmatikai kihasználásával a beszélőváltás rugalmas, közös és egyezkedéses alakítását teszi lehetővé a résztvevők között. Eközben lehetőséget ad a beszélőváltás gördülékenységének megőrzésére, és társas szempontoktól sem független.

A szóátvételkor gyakori *'hát', 'és', 'de'* diskurzusjelölők közül a *'hát'* erős multifunkcionalitását középpontba állítva mutatjuk meg, hogy e vonása ideális fordulókezdő elemmé teszi: a beszélő képes vele felvezetni a téma kidolgozását, a saját véleményét és érzelmeit, csökkenteni az ellentmondás erejét (vö. *'igen de hát'*), valamint tervezési időt biztosítani maga számára – akár egyazon használat során. A *'hát'* feltűnően gyakori az olyan társalgásrészekben, ahol ismerős beszédpartnerek fordulói váltják egymást gyorsan, és informális jellege is elősegíti a közös jelentéskonstruálás folyamatát.

Irodalom

[1] Degand, L. & van Bergen, G. (2016). ’Discourse Markers as Turn-Transition Devices: Evidence From Speech and Instant Messaging’. *Discourse Processes* 2016. Vol. 00, No. 0., 1–25. old.

[2] Dér, Cs. I. (2012). ’Beszélőváltások során használt diskurzusjelölők a magyar spontán beszédben’. *Beszédkutatás,* 2012, 130–141. old.

[3] Duncan, S. (1972). ’Some signals and rules for taking speaking turns in conversations’. *Journal of Personality and Social Psychology 23*(2), 283–292. old.

[4] Gravano, A. & Hirschberg, J. (2009). ’Turn-Yielding Cues in Task-Oriented Dialogue’. *Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue.* Queen Mary University of London. Association for Computational Linguistics. 253–261. old.

[5] Local, J. K., Wells, B. H. G. & Sebba, M. (1985). ’Phonology for conversation. Phonetic aspects of turn delimitation in London Jamaican’. *Journal of Pragmatics* 9, 309–330. old.

[6] Schegloff, E. A. (1996). ’Turn organization: One intersection of grammar and interaction’. *Interaction and grammar*. Szerk. Ochs, E., Schegloff, E. A. & Thompson, S. A. Cambridge, England: Cambridge University Press. 52–133. old.

A kutatást az NKFIH K-128810 számú pályázata támogatta.

Language specific representations in word stress perception: ERP evidence

Ferenc Honbolygó[1,2]
[1]Research Centre for Natural Sciences, Budapest, Hungary
[2]Institute of Psychology, Eötvös Loránd University, Budapest, Hungary
honbolygo.ferenc@ttk.hu

During speech perception, segmental and suprasegmental or prosodic information is extracted from the speech input simultaneously and are encoded in separate memory representations. Thus, when listeners recognize speech, they are processing a prosodically determined variant [1]. We can identify several different types of prosodic information, the most important ones being length, rhythm, intonation and stress [3]. Stress, the focus of the present talk is a relative emphasis given to certain syllables within words or to certain words in sentences [7]. Word stress potentially contributes to the segmentation of continuous speech into words [2]. Languages differ considerably in the use of word stress: in the position of the stressed syllable within multisyllabic words (initial, final, penultimate, etc.); in the variability of the stressed syllable's position (free or fixed); and whether stress can distinguish lexical meaning (contrastive or non-contrastive). Fixed-stress languages (like Hungarian) mandatorily assign syllable stress to a specific position within a word, and stress is non-contrastive. Therefore, it can be assumed that stress processing demonstrates language specificity.

In this talk, I will present event-related brain potential (ERP) results related to the processing of word stress patterns from three different studies [4, 6, 5], using meaningless pseudowords, meaningful words and pseudowords spoken by foreign speakers. The basic premise of the studies was that words stress is processed in relation to long-term representation, which are pre-lexical and language specific. Consequently, we can assume that both words and pseudowords are processed similarly, and foreign words are processed differently from native words. We applied the passive oddball paradigm to elicit the Mismatch Negativity (MMN) ERP component, a fronto-centrally negative waveform appearing to the pre-attentive detection of violation of simple or complex regularities [8].

In the experiments, Hungarian participants heard disyllabic words (bɑˈbɑ, meaning 'baby'), pseudowords (bɛˈbɛ) and pseudowords pronounced either by a Hungarian or a German speaker (beːˈbeː). Stress could be either on the first (legal stress) or on the second (illegal stress) syllable. The experimental design in all three experiments was similar. We used two conditions: in the first, stimuli with the legal stress were standards and stimuli with illegal stress were deviants; in the second condition, the standards and the deviants were reversed. This allowed us to calculate the MMN component by subtracting ERPs to the standard from the ERPs to the deviant using physically identical stimuli.

Results showed that the pseudoword deviant having an illegal stress pattern elicited two consecutive ERP components that were considered as MMN, whereas the deviant having a legal stress pattern did not elicit MMN. Moreover, pseudowords with a legal stress pattern elicited very similar ERP responses irrespective of their role in the oddball sequence, i.e., if they were standards or deviants, demonstrating that their processing relied on long-term instead of short-term (stimulus sequence related) memory traces (see Fig. 1.). Meaningful words elicited similar

ERP deflections, but the lexical status slightly modulated the processing of stress patterns. This modulation was different for the legal and illegal stress patterns (see Fig. 2.). Finally, the comparison of pseudowords with native and non-native pronunciation showed that all pseudowords in the deviant position elicited an Early Differentiating Negativity (EDN) and a Mismatch Negativity (MMN) component, except for the Hungarian pseudowords stressed on the first syllable (see Fig. 3.). This suggests that Hungarian listeners did not process the native legal stress pattern as deviant, but the same stress pattern with a non-native accent was processed as deviant.

In conclusion, our data show that word stress pattern change is processed preattentively by the human brain, and this is slightly modulated by the lexical status of the word. Words and pseudowords with legal stress pattern do not initiate an error detection mechanism, as indicated by the lack of MMN component, but only if they are native sounding. We suggest that this is an evidence that word stress is processed based on prelexical language specific stress representations.

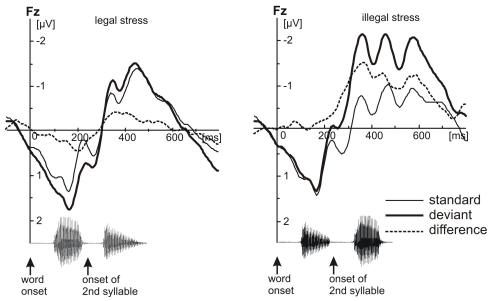


Figure 1: ERPs elicited by the pseudoword with legal and illegal stress pattern in the standard and deviant positions. The figures depict ERPs to the same stimulus in two different positions, and the difference wave obtained by subtracting the ERPs to the same pseudoword in the deviant and standard positions.

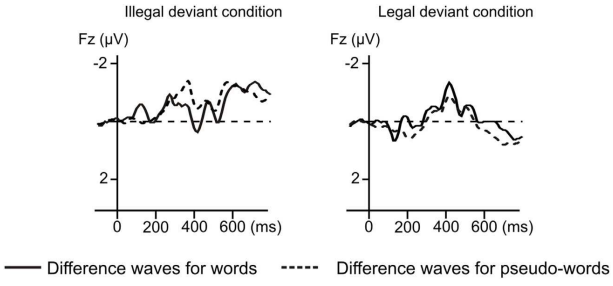


Figure 2: Difference wave ERPs for words and pseudowords.

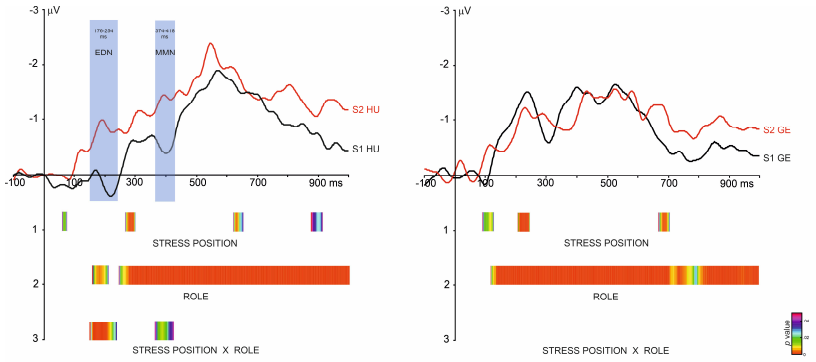


Figure 3: Difference wave ERPs cross-linguistic study. Difference waves in the four different conditions split by language (Hungarian: left side; German: right side) and the TANOVA results. The colored intervals show those time windows where the TANOVA indicated significant main effects of Stress Position, Role, and Stress Position * Role interaction. Note: S1: stimuli with stress on the first syllable; S2: stimuli with stress on the second syllable; HU: Hungarian stimuli; GE: German stimuli. EDN: Early Differentiating Negativity.

References

[1] Cutler, A., Dahan, D., & Van Donselaar, W. (1997). 'Prosody in the comprehension of spoken language: A literature review'. In: *Language and Speech,* 40(2), pp. 141–201.

[2] Cutler, A., & Norris, D. (1988). 'The Role of Strong Syllables in Segmentation for Lexical Access'. In: *Journal of Experimental Psychology: Human Perception and Performance,* 14(1), pp. 113–121. https://doi.org/10.1037/0096-1523.14.1.113

[3] Fox, A. (2000). *Prosodic features and prosodic structure: the phonology of suprasegmentals.* Oxford University Press, USA.

[4] Garami, L., Ragó, A., Honbolygó, F., & Csépe, V. (2017). 'Lexical influence on stress processing in a fixed-stress language'. In: *International Journal of Psychophysiology,* 117, pp. 10–16. https://doi.org/10.1016/j.ijpsycho.2017.03.006

[5] Honbolygó, F., & Csépe, V. (2013). 'Saliency or template? ERP evidence for long-term representation of word stress'. In: *International Journal of Psychophysiology,* 87(2), 165–172. https://doi.org/10.1016/j.ijpsycho.2012.12.005

[6] Honbolygó, F., Kóbor, A., German, B., & Csépe, V. (2020). 'Word stress representations are language-specific: Evidence from event-related brain potentials'. In: *Psychophysiology,* 57(5). https://doi.org/10.1111/psyp.13541

[7] Kager, R. (2010). 'Feet and metrical stress'. In: *The Cambridge Handbook of Phonology.* Ed. by de Lacy, P. Cambridge: Cambridge University Press, pp. 195–228. https://doi.org/10.1017/cbo9780511486371.010

[8] Winkler, I., Denham, S. L., & Nelken, I. (2009). 'Modeling the auditory scene: predictive regularity representations and perceptual objects'. In: *Trends in Cognitive Sciences,* 13(12), pp. 532–540. https://doi.org/10.1016/j.tics.2009.09.003

A beszélőváltások dinamikus jellemzői I.: Fonetikai aspektus

Horváth Viktória, Huszár Anna, Krepsz Valéria és Gyarmathy Dorottya
Nyelvtudományi Intézet

A beszédfordulók és a beszélőváltások szerveződése a konverzációelemzés egyik központi kérdése a kezdeti kutatások óta [1]. Az elemzések középpontjában leggyakrabban az a kérdés áll, hogy miként képesek a résztvevők a beszélőváltások gördülékeny lebonyolítására olyan módon, hogy az egyik beszélő közlésének végét követően a partner igen gyorsan megszólal. Az első időzítési modellek (amelyek azt feltételezték, hogy a szóváltás a két megnyilatkozás közötti néma szünet formájában valósul meg) legnagyobb kritikája az volt, hogy nem kezeli sem az egyszerre beszélést mint kommunikációs jelenséget, sem a néma szünetek által betöltött kommunikációs funkciókat (pl. egyet nem értés, figyelemfelhívás). Az újabb kutatások eredményei szerint a beszédjog, valamint a beszélői és hallgatói szerepek rendkívül gyors átadása és cseréje úgy válik lehetővé, hogy a beszédpartnerek képesek különböző nyelvi jelekből valószínűsíteni a beszélő fordulójának végét, és azonosítani a lehetséges szóátvételi pontokat bizonyos szemantikai, szintaktikai, grammatikai, illetve fonetikai jellemzők, kulcsok alapján. Ilyen módon egyidejűleg történik a feldolgozás és a következő, kapcsolódó megszólalás, valamint a szerepváltás tervezése [vö. pl. 2, 3, 4, 5, 6].

A korábbi magyar, korpusz-alapú kutatások eredményei szerint a beszélőváltások gyakorisága növekszik a beszélgetések első és utolsó szakasza között [7]. Emellett a beszélgetések elején nagyobb a külválasztásos beszélőváltások gyakorisága, majd az idő előrehaladtával a társalgás folyamán nő az önkiválasztásos beszélőváltások aránya [8].

A jelen kutatás célja a. egyes beszédparaméterek (szünetek, artikulációs tempó) fonetikai szempontú vizsgálata és a háttércsatorna-jelzések sajátosságainak önálló elemzése, valamint ezen, a társalgás szerkezetét meghatározó fonetikai és pragmatikai paraméterek együtt járásának elemzése, b. ezen tényezők dinamikus változásának elemzése a megelőző és a követő váltás távolságának függvényében, c. a teljes felvételben elfoglalt pozíciójának függvényében.

Hipotéziseink szerint: 1. A különféle típusú beszélőváltások gyakorisága változik a beszélgetés kezdete és vége között: a beszélgetések elején jellemzőbb a néma szünettel, a vége felé az együtt beszéléssel történő szóátvétel. 2a. Különbség mutatkozik a beszélőváltáshoz közeli és az attól távoli beszédjellemzők között, 2b. a váltások előtti és az azokat követő jellemzők alakulásában.

A vizsgálathoz 10 háromfős társalgást választottunk ki a BEA adatbázis társalgási részfeladatából [9]. Az adatközlők 20 és 35 év közötti beszélők voltak (5 ffi és 5 nő). Az interjúkészítő és a társalgó partner azonos volt minden felvétel esetén (a felvételkészítés időtartama alatt 28 és 35 év közöttiek voltak). A beszélők közléseinek meglévő beszédszakasz szintű annotációját kiegészítve jelöltük a háttércsatorna-jelzéseket, a beszélőváltásokat, az egyszerre beszéléseket, valamint a narratív és dialogikus szakaszokat a társalgásokban [10] alapján). Elemeztük a háttércsatorna-jelenségek, az egyszerre-beszélések és a néma szünetek gyakoriságát és időtartamát, az artikulációs tempót minden beszédszakaszban a beszélőváltások közelében és azoktól függetlenül, megkülönböztetve a váltások előtti és az azokat követő szakaszok jellemzőit egymástól. Emellett megvizsgáltuk a beszélőváltások jellemzőit: azok típusát, gyakoriságát és időtartamát. A statisztikai elemzés során

kevert modellt építettünk, post hoc elemzést és cluster elemzést végeztünk a vizsgált jelenségek beszélőváltási pontoktól való időbeli távolsága alapján.

A kutatás eredményei szerint átlagosan 3 szóátvétel történt percenként a társalgásokban, de jelentősek voltak az egyéni különbségek (a legtöbb váltást megvalósító közlésben a percenkénti előfordulás 5,39 volt, a legkevesebbet tartalmazóban pedig csupán 1,24 fordult elő átlagosan). A társalgások elején a szóátvételek nagyobb arányban valósultak meg néma szünetként, míg a társalgások vége felé közeledve egyre gyakoribb volt az egyszerre beszélésként realizálódó beszélő-hallgató szerepváltás, tehát összefüggés mutatkozott a váltások pozíciója és típusa között. Az artikulációs tempó szignifikánsan eltért az egyes beszélők (B) között, valamint az interjúkészítő (IK) artikulációs tempója is szignifikánsan különbözött az egyes felvételek, azaz az egyes beszélőkkel folytatott társalgásokban. Az interjúkészítő által megvalósított szünetek jelentős variabilitást mutattak, időtartamukat szignifikánsan meghatározza a beszélgetőpartner személye, a szünet típusa, illetve ezek együttes hatása. Mindhárom beszélő szünettartási stratégiáinak vizsgálata rávilágított arra, hogy a beszélői szerepkör, a szünet típusa, valamint az adott szünetnek a megelőző beszélőváltástól való távolsága jelentős befolyással bír annak időtartamára. A szóátvételt megelőzően jelentősen több háttércsatorna-jelzés valósult meg, mint a szóátadást követően függetlenül attól, hogy azok a váltás szűk vagy tágabb környezetéhez tartoznak-e. A háttércsatorna-jelzések időtartamát és gyakoriságát sem befolyásolta a pozíciójuk, kizárólag azok típusa (a legrövidebbek és leggyakoribbak a hümmögés típusúak voltak, a legkisebb arányú és leghosszabb időtartamúak a kevertek (verbális, nonverbális és hümmögések kombinációi)). Az előfordulási gyakoriságra hatással volt a beszélői szerep is: a legtöbb háttércsatorna-jelzés az interjúkészítő közlésében jelent meg, fő funkciója ugyanis a figyelem fenntartásának, a támogatásnak jelzése,

Az egyes tényezők összefüggése szerint az eredmények azt mutatták, hogy az artikulációs tempó nem változott, a szünettartások időtartama rövidült, a háttércsatorna-jelzések és az egyszerre beszélés gyakorisága pedig fokozatosan nőtt a társalgások folyamán az idő előrehaladtával. Ennek oka részben az egyes paraméterek saját jellemzőiben keresendő (pl. az artikulációs tempó változtatása és hosszú távon annak fenntartása nehezített a beszélők számára), részben pedig a beszélők összeszokását, a témákba való involválódását mutatja. A beszélőváltásokhoz közel a tempó lassabb, a szünetek időtartama pedig hosszabb, a háttércsatorna-jelzések száma pedig szignifikánsan nagyobb volt, mint a váltásoktól távol. A lassabb artikulációs tempó utalhat arra, hogy a beszélő fokozatosan „kifogy a mondanivalójából”, így ez egyfajta jelzésként szolgál a szóátadási szándékra. A háttércsatorna-jelzések megnövekedett sűrűsége a beszélőváltások előtt jelezheti a partnerek megszólalási szándékát – eltérően a jelenség hagyományosan alkalmazott definíciójától. A gyakoribb háttércsatorna-jelzés, valamint a rövidebb szünettartások, és a lassabb artikulációs tempó mutattak együttjárást, amely alátámasztja a feltételezést, miszerint mind az egyes, beszélőváltások, mind a társalgás kezdete és vége között dinamikusan változnak a vizsgált társalgási paraméterek, ezzel segítve a beszélőket a szóátvételi pontok kijelölésében és bejósolásában.

Irodalom

[1] Sacks, H., Schegloff, E., Jefferson, G. (1974). A simplest systematic for the organization of turn-taking for conversation. Language 50, 696–735.

[2] Ford, C., Thompson S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In Schegloff, E. A., Thompson, S. A. Interaction and grammar. Cambridge: Cambridge University Press. 135–184.

[3] Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn constructional units and turns in conversation. Pragmatics, 6(3), 371–388.

[4] de Ruiter, J. P., Mitterer, H., Enfield, N. J. (2006). Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. Language 82. 515–535.

[5] Couper-Kuhlen, E., Selting, M. (1996). Prosody in conversation. Cambridge: Cambridge University Press.

[6] Levinson, S. C., Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. Frontiers in Psychology. 6(731).

[7] Gráczi, T. E., Bata, S. (2010). The effect of familiarization on temporal aspects of turn-taking: a pilot study. Acta Linguistica Hungarica, 57/2–3, 307–328.

[8] Markó A., Gósy M. (2015). A megszólalás stratégiái társalgásban. In Bárdosi V. (szerk.) A nyelvi pragmatika kérdései szinkrón és diakrón megközelítésben. 159–168.

[9] Horváth, V., Krepsz, V., Gyarmathy, D., Hámori, Á., Bóna, J., Dér, Cs. I., Weidl, Zs. (2019). Háromfős társalgások annotálása a BEA-adatbázisban: elvek és kihívások [The annotation of three-participated conversations in Hungarian Spontaneous Speech Database: principles and challenges]. Nyelvtudományi Közlemények 115. 255–274.

[10] Hutchby, I., Woofitt, R. (2006). Conversation analysis: principles, practices and applications. Cambridge: Polity Press.

A kutatást a Nemzeti, Kutatási, Fejlesztési és Innovációs Hivatal K-128810 számú pályázata és a Bolyai János Kutatási Ösztöndíj támogatta.

A ritmusérzék, a fonológiai tudatosság és az olvasás kapcsolata iskolakezdő gyerekeknél

Kertész Csaba [1] és Honbolygó Ferenc[2]
[1] ELTE PPK Pszichológiai Doktori Iskola, Budapest
[2] Természettudományi Kutatóközpont, Agyi Képalkotó Központ, Budapestt;
ELTE PPK Pszichológiai Intézet, Budapest

A ritmikai és a nyelvi, illetve olvasási képességek közötti kapcsolatot a gyakorlati megfigyeléseken túl számos kutatás eredménye látszik alátámasztani. Ezt a kapcsolatot megfigyelték tipikus [3, 6] és atipikus nyelvi fejlődésű, elsősorban diszlexiával [1, 4], illetve specifikus nyelvi zavarral küzdő gyerekek körében egyaránt [2].

Az óvodás, illetve iskolakezdő gyerekekkel végzett ritmusészleléses, ritmusreprodukciós és szinkronizációs feladatok számos esetben bizonyultak jó előrejelzőnek az olvasás későbbi színvonala, illetve a fonológiai tudatosság tekintetében [5, 12]. Ozernov-Palchik és Patel [13] metaanalízisükben a ritmikai képességek és az olvasás kapcsolatát vizsgáló kutatásokat elemezve megállapították, hogy az elsősorban a tempóval kapcsolatos képességet mérő („beat-based") tesztek járnak együtt stabilan az olvasás színvonalával. Ugyanakkor a magyar anyanyelvű gyerekeket vizsgáló munkák száma ez idáig csekély, különösen a nemzetközi szakirodalomban elterjedt szinkronizációs feladatok esetében [10].

Jelen kutatásban 39 tipikus fejlődésű, 6-7 éves iskolakezdő gyerekkel végeztünk különböző ritmikai teszteket: szenzomotoros szinkronizációs (sensorimotor synchronization, SMS), spontán motoros tempó (spontaneous motor tempo, SMT), valamint ritmusreprodukciós feladatokat tanévkezdéskor. Az SMS feladat során a vizsgálati személynek valamilyen ritmikus, tempóval rendelkező hangingerrel (általában metronóm) kell a mozgását (általában kopogás) szinkronba hoznia, majd a hang megszűnése után a felvet tempót megtartania. Az ütéseket a referenciaértékektől való távolságuk, belső konzisztenciájuk, illetve a kezdőtempótól való eltávolodás alapján elemezzük. Az SMT feladatban nem használunk referenciaként hangingert, mindössze arra kérjük a vizsgálati személyt, hogy számára kényelmes tempóban kopogjon meghatározott ideig. Ekkor a spontán módon kialakított tempó nagyságát, annak változását és az ütések belső konzisztenciáját (szórás) vizsgáljuk [14].

Az SMS és SMT feladatokhoz digitális MIDI eszközt használtunk, valamint az SMS feladatban komplex zenés ingeranyagot alkalmaztunk, amely a vizsgált korosztály motivált feladatvégzése szempontjából ideálisnak bizonyult. Az ingeranyag három különböző tempójú (80, 120, 150 bpm) instrumentális zenei részletből állt, amelyekhez a gyerekeknek szinkronban kellett kopogniuk, majd a zene végeztével megtartani a felvett tempót.

A ritmusreprodukciós feladat az ismert "visszatapsolásos" módszerrel történt, amely sikerességét szakértők – diplomás zeneművészek és pedagógusok – pontozták megadott kritériumok alapján.
A tanév végén felmértük a tanulókat a fonológiai tudatosság, valamint az olvasás területein a Fonológiai Tudatosság Teszt [8], valamint a korosztálynak megfelelő Meixner olvasólap [15] segítségével. A FTT öt szubtesztetjét (rímtalálás, szótagolás, hangszintézis, hasszú hang megnevezés és hangmanipuláció), a Meixner olvasólapnak pedig a szóolvasás feladatát

használtuk. Az olvasás színvonala (hibák száma és fluencia) és a fonológiai tudatosság egyaránt szignifikáns korrelációt mutattak a szinkronizációs és a spontán tempó feladat több mutatójával (1. táblázat), ugyanakkor a ritmusreprodukciós feladat nem rendelkezett hasonló prediktív erővel.

Az SMS feladatok folytatásos (hang nélküli) szakaszában mért eltávolodás a kezdőtempótól („Eltávolodás") a Fonológiai Tudatosság Teszt összpontszámával (r = 0,53), a szóolvasási hibák számával (r = -0,41) és fluencia mutatóval (r = -0,33) is szignifikáns korrelációt mutatott. A spontán tempó nagysága szintén a szóolvasási hibák számával (r = 0,54), a fluenciával (r = 0,45), valamint az FTT hosszú hang megnevezés feladatának pontszámával (r = -0,38) állt kapcsolatban. Összességében megállapítható, hogy az alacsony aszinkronitás, kisebb mértékű inkonzisztencia és eltávolodás a kezdő– vagy felvett tempótól és az alacsonyabb spontán motoros tempó járt együtt a fonológiai tudatosság és az olvasás magasabb színvonalával.

Bár a minta nagysága messzemenő következtetések levonására nem alkalmas, az eredmények egy irányba mutatnak a külföldi szakirodalomban találhatóakkal, és felhívják a figyelmet a korai ritmikai fejlesztés fontosságára, valamint a szenzomotoros szinkronizációs és spontán motoros tempó feladatok esetleges alkalmazhatóságára a nyelvi és olvasási nehézségek előrejelzésében. További kérdésként merül fel, hogy a ritmikai, illetve nyelvi teljesítmény kapcsolatának hátterében olyan területáltalános tényezők állnak, mint a figyelemirányítás [9], vagy a végrehajtó funkciók [11], vagy olyan területspecifikusak, például a hallási észlelés [7].

1. táblázat. A fonológiai tudatosság és a szóolvasás pontszámainak korrelációi a ritmikai szinkronizációs feladatok mutatóival

Változók	Abszolút aszinkronitás	Eltávolodás	Folytatásos inkonzisztencia	SMT	SMT inkonzisztencia	Ritmusrep.
Szótagolás[a]	,30+	**,47****	**-,50****	-,29+	-,06	,17
Rímtalálás	,13	,29+	-,15	,04	**-,33***	-,10
Hangszintézis	,04	,17	-,14	,06	,01	,26
Hosszú hang megnevezés	-,13	**,35***	**-,42****	**-,38***	-,18	,24
Hangmanipuláció	,20	**,35***	-,19	-,30+	-,19	,06
FT összpontszám	,12	**,53****	**-,40***	-,29+	-,23	,18
Szóolvasás hiba	-,26	**-,41***	,18	**,54****	-,05	-,00
Szóolvasás fluencia	**-,40***	**-,33***	,21	**,45****	-,05	,13

Irodalom

[1] Colling, L. J., Noble, H. L., & Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An EEG study. *Frontiers in Neuroscience*, *11*(JUL). https://doi.org/10.3389/fnins.2017.00360

[2] Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, *45*(1), 119–130. https://doi.org/10.1016/j.cortex.2007.09.008

[3] David, D., Wade-Woolley, L., Kirby, J. R., & Smithrim, K. (2007). Rhythm and reading development in school-age children: A longitudinal study. *Journal of Research in Reading*, *30*(2), 169–183. https://doi.org/10.1111/j.1467-9817.2006.00323.x

[4] Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS ONE*, *10*(9), 1–17. https://doi.org/10.1371/journal.pone.0138715

[5] Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., Monasta, L., Montico, M., Sila, A., Ronfani, L., & Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*, 392. https://doi.org/10.3389/fnhum.2014.00392

[6] Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin Mcauley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*(4), 635–644. https://doi.org/10.1111/desc.12230

[7] Goswami, U. (2018). A Neural Basis for Phonological Awareness? An Oscillatory Temporal-Sampling Perspective. *Current Directions in Psychological Science*, *27*(1), 56–63. https://doi.org/10.1177/0963721417727520

[8] Jordanidisz, Á. (2009). A fonológiai tudatosság fejlődése az olvasástanulás időszakában. *Anyanyelv Pedagógia*, *4*.

[9] Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119.

[10] Maróti, E., Barabás, E., Deszpot, G., Farnadi, T., Norbert Nemes, L., Szirányi, B., & Honbolygó, F. (2019). Does moving to the music make you smarter? The relation of sensorimotor entrainment to cognitive, linguistic, musical, and social skills. *Psychology of Music*, *47*(5), 663–679. https://doi.org/10.1177/0305735618778765

[11] Moreno, S., & Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research*, *308*, 84–97. https://doi.org/10.1016/j.heares.2013.09.012

[12] Moritz, C., Yampolsky, S., Papadelis, G., Thomson, J., & Wolf, M. (2013). Links between early rhythm skills, musical training, and phonological awareness. *Reading and Writing*, *26*(5), 739–769. https://doi.org/10.1007/s11145-012-9389-0

[13] Ozernov-Palchik, O., & Patel, A. D. (2018). Musical rhythm and reading development: does beat processing matter. *Ann. NY Acad. Sci*, *1423*, 166–175.

[14] Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. Psychonomic bulletin & review, 12(6), 969-992.

[15] Sipos, Z. (2015). A 3. évfolyamosok olvasásának vizsgálatára kidolgozott Meixner-olvasólap sztenderdizálásának első eredményei. https://docplayer.hu/105487901-A-3-evfolyamosok-olvasasanak-vizsgalatara-kidolgozott-meixner-olvasolap-sztenderdizalasanak-elso-eredmenyei.html

Lingual coarticulation in voiced and voiceless postalveolar fricatives in cochlear implant users: EPG evidence from Croatian

Dora Kolarić[1] and Marko Liker[1]
[1] Department of Phonetics, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia

There are several reasons for investigating coarticulation in atypical speech - it can reveal new insights into typical sensorimotor behaviour, enhance existing coarticulatory models and theories and finally, improve the diagnosis and the treatment of atypical speech [4]. Researchers have so far focused on several groups of patients and hearing-impaired persons are one of the most frequently targeted groups in this respect [4, 7]. The common finding is that prelingually deaf persons with cochlear implants (CI) might have different gestural organization from persons with typical hearing (TH). The nature and details of that difference is still largely unknown and further research is needed. Lingual aspects of gestural organization are the most important aspects of speech production and electropalatography (EPG) is currently the only available technique specifically designed to investigate linguopalatal contact patterns [2, 3].

Therefore, in this investigation we use EPG to investigate lingual coarticulation in voiced and voiceless postalveolar fricatives /ʃ/ and /ʒ/ in prelingually deaf CI users. We shall compare the results with previously published findings on TH speakers of Croatian [5, 6].

It is our hypothesis that analysing coarticulation in speech sounds which require complex gestural organization and auditory control will reveal some differences in the nature of gestural organization between CI and TH persons. Research into speech of CI users is challenging in several important aspects – small and heterogeneous groups, which are hard to identify and recruit. This investigation is no different in this respect, so we treat each speaker as a separate experiment.

Three prelingually deaf CI users participated in this investigation. All three participants were female and they were native speakers of Croatian, aged 19 (I1), 23 (I2) and 27 (I3). They differed with respect to age at implantation (I1=3; I2=5, I3=10) and age at rehabilitation (I1=2; I2=5.7; i3=3.6) and speech audiometry results (I1=90% at 45 dB, i2=70% at 65Db, I3=60% at 50dB). Target sounds were transcribed by broad phonetic transcription by the two authors independently. The agreement in judgements was well above 90% and all cases in which the judgements differed were discussed and re-transcribed. Speech material was collected by means of a modified map-task experiment whereby quasi-spontaneous speech was elicited. In such a task speakers repeated each fricative (/ʃ/ and /ʒ/) four times in three symmetrical vowel contexts (/ii/, /aa/, /uu/). Recording, annotation and data analysis were performed via Articulate Assistant software ([1] Articulate Instruments Ltd., 2008), while statistical analysis and data visualization were done via MS Excel. Four EPG parameters were measured (CoG, amount of contact, central groove, fricative duration) and they were analysed at temporal midpoint of the fricative (in order to analyse overall articulatory configuration) and at 5 equally spaced sample points throughout the fricative (in order to analyse articulatory dynamics in normalized duration).

The results showed that there are observable inter-speaker differences in the way they articulate and coarticulate voiced as opposed voiceless postalveolar fricatives. These differences and their relation to participants' hearing status will be discussed Lingual correlates of voicing difference are similar to those reported for TH persons when measured at fricative mid-point, but observable difference between CI and TH persons can be seen when articulatory dynamics is taken into account.

Key words: cochlear implant (CI), coarticulation, fricatives, electropalatography (EPG)

References

[1] Articulate Instruments (2008). Articulate Assistant User Guide, ver. 1.17. Musselburgh: Queen Margaret University.

[2] Gibbon, F. E. & K. Nicolaidis (1999). Palatography.In: Coarticulation: theory, data and techniques. Ed by Hardcastle, W. J. & N. Hewlet. Cambridge: Cambridge University Press, pp 229-245.

[3] Gibbon, F. E. 2008. Instrumental analysis of articulation in speech impairment. In: The handbook of clinical linguistics. Ed by Ball, M. J., Perkins, M. R., N. Müller & S. Howard Oxford: Blackwell publishing. pp 311-331.

[4] Hardcastle, B. & K. Tjaden (2008). Coarticulation and Speech Impairment. The handbook of clinical linguistics, 506-524. Blackwell Publishing.

[5] Liker, M. & F. E. Gibbon (2013). Differences in EPG contact dynamics between voiced and voiceless lingual fricatives. Journal of the International Phonetic Association, 43(1), 49-64.

[6] Liker, M. & F. Gibbon (2011). Groove Width in Croatian Voiced and Voiceless Postalveolar Fricatives. In ICPhS, 1238-1241.

[7] Pratt, S. R. & N Tye-Murray (1997). Speech impairment secondary to hearing loss. In: Clinical Management of Sensorimotor Speech Disorders. Ed by M. R. McNeil, New York: Thieme, pp. 345–87

Az élekor szerepe a verbális mondatemlékezet működésében

Laczkó Mária
ÉSZC Eötvös József Technikum

A pontos hallás utáni mondatismétlésben lényeges a mondat fogalmi jelentésének reprezentációja, az előhívott lexikai szerkezetek/lexémák jellemzői, a mondatszerkezet, a munkamemória működése [1, 5, 10]. Közülük meghatározó a rövid idejű emlékezet. Baddley modellje alapján a „nyelvelsajátítás motorjaként" definiálható, hiszen működése szorosan összefügg az anyanyelv-elsajátítás ütemével, a lexikális, morfoszintaktikai fejlődéssel és az olvasással [7, 9]. Más megfogalmazásban: a nyelvi folyamatokban a kialakuló nyelvi reprezentáció színtere [4], ami a beérkező nyelvi anyagtól függően folyamatosan módosul és játszik szerepet a szókincs alakulásában, a morfoszintaktikai fejlődésben, a mondatértésben és a szövegértésben [8, 7]. A Baddley-féle munkamemória legfontosabb része a beszédalapú információk megtartásáért felelős fonológiai hurok [1, 3]. Kapacitásbeli különbségei meghatározzák a hosszú távú emlékezeti reprezentációkat, ami befolyásolja a mondatismétlés sikerességében szerepet játszó faktorokat, a mondatok terjedelemét szemantikai, szintaktikai szerkezetét. A pontos mondatismétlés tehát kulcsfontosságú az anyanyelvre épülő folyamatokban, mint például a tanulás, s az iskolai kommunikáció különböző formáiban, így az írásbeli kifejezőkészség, a vázlatkészítés és jegyzetelés terén is lényeges lehet a szerepe.

A mondatismétlés felmérése az iskoláskorú gyermekek esetében diagnosztikus értékű. Utalhat a tanulók beszédészlelési működésre, szókincsének alakulására, a morfoszintaktikai fejlődésre, s így jelezheti beszédfeldolgozási és beszédprodukciós szintjüket, várható iskolai előmenetelüket, feltételezhető mondatértési és szövegértési teljesítményüket.

Korábbi kutatások azt mutatták, hogy az óvodások, alsó tagozatosok és felső tagozatosok mondatismétlései [4, 2] az életkor előrehaladtával egyre pontosabbak voltak, de az idősebb tanulók sem nyújtottak hibátlan teljesítményt. Középiskolai tanulók teljesítményére hatással volt az iskolatípus is [6].

A jelen kutatás központi kérdése az, miképpen alakul különböző életkorú általános iskolai tanulók verbális mondatemlékezete. Feltételezem, hogy az életkorral lineáris változás mutatkozik a mondatismétlésben, tehát minél fiatalabb a gyermek, annál gyengébb eredményt mutat. Kérdés, hogy az egyes életkorokban mekkorák a különbségek és ezek minőségi vagy mennyiségi eltérések-e. Feltételezem, hogy a mondatok helyes ismétlésében a mondatok szerkezeti és tartalmi jellemzői is tükröződnek és összefüggést mutatnak az életkorral.

A kutatásban 2. 4. 6. és 8. osztályosok vettek részt, minden korcsoportban 10-10 tanuló. A mondatismétlési feladat 20 általam összeállított mondatból áll (70% egyszerű, 30% összetett mondat). Egyszerű mondat például: *Már fellobbant az olimpiai láng a sportesemények helyszínén.* Összetett mondat például*: Hallani, amint megzizzen az őszi avar a lábad alatt.* A tesztanyag a modalitás alapján többségében kijelentő mondatokból áll, de van két kérdő és három felszólító mondat is közöttük. Például: *Melyik terminálról fog indulni a repülőgépetek? Szedjétek össze a szemetet a földről!*

A teszt mondatainak összeállításakor figyelembe vettem továbbá a mondatok terjedelmét, amit a szó és a szótagszám alapján egyaránt számítottam, valamint a mondatok szerkezeti felépítését, a mondatokat alkotó szavak feltételezhető gyakoriságát, stílusértékét [6].

A mondatok összesen 154 szót tartalmaznak. A mondatokat felépítő szavak száma 5 és 11, a szótagoké 12 és 31 közötti. Az átlagos mondathosszúság 6,75 szó/mondat, illetve 16,4 szótag/mondat. A legrövidebb mondatban 5 szó fordult elő (*Kifürkészhetetlen titkokat őriznek még mindig.*), a leghosszabban 11 (*Minden este bömböl a rádió és üvölt a zene a szomszédban*). Mivel a legkevesebb szót tartalmazó mondat nem feltétlen a legrövidebb, a szótagszám alapján is megadom a terjedelmi határokat. A legrövidebb mondat 12 szótagos (*Szedjétek össze a szemetet a földről*!), a leghosszabb 31. (*A seregély óvatosan belopakodott a szőlőlugasba, ahol megdézsmálta a szőlőfürtöket*.) (Az utóbbi a leghosszabb szótagszámú mondat, ebből mindössze egy van a tesztanyagban.)

A tanulók feladata a mondatok egyszeri elhangzása után azok pontos ismétlése volt. Az elemzés kiterjedt a helyes/helytelen ismétlések arányára. Elemeztem, hogy a mondatok helyességét a mondatok szerkezete és tartalma miképpen befolyásolja. A helyesen visszamondott mondatokban megvizsgáltam az előforduló néma szünetek gyakoriságát és időtartamát. A hibásan ismételt mondatokban kategorizáltam a hibatípusokat (grammatikai hiba, csere, kihagyás, betoldás, metatézis, módosított szó, mondatszintű hibák (kihagyás és átalakítás)). Elemeztem a mondatok elhangzása és mondatok ismétlésének kezdete közötti időtartamot (reakcióidőt) is.

Az elemzések eredményei egyértelműen igazolták az életkor meghatározó szerepét a mondatismétlésben, s így az eredmények korreláltak a szakirodalmi adatokkal, noha az egymáshoz közel álló korcsoportok között nem volt minden esetben szignifikáns eltérés. Ugyanakkor valamennyi vizsgált szempont alapján jól látszódott az alsó és a felső tagozatosok közötti mennyiségi és minőségbeli különbség. Így a helyesen visszamondott mondatok aránya a két alsó tagozatban közel azonos volt. Hasonlóan közeli értékeket kaptunk a két felső tagozatos korcsoportban is, ám náluk a helyesen visszamondott mondatok aránya másfélszer annyi, mint az alsó tagozatosoké. A mondatszerkezet hatása valamennyi korcsoportban azonos tendenciát mutatott: az egyszerű mondatok ismétlése sokkal pontosabb és eredményesebb volt, mint az összetett mondatoké, s a bár az életkorral haladva lineáris növekedés volt tapasztalható mindkét mondattípus esetén, az alsó és a felső tagozatosok eredményében szignifikáns eltérés az összetett mondatokban volt követhető. A hibatípusok osztályozásakor életkortól függetlenül a csere volt a leggyakoribb, a betoldás a legritkább, a többi hibatípus az egyes életkorokban eltérő mintázatot mutatott, némelyik az életkor függvényében alakult. Jelentős különbségek mutatkoztak az életkor szerinti reakcióidőkben is.

Az előadás részletesen mutatja be a kapott eredményeket a vizsgált paraméterek mentén, rámutatva nemcsak a vizsgálat nyelvészeti hozadékára, de annak pedagógiai jelentőségére is, hangsúlyozva azokat a tananyagba illeszthető feladatokat, amelyek a munkamemória fejlesztését segítik.

Irodalom

[1] Baddeley, A. (2005). *Az emberi emlékezet.* Osiris Kiadó. Budapest.

[2] Bóna J. (2016). ’Hallásalapú és vizuális közlések feldolgozása 3-7.osztályos korban’. *Anyanyelv-pedagógia*, 2016/4. http://www.anyanyelv-pedagogia.hu/cikkek.php?id=650

[3] Eysenck, M. W. & M. T. Keane (2003). *Kognitív pszichológia*. Hallgatói kézikönyv. Budapest. Nemzeti Tankönyvkiadó.

[4] Fejes A. (2016). *Mondatészlelés az életkorok függvényében*. In: Balázs G. – Veszelszki Á. (szerk.) Generációk nyelve. Tanulmánykötet. ELTE BTK Mai Magyar Nyelvi Tanszék. Budapest. 63–72.

[5] Jefferies, E., M. A. L. Ralph, & A. D. Baddeley, (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language* 51/4, 623–643.

[6] Laczkó M. (2019). A verbális mondatemlékezet sajátosságai középiskolás korban. *Anyanyelv kultúraközvetítés* 2/1, 43–61.

[7] Mohai K. & Szabó Cs. (2017). 'A munkamemória vizsgálata'. https://epa.oszk.hu/03000/03047/00065/pdf/EPA03047_gyosze_2014_3_226-232.pdf

[8] Németh D. (2002). 'Munkamemória, fejlődés, nyelv'. In: Racsmány M. & Kéri Sz. (szerk.): Architektúra és patológia a megismerésben. Books in Print Kiadó, Budapest.

[9] Racsmány M., Lukács Á., Németh D. & Pléh Cs. (2005). A verbális munkamemória magyar nyelvű vizsgálóeljárásai. Magyar Pszichológiai Szemle, 4, 479–505.

[10] Tsiamtsiouris, J. & H. Smith Cairns (2013). Effects of sentence-structure complexity on speech initiation time and disfluency. *Journal of Fluency Disorders* 38/1, 30–44.

Primary functions in infant-directed speech and their longitudinal development

Katalin Mády[1], Uwe D. Reichel[1], Anna Kohári[1], Andrea Deme[2,3], Ádám Szalontai[1]

[1]Hungarian Research Institute for Linguistics
[2]Eötvös Loránd University
[3]MTA–ELTE Lendület Lingual Articulation Research Group

Introduction. In most cultures, caregivers alter their speech when they talk to children. Traditionally, differences have been attested to two main functions: (1) expression of positive emotions such as love, comfort, attention and approval, and (2) exaggerating linguistic forms in order to aid language acquisition. Findings indicate a longitudinal change with respect to characteristic cues both in caregivers' speech and in infants' responses [4]. It seems that in the preverbal age of the infant, acoustic cues are primarily used for expressing emotions, while language-relevant cues gain importance with the increasing age of the infant. Importantly, the production of cues related to infant-directed speech (IDS) and infants' responses are strongly interconnected: infants show more attention to IDS, especially before the age of 6 months, and mothers partly adapt their linguistic behaviour according to the infants' needs.

This paper analyses potential changes in mothers' speech during the first 18 months of the infant. We investigate whether acoustic features traditionally associated with expressing positive emotions and guiding language acquisition change throughout this period. Two types of measures were selected: higher fundamental frequency (f0) and higher energy have been shown to signalise positive emotions, while larger vowel space is often linked to hyper-articulation and thus aiding language acquisition [4]. However, the presence of these features is not associated with exclusive functions, instead, they seem to serve to gain the infants' attention and thus enhance communication between caregivers and children.

Methods and materials. 16 mothers and their infants participated in the study. All mothers were native speakers of Hungarian without speech or language impairment, and all babies were raised in a monolingual Hungarian environment. Mothers were given a colourful booklet with images of a hide-and-seek story of four pixies. The story was presented partly only in pictures, partly also in form of written utterances (17 sentences). Mothers were asked to tell the story first to the experimenter, then to their own baby in their own words, but replicating the fixed utterances if present. Recordings took place at 0, 4, 8 and 18 months of the baby. Recordings with 0 months were carried out at the Birth Centre of the Military Hospital in Budapest, those between 4 and 18 months in the baby lab of RCNS HAS. Acoustic analysis was based on altogether 2111 utterances at four time points in tow registers: adult-directed (AD) and ID speech.

Prosodic features that had been shown to participate in the expression of positive emotions [4, 5] were investigated, such as f0 mean (m), maximum (max) and standard deviation (sd), accompanied by energy (en) m, max and sd in each target sentence. Parameters were extracted with the CoPaSul tool [3] that normalises f0 (semitones) to speakers and energy (root mean square, RMS) to local context. One feature that serves to enhance language learning of the infant is the size of the vowel space. Vowel formant dispersion (vfd) was

calculated similarly to [1] for each speaker as the mean Euclidean distance in the F1–F2 space for all vowels to the formant centroid in Bark. [1] found that this metrics was well correlated with speech intelligibility.

Results. The 7 parameters f0_m|max|sd, en_m|max|sd and vfd were first analysed for the pooled data (all utterances of all speakers in both registers). Linear mixed-effect models with random slopes (fixed effect: register ADS vs. IDS, random effect: speaker) were fitted to the data in R. P-values were calculated using the `Anova()` function in the `car` package. All parameters had significantly higher values in IDS ($p < 0.01$), except for f0_sd ($p = 0.98$). Thus, IDS in Hungarian by and large shows the same acoustic features as in other languages.

Longitudinal changes are described without using statistical models in order to detect general trends in this exploratory study. Two general tendencies were found. First, IDS at 0 months, i.e. with newborn babies differed clearly from the other three time points. All values (mean, maximum for f0 and energy along with vfd) were lower in IDS than between 4 and 18 months of the infant, except for f0_sd that did not increase within the analysed period. Differences between ADS and IDS with 0 months were less pronounced altogether (see Fig. 1), especially for vfd that did not show any hyperarticulation with newborn babies (Fig. 2). Second, none of the investigated measures for f0, energy or vowel space showed a clear trend between the age of 4 and 18 months. This means that they do not reflect a functional change in IDS going from emotion expression to the highlighting of linguistic features during the first 18 months.

Conclusions. Higher f0 and energy and an enlarged vowel space were observed in Hungarian IDS, similarly to other languages. The difference between IDS and ADS was present to a lesser extent with newborns, becoming more distinct and stable with the growing age of the infant between 4 and 18 months. However, several issues need to be raised. It is not clear whether IDS with 0 months is different due to missing communication experience of mothers with their newborn baby or to external factors (e.g. fatigue after giving birth, the location being a hospital instead of a baby lab etc.). A comparison with 22 multipara mothers of newborn babies giving birth to their second or third child supplied slight evidence for the lack of entrainment between mothers and their babies [2]: multipara mothers used more IDS-like prosodic features than primipara ones in the same experimental setting.

The tendency described in [4] for f0 to increase from birth and decrease until the age of two years is not reflected in our data. Values for all investigated parameters remained stable between 4 and 18 months of the infant. It is important to recall that the role of prosodic features cannot be reduced to the function of signalising positive emotions. They also serve to gain the infant's attention, and their presence has been shown to actually trigger more intensive reactions of the infant [4].

Acknowledgements. This work was funded by the NRDI grants 115385 and 135038. The fourth author was supported by a Bolyai János Research Scholarschip of HAS, the ÚNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology (MIT) from ressources of the NRDIF and the Thematic Excellence Program of the MIT.

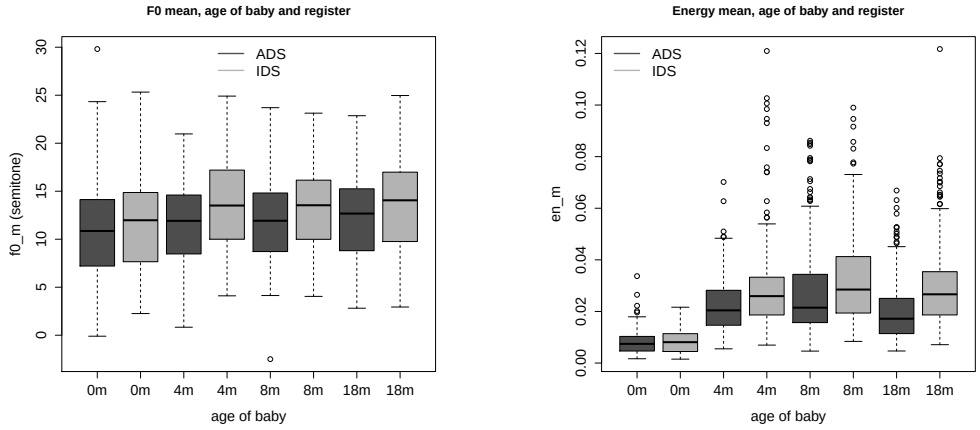


Figure 1: Adult- and infant directed speech at the age of 0, 4, 8 and 18 months of the infant. Left: f0 median, right: energy median.

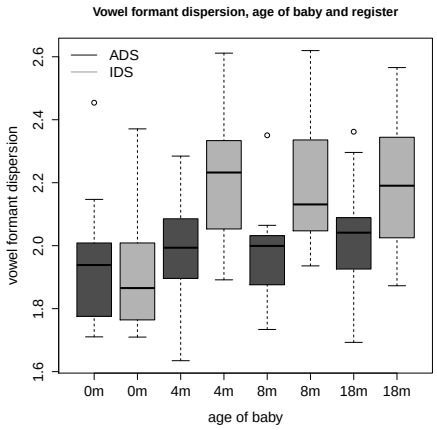


Figure 2: Adult- and infant directed speech at the age of 0, 4, 8 and 18 months of the infant: vowel formant dispersion.

References

[1] Bradlow, A., G. Torretta & D. Pisoni (1996). 'Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics'. In: *Speech Communication*.

[2] Mády, K. (2018). In: *Proc. Speech Prosody, Poznań*.

[3] Reichel, U. D. (2017). *CoPaSul Manual – Contour-based parametric and superpositional intonation stylization*. https://arxiv.org/abs/1612.04765. RIL, MTA. Budapest, Hungary.

[4] Saint-Georges, C. et al. (2013). 'Motherese in interaction: at the cross-road of emotion and cognition? (A systematic review)'. In: *PLoS ONE*.

[5] Trainor, L. J., C. M. Austin & R. N. Desjardins (2000). 'Is infant-directed speech prosody a result of the vocal expression of emotion?' In: *Psychological Science*.

Magánhangzók eltérése a hangsúly függvényében – artikulációs és akusztikai adatok

Markó Alexandra[1,2], Bartók Márton[1,2], Csapó Tamás Gábor[2,3], Gráczi Tekla Etelka[2,4] és Deme Andrea[1,2]
[1] ELTE Eötvös Loránd Tudományegyetem, Budapest
[2] MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport, Budapest
[3] Budapest Műszaki és Gazdaságtudományi Egyetem, Budapest
[4] Nyelvtudományi Intézet, Budapest

A nemzetközi szakirodalom szerint a hangsúlyos és a hangsúlytalan helyzetben megjelenő magánhangzók eltérnek. Ez az eltérés artikulációs szempontból a szonoritás (vagy hangzósság) növelésében és/vagy ún. lokális túlartikulálásban (hyperarticulation) érhető tetten, azaz abban, hogy hangsúlyos helyzetben az artikuláció jobban megközelíti az artikulációs célt [5].

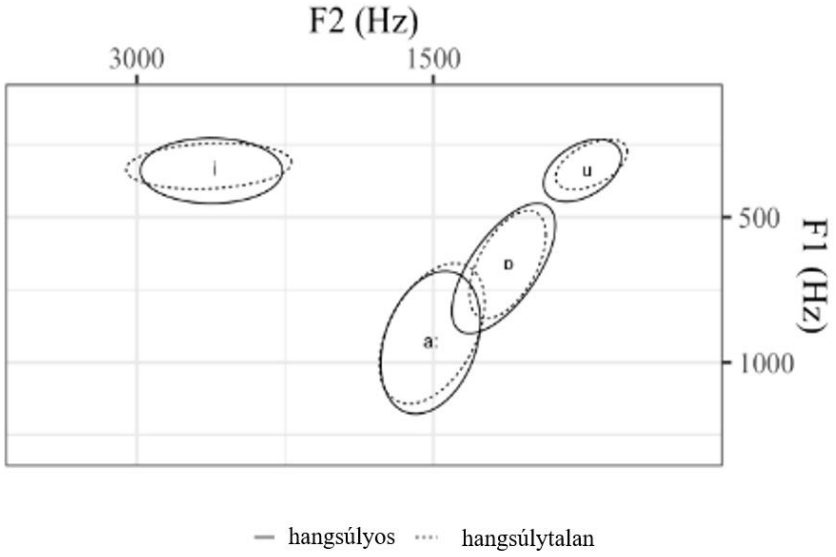
A túlartikulálás voltaképpen a magánhangzótér széleihez közelítő, azaz periferikusabb ejtést jelent, ennélfogva a túlartikulálás értelmezése a magánhangzók esetében az artikulációs célnak a magánhangzótéren belül elfoglalt helyzetétől függ. Az alsó és legalsó nyelvállású magánhangzók esetében a túlartikulálás alacsonyabb nyelvhelyzetet jelent a hangsúlyos helyzetben a hangsúlytalanhoz képest, míg a felső nyelvállásúaknál magasabbat. A hangsúlyos elülső magánhangzók esetében a szájüregi térben előrébb helyezett nyelvhelyzet jelenti a túlartikulálást a hangsúlytalanhoz képest, a hátulsóknál pedig ugyanezt a hátrébb húzott nyelv jelenti [3-5]. A szonoritás növelése az ajaknyílás növelésével (elsősorban az alsó nyelvállású magánhangzók esetében), illetve a szegmentumok időtartam-növekedésével érhető el [5]. A jelen kutatás kérdése az volt, hogy artikulációs és/vagy az akusztikai adatok utalnak-e akár lokális hiperartikulációra, akár megnövelt szonoritásra a hangsúlyos szótagokban a hangsúlytalanokhoz képest a magyarban.

A kísérletben négy szótagos értelmetlen hangsorokat vizsgáltunk önálló megnyilatkozásként, amelyben minden szótag CV szerkezetű volt, és azonos szegmentumokat tartalmazott. A mássalhangzó mindig a /p/ volt, a magánhangzó pedig az /u/, /i/, /aː/, illetve /ɒ/ valamelyike. Egymást követően ejtett hangsúlyos (mind mondat-, mind szóhangsúlyos, első szótagbeli) és hangsúlytalan (második szótagbeli) magánhangzókat vetettünk össze. Minden logatomból (legalább) 6-ot rögzítettünk beszélőnként, random sorrendben, disztraktorok között. Az artikulációs vizsgálathoz elektromágneses artikulometriát alkalmaztunk. Az akusztikai jelet a szájzugba helyezett omnidirekcionális mikrofonnal vettük fel, szinkronban az artikulációs méréssel. Kilenc beszélő adatait elemeztük.

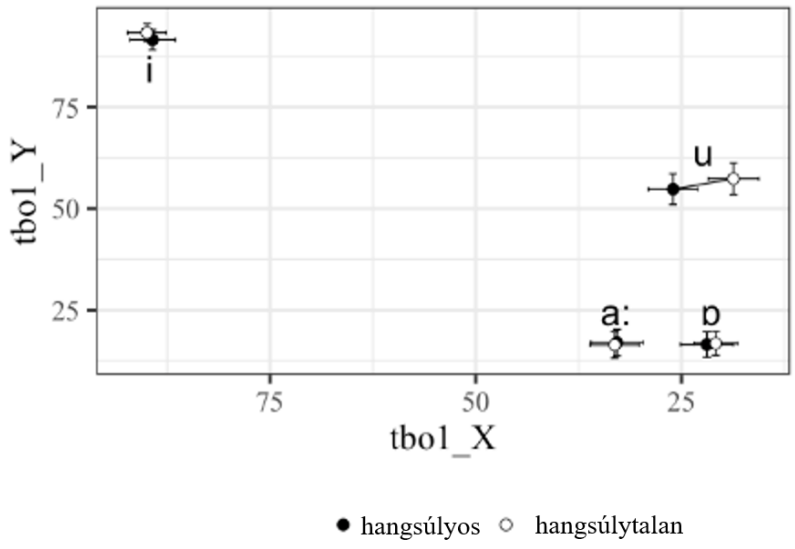
Meghatároztuk és elemeztük az összes célszóbeli magánhangzó időtartamát, valamint f_0-, F_1- és F_2-értékét, az utóbbiak alapján a magánhangzótér közepétől mért euklideszi távolságokat is meghatároztuk (1. ábra). A nyelvhelyzet esetében az elülső nyelvháti szenzor vízszintes (x) és függőleges (y) tengelyen mért (normalizált) értékeit vettük figyelembe ([2] alapján) (2. ábra). Az ajaknyílás mértékét (Lip Aperture Index, LAI) [1] alapján számszerűsítettük a felső- és az alsóajak-szenzor euklideszi távolságával, amit beszélőnként normalizáltunk (3. ábra).

A hangsúlyos szótagban a hangsúlytalanhoz képest várható lokális hiperartikuláció létét a nyelvszenzor-adatok (2. ábra) nem támasztották alá, ugyanakkor az akusztikai magánhangzótér közepétől mért euklideszi távolságokban (1. ábra) szignifikáns volt a különbség a hangsúlyos és hangsúlytalan helyzetű magánhangzók között, a várt irányban. A szonoritás növelésében is

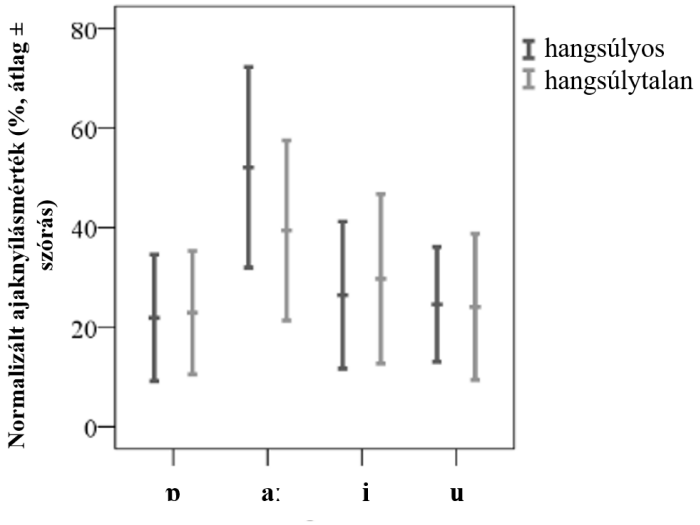
szerepet játszó ajaknyílás tekintetében egyedül az /aː/ esetében találtuk a várt tendenciát, azaz nagyobb szonoritást hangsúlyos helyzetben (3. ábra).



1. ábra. A vizsgált magánhangzók $F_1 \times F_2$ tere a hangsúlyosság függvényében



2. ábra. A vizsgált magánhangzók ejtésekor mért nyelvhelyzet a hangsúlyosság függvényében



3. ábra. A vizsgált magánhangzók ejtésekor mért ajaknyílás normalizált értéke a hangsúlyosság függvényében

Irodalom

[1] Byrd, D. (2000). ’Articulatory vowel lengthening and coordination at phrasal junctures.’ *Phonetica* 57, 3–16. old.

[2] Cho, T. (2004). ’Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English.’ *Journal of Phonetics* 32, 141–176. old.

[3] Cho, T. (2005). ’Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English.’ *Journal of the Acoustical Society of America* 117, 3867–3878. old.

[4] de Jong, K. (1995). ’The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation.’ *Journal of the Acoustical Society of America* 97, 491–504. old.

[5] Mücke, D., Grice, M. (2014). ’The effect of focus marking on supralaryngeal articulation – Is it mediated by accentuation?’ *Journal of Phonetics* 44, 47–61. old.

A kutatás a Bolyai János Kutatási Ösztöndíj, a Tématerületi Kiválósági Program és az Innovációs és Technológiai Minisztérium ÚNKP-20-5 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

How does an AI recognize speech?
– About end-to-end deep neural network based speech recognition

Péter Mihajlik[1,2]
[1] Budapest University of Technology and Economics, Hungary
[2] SpeechTex Inc., Hungary

In the past decade artificial neural networks and deep learning have revolutionized automatic speech recognition (ASR) – not to mention speech synthesis (TTS) or image recognition or even autonomous driving. We can use voice search even in a noisy environment or watch television with automatic subtitling or dictate emails. Many of us, though, have little knowledge on how these neural units work together and merely think of it as a mysterious "AI" (Artificial Intelligence) solving problems that used to be human privileged. In the presentation we are going to make an attempt to uncover the myths about the entirely neural network based – called "end-to-end" – speech recognition. We point out that this new technology is the natural continuation of the preceding several decades long speech recognition and machine learning research that focuses more and more on data instead of human analogous thinking.

One of the first automatic speech recognizers in the world was the "Shoebox" system of IBM [3] in 1961. It could recognize a few spoken words and the digits: 0 through 9. The principle of operation was to apply 3 different *frequency band filters* and then use simple logic according to the available phonetic knowledge to decide on the recognition result. This approach, however, was not viable in real life.

The first approach that had some practicality required the *recording of reference waveforms* for all vocabulary items and then applied similarity measurement between the references and the input speech signal to select the most likely one. For the comparison, waveforms were converted into time-sequential Mel-filter band feature representations and temporal variations were smoothed out by the DTW (Dynamic Time Warping) algorithm [7]. A limitation of the approach was that only isolated words could be recognized, besides, the acoustic conditions between references and test recording – including the identity of the speaker – must have been matched.

The next big step was to generalize similarity measurement towards a probabilistic framework by the introduction of Hidden Markov-Models (HMM), starting from 1975 [4]. This approach allowed a linguistically appealing hierarchical model structure: acoustic models (on the phonetic level), pronunciation models (phonological level), and language models (syntactic level). One of the most important achievements of HMM, however, was that *it used statistics to calculate acoustic similarity* (called likelihood). Instead of having separate utterance-level models for each word and for each speaker, universal phoneme-based acoustic models were built on natural language speech recordings from hundreds of speakers. The technique that captured the acoustic variability of phones (within the HMM framework) was called GMM (Gaussian Mixture Model) which modeled the probability distribution of the acoustic data of a given phone directly. All in all, HMM made large vocabulary continuous speech recognition (LVCSR) possible thanks to the statistics-based machine learning techniques applied for both low (acoustic) and high (language) levels.

The only area that still needed language specific expert knowledge was pronunciation modeling: how orthographic words can be mapped to phoneme sequences. For languages with loosely coupled orthography and phonemics – such as English or French – hand-crafted pronunciation dictionaries were unavoidable. In case of the more phonemic languages – e.g., for Slavic or Hungarian – a couple of rules could be used to generate automatically the majority of the pronunciation dictionary elements. However, all the manually derived rules or dictionaries suffered from an inherent inconsequence (due to multiple editors, for example) and erroneous forms (e.g. from typos). Moreover, the phoneme inventory sets applied were typically ad-hoc to some extent. Therefore, it seemed desirable to avoid their explicit usage. It turned out that *building acoustic models directly on graphemes* (letters) did not harm speech recognition performance at all for many languages, including Hungarian [5]. The "trick" was to consider the phonetic (graphemic) context – what has already been used for conventional phoneme-based approaches widely.

In the meanwhile, artificial neural networks (ANN) – built from *simple mathematical model* [6] of a real neuron – evolved and performed well in classification and prediction tasks, but gained only minor interests. Merely they were considered as replacements for GMM models in the HMM framework (called hybrid ANN-HMM approach). Theoretically, ANN should have performed better then GMM because of its discriminative ability but according to the actual computational infrastructures, training – the estimation of free model parameters – was slow, complicated and in turn the result was only marginally better than that of the GMM-based systems. Until the "deep learning revolution" in the 2010s.

Technically speaking a deep neural network (DNN) means a regular ANN with the restriction of having at least two (but possibly tens or hundreds of) "hidden" layers between the input and output layers. Note, that any layer means by default simple linear (weighted sum) operations on the input and then applying a well-defined nonlinear function to get the output. So, what has had happened resulting later in ASR systems competing human speech recognition accuracies? A lot… but essentially, the parameters of the deeper neural network structures became tunable. Thus, *artificial neural networks could be scaled up* to more and more data using deeper and more complex structures. But the basic operations – weighted sum, simple nonlinearities, various normalizations – remained. What has been changed: dramatically improved classification and recognition accuracies over the GMM (or shallow ANN) baseline and much more computational requirements, especially during training (parameter tuning).

Soon the Connectionist Temporal Classification (CTC) [2] was (re)discovered. This technique allows *training neural networks directly on acoustic and orthographic transcription data* by moving time alignment calculation outside of the network. Essentially, the same technique (forward-backward calculation) is applied as in the case of HMM-based time alignment, but now it is used only as a cost function (which tells us how well a neural net is trained). Having a large acoustic context, grapheme acoustic models can now be trained and with the introduction of a special "blanc" character, they can turn speech input directly into text – even for English. Dictionaries with or without pronunciations and language models are no longer essential components of an end-to-end neural network based ASR system, but they still can be applied as complementary modules.

The evolution of deep end-to-end speech recognition has not stopped. The introduction of the "attention mechanism" can provide time alignment without the CTC algorithm and a recurrent neural layer can now be used as an internal language model [1]. Possibly one of the main benefits of using end-to-end speech recognition is that no dedicated speech recognition software is needed beyond the deep learning framework.

Our conclusion is that end-to-end deep neural network ASR systems are natural continuation of "conventional" ASR approaches. Learning from data was the key for practical ASR. Using statistics was the first step towards general speech-to-text conversion. Modeling graphemes acoustically preceded deep learning and statistics-based machine learning has always been the core of speech-to-text conversion. The main contribution of deep neural networks is allowing more and more complex nonlinear computational structures utilizing more and more data resulting in unprecedented classification/prediction accuracies. This is indeed a big achievement, still in progress – the appearance of novel neural structures and approaches seems unstoppable. However, the basic concept of ASR has not been changed from the beginnings: machines do mechanical pattern matching and choose the most likely hypothesis – without understanding any of the words they transcribe.

References

[1] Chan, W.; Jaitly, N.; Le, Q.; and Vinyals, O. (2016), "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," IEEE - ICASSP 2016, pp. 4960-4964. DOI: 10.1109/ICASSP.2016.7472621

[2] Graves, A.; Fernández, S.; Gomez, F., Schmidhuber, J. (2006). "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks". ICML 2006, pp. 369–376, DOI: 10.1145/1143844.1143891

[3] https://www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html. – accessed on 13.11.2020

[4] Jelinek, F.; Bahl, L.; Mercer, R. (1975). "Design of a linguistic statistical decoder for the recognition of continuous speech". *IEEE Transactions on Information Theory*. 21 (3), pp. 250. DOI:10.1109/TIT.1975.1055384

[5] Mihajlik, P.; Tüske, Z.; Tarján, B.; Németh, B.; Fegyó T. (2010). "Improved recognition of spontaneous Hungarian speech - morphological and acoustic modeling techniques for a less resourced task", *IEEE Transactions on Audio, Speech, and Language Processing*, 18 (6), pp. 1588-1600, DOI: 10.1109/TASL.2009.2038807

[6] Rumelhart, D.; Hinton, G.; Williams, R. (1986). "Learning representations by back-propagating errors", *Nature*, 323 (6088): pp. 533–536, DOI:10.1038/323533a0.

[7] Sakoe, H.; Chiba, S. (1978). "Dynamic programming algorithm optimization for spoken word recognition". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 26 (1), pp. 43–49. DOI: 10.1109/tassp.1978.1163055

A Narrative Assessment Protocol alkalmazási lehetőségei magyar nyelven – 5–6 évesek által létrehozott narratívák elemzésében

Murányi Sarolta
Eötvös Loránd Tudományegyetem, Nyelvtudományi Doktori Iskola

A gyermekek nyelvfejlődésének fontos területe a történetmesélési készség [5]. Az utóbbi években a nemzetközi gyakorlatban több olyan elemző, vizsgálati módszer került bevezetésre melyek a narratív készség szintjét mérik. Ezek nem csak a gyermek narratív tudásáról, hanem kognitív és nyelvi szintjéről is releváns információt hordoznak (NAP 2010, NAP2 2019, ENNI 2006, Narrative Language Measure 2010, Test of Narrative Language 2004, Bus Story 1994).

A NAP egy angol nyelvű vizsgálati protokoll. A kutatók fő célja az volt, hogy létrehozzanak egy olyan kiegészítő mérést a meglévő nyelvi tesztek mellé, amely kifejezetten a történetmesélés módszerével (előzetesen meghallgatott történet visszamondása) méri a gyermekek nyelvi állapotát. A vizsgálatot tipikus nyelvfejlődésű, 3-6 éves gyermek részvételével tesztelték [2]. A vizsgálat fejlesztői a NAP megalkotásának előzményeként azt a felismerést hangsúlyozták, hogy annak ellenére, hogy egyértelműen kimondható, hogy a narratív készség a gyermekek nyelvfejlődésének egy kiemelten fontos területe, nagyon kevés megbízható teszt áll rendelkezésre ennek a kompetenciának az objektív mérésére [3].

Ellentétben a nemzetközi gyakorlattal a hazai vizsgálatok kevesebb figyelmet fordítanak a narratív készség vizsgálatára. Az óvodáskorú gyerekek narratíváiban, spontán történeteiben az eddigi magyar nyelvészeti kutatások főként az aktivált szókincset, a szöveghosszúságot, a grammatikai szerkezeteket vizsgálták [1, 4]. Jelen vizsgálat célja a NAP módszertanának és elemzési szempontjainak alkalmazása magyar nyelven.

A vizsgálatban 13 ötéves és 13 hatéves magyar anyanyelvű, tipikus nyelvfejlődésű óvodás vett részt (10 fiú, 16 lány). A kísérletben használt, 16 oldalas képsort a gyermekek könyv formátumban kapták meg. A mese meghallgatása és a képek megnézése után, a NAP módszertanát követve a résztvevők visszamondták a hallott mesét.

A protokoll vizsgálati lapját a magyarra fordított történethez igazítottam. A gyermekek által létrehozott narratívák elemzése, pontozása során ezt használtam. A szempontok között egyaránt szerepelnek olyanok, amelyek az elbeszélések makroszerkezetére (pl. történetegységek megléte, kezdés és befejezés) és mikroszerkezetére (pl. grammatikai szerkezet összetettsége, szóhasználat) vonatkozó vizsgálati szempontok. Az előadásban a NAP módszertanát, az adaptálás lehetőségeit és a két korosztály eredményeit ismertetem.

A NAP magyar nyelvű adaptálásához a vizsgálati módszernek a többi korosztály bevonásával történő kipróbálása újabb szempontokat és objektív mérési lehetőséget adhat.

Irodalom:

[1] Horváth Viktória (2017): Közlések grammatikai szerkesztettsége 6–9 éves gyermekek narratíváiban. Anyanyelv-pedagógia, 10. 4. sz. 5–18. URL: http://anyanyelv-pedagogia.hu/cikkek.php?id=703

[2] Justice L. M., Bowles R., Pence K., Gosse C. (2010): A scalable tool for assessing children's language abilities within a narrative context: The NAP (Narrative Assessment Protocol), Early Childhood Research Quarterly, Volume 25, Issue 2, Pages 218–234.

[3] Merritt, D.D. – Liles, B.Z. (1987): Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. Journal of Speech and Hearing Research 30. 539–552.

[4] Neuberger Tilda (2014): A spontán beszéd sajátosságai gyermekkorban. ELTE Eötvös Kiadó. Budapest.

[5] Westby, C. (1989): Assessing and remediating text comprehension problems. In: A. Kahmi, A. – Catts, H. (eds): Reading disabilities: A developmental language and reading disabilities. Boston. 259–324.

Bilingual speech production

Judit Navracsics
Pannon University

The speech production of bilinguals might be 'strange' for monolinguals, and this often results in a false judgement of monolinguals about the capability of bilinguals to speak perfectly in either of their languages. The well-known maximalist view [1] stigmatizes bilinguals by neglecting a most natural fact, namely that the bilingual person can never deactivate either of their languages. Even in the monolingual mode [1], the other language is there and ready to interfere for whatever reason. The esssence of the functionalist (wholistic) view of bilingualism and the Complementarity Principle is that there are hardly any balanced bilingual people as what makes them bilingual is the need for the alternating use of two different languages in their everyday life depending on the situation, the partner in communication, the topic, etc.

Once the two languages are always there, it is inevitable that implicit and/or explicit cross-linguistic interferences may come up, code-switches may occur, or simply disfluencies indicate that something is holding up the planning or execution phase of speech production.
In my talk, I will be discussing how the bilingual mental lexicon is structured, what memory systems contribute and are responsible for its operation. Through semantic and phonological paraphasia I will show the TOT phenomenon, the complexity of lexical access. The shared or non-shared character of the conceptual level will also be highlighted. Apart from the problems related to the declarative memory, I will also give examples of violated phrases, as a result of contamination of structures from the two languages, which demonstrate the flaws of the procedural memory.

The examples will be taken from the interviews and guided narratives carried out with 125 bilingual people, whose L1 or L2 is Hungarian.

References

[1] Grosjean, F. (2010) *Bilingual. Life and Reality.* Cambridge: Harvard University Press.

A nemlexikális *öö* hang beszélőváltásban betöltött szerepe magyar nyelvű társalgásokban

Németh Zsuzsanna
SZTE BTK Általános Nyelvészeti Tanszék
MTA-DE-SZTE Elméleti Nyelvészeti Kutatócsoport

Az *öö* hangot a magyarban széleskörűen kutatta, ill. kutatja a pszicholingvisztikai és fonetikai szakirodalom (l. pl. [4, 5, 1, 7, 8]), és a hezitációs jelenségek körébe sorolja. A hezitációs jelenségek átfogó pszicholingvisztikai vizsgálatát nyújtja [8]. [10] az *öö* társalgásban betöltött funkcióit a társalgáselemzés elméleti keretében vizsgálja, és arra a következtetésre jut, hogy az *öö* hang használata a forduló- és szekvenciaalkotás részét képezi.

Jelen előadásban meg kívánom mutatni, hogy bizonyos kontextusokban, amikor az *öö* részt vesz a beszélőváltási rendszer működésében, diskurzusszervező funkcióval is rendelkezhet. Ez motiválja, hogy az *öö* beszélőváltásban betöltött szerepét megvizsgáljam egy új, konverzációelemzéses-pragmatikai keretben. Amellett érvelek, hogy a beszélőváltásra alkalmas helyeken megjelenő *öö* szerepet játszhat a beszélőváltásban. [2] jellemzi a diskurzusjelölők egy, a megnyilatkozás szintaktikai szerkezetétől független csoportját. Ezek a diskurzusjelölők a kommunikációs helyzetek összekapcsolásában játszanak szerepet, lehetnek nemlexikálisak is, és érinthetik a társalgás szekvenciális szerkezetét ([3]). Ide tartoznak azok a diskurzusjelölők, amelyek a beszélőváltási rendszer szabályozásában vesznek részt ([2]). Célom annak megmutatása, hogy az *öö* részt vehet a beszélőváltási rendszer szabályozásában, és így bizonyos előfordulásai értelmezhetőek diskurzusjelölőként is.

A konverzációelemzés (társalgáselemzés) a társas interakciót annak természetes szerveződésében tanulmányozza, természetes, verbális interakciókat vizsgál hang- és videofelvételek segítségével ([12, 9]). A beszélőváltás normatív szabályait [11] tárta fel és írta le. Eszerint a beszélőváltást a résztvevők maguk irányítják: minden beszélőváltásra alkalmas helyen ugyanaz a szabálysor lép életbe ugyanabban a sorrendben. Ha a beszélő kiválasztja a következő beszélőt, akkor a jelenlegi beszélőnek be kell fejeznie a beszédet, a kijelölt beszélőnek pedig meg kell szólalnia (1a szabály). Ha a beszélő nem választja ki a következő beszélőt, akkor bármely más résztvevő átveheti a szót; az első megszólaló szerzi meg a jogot a következő fordulóhoz (1b szabály). Ha a beszélő nem választja ki a következő beszélőt, és más résztvevő sem szólal meg a b) lehetőség szerint, akkor az eredeti beszélő folytathatja (de nem köteles folytatni) a fordulót, azaz ő szerzi meg a jogot a forduló egy további szerkezeti egységéhez (1c szabály). Ezek a szabályok minden soron következő váltásreleváns helyen ugyanebben a sorrendben lefutnak, azaz működésük rekurzív.

Vizsgálatom legfontosabb eredménye annak megmutatása, hogy amikor az *öö* a beszélőváltásra alkalmas ponton jelenik meg, a beszélőváltás mindhárom szabályának alkalmazásakor szerepet játszhat, és ezekben a funkciókban értelmezhető diskurzusjelölőként. Három ilyen funkciót különítek el:

(i) az öö-t a kijelölt új beszélő használja annak jelzésére, hogy tisztában van az őt érintő normatív elvárással, hogy egy második párrészt produkáljon, és ezt az elvárást ki is kívánja elégíteni (a beszélőváltás 1a szabálya);
(ii) egy önmagát kijelölő új beszélő az *öö* használatával jelzi, hogy jogot formál a következő fordulóra, és ezzel megakadályozza, hogy az eredeti beszélő folytassa a fordulót vagy más résztvevők lépjenek be (1b szabály);

(iii) az aktuális beszélő arra használja az *öö*-t, hogy a szóátvételt megakadályozza, és további egység(ek)et fűzhessen a fordulójához (1c szabály).

A vizsgálat során bemutatott példák egy magyar nyelvű, hétköznapi társalgásokat tartalmazó korpuszból származnak, amely két részre osztható. Az egyik rész 9 db, egyenként kb. 20 perces, 3 résztvevős társalgást tartalmaz. Ezeket a Szegedi Tudományegyetem Pszichológiai Intézetében rögzítették. A korpusz további 8, egyenként kb. 15 perces, szintén 3 fős beszélgetése a BEA beszélt nyelvi adatbázis részét képezi [6]. Noha az *öö* valamennyi előfordulását egyenként megvizsgáltam a korpuszban, vizsgálatom szigorúan kvalitatív jellegű, kvantitatív elemzést nem tartalmaz. Ennek két oka van. Egyrészt vannak a korpuszban az *öö*-nek olyan előfordulásai, amelyek „ellenállnak az elemzésnek" (vö. [14]), azaz a rendelkezésre álló információink nem teszik lehetővé, hogy interakciós funkciójukat megállapítsuk. Ezen előfordulások számbavétele nélkül félrevezető lenne számadatokban megadni a különböző funkciókban megjelenő *öö*-ket. A tisztán kvalitatív analízist másrészt az indokolja, hogy a konverzációelemzés módszertani alapelve az előfordulásról előfordulásra történő induktív kvalitatív elemzés ([15]). E megközelítés szerint csak ezzel a módszerrel írható le és magyarázható a társas interakció szerkezete. Amellett érvelnek, hogy az induktív általánosításon túlmutató kvantitatív vizsgálatok torzíthatják az egyedi esetek észlelését, mert az előre meghatározott változók alapján a kutató „beleerőlteti" őket egy előre meghatározott kategóriába. A túl nagy tömegű adat pedig lehetetlenné teszi az egyedi előfordulásokra való újbóli és újbóli visszatérést, ami pedig szintén kulcsfontosságú a társalgáselemzésben ([13]).

Irodalom

[1] Deme, A. – A. Markó (2013). 'Lengthenings and filled pauses in Hungarian adults' and children's speech'. *Proceedings of DISS 2013. The 6th Workshop of Disfluency in Spontaneous Speech.* Szerk. Eklund, R. Stockholm: KTH Royal Institute of Technology, 21–24. old.

[2] Diewald, G. (2013). '"Same same but different" – Modal particles, discourse markers and the art (and purpose) of categorization'. *Discourse Markers and Modal Particles: Categorization and Description.* Szerk. Degand, L., B. Cornillie & P. Pietrandrea. Amsterdam: John Benjamins, 19–46. old.

[3] Fischer, K. (2006). 'Towards an understanding of the spectrum of approaches to discourse particles: Introduction to the volume'. *Approaches to Discourse Particles.* Szerk. Fischer, K. Amsterdam: Elsevier, 1–20. old.

[4] Gósy M. (1993). 'A lexikális hozzáférés: Szófelismerési stratégiák'. *Beszédkutatás 1993*. Szerk. Gósy, M. Budapest: MTA Nyelvtudományi Intézet, 14–32. old.

[5] Gósy M. (2006). 'A semleges magánhangzó nyelvi funkciói'. *Beszédkutatás 2006*. Szerk. Gósy, M. Budapest: MTA Nyelvtudományi Intézet, 8–22. old.

[6] Gósy M. (2012). 'Multifunkcionális beszélt nyelvi adatbázis – BEA'. *Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai Kutatások.* Szerk. Prószéky, G. & T. Váradi. Budapest: Akadémiai Kiadó, 329–349. old.

[7] Gósy M., Bóna J., Beke A. & Horváth V. (2013). 'A kitöltött szünetek fonetikai sajátosságai az életkor függvényében'. *Beszédkutatás 2013*. Szerk. Gósy, M. Budapest: MTA Nyelvtudományi Intézet, 121–143. old.

[8] Horváth V. (2014). *Hezitációs jelenségek a magyar beszédben*. Budapest: ELTE Eötvös Kiadó.

[9] Mondada, L. (2013). 'The conversation analytic approach to data collection'. *The Handbook of Conversation Analysis.* Szerk. Sidnell, J. & T. Stivers. Oxford: Wiley-Blackwell, 32–56. old.

[10] Németh Zs. (2020). ‘A nemlexikális *öö* hang interakciós szerepének elemzése magyar nyelvű társalgásokban’. *Jelentés és Nyelvhasználat* 7.1, 23–50.
[11] Sacks, H., E. A. Schegloff & G. Jefferson (1974). ‘A simplest systematics for the organization of turn-taking for conversation’. *Language* 50.4, 696–735.
[12] Schegloff, E. A. (1996). ‘Confirming allusions: Toward an empirical account of action’. *American Journal of Sociology* 102.1, 161–216.
[13] Schegloff, E. A. (2009). ‘One perspective on Conversation Analysis: Comparative Perspectives’. *Conversation Analysis: Comparative Perspectives.* Szerk. Sidnell, J. Cambridge: Cambridge University Press, 357–406. old.
[14] Schegloff, E. A. (2013). ‘Ten operations in self-initiated, same-turn repair’. *Conversational Repair and Human Understanding*. Szerk. Hayashi, M., G. Raymond & J. Sidnell. Cambridge: Cambridge University Press, 41–70. old.
[15] Stivers, T. & J. Sidnell (2013). ‘Introduction’. *The Handbook of Conversation Analysis*. Szerk. Sidnell, J. & T. Stivers. Oxford: Wiley-Blackwell, 1–8. old.

Non-Parallel Voice Conversion Incorporating Sinusoidal Model with Adversarial Learning

Mohammed Salah Al-Radhi[1], Tamás Gábor Csapó[1,2], and Géza Németh[1]
[1] Department of Telecommunications and Media Informatics
Budapest University of Technology and Economics, Budapest, Hungary
[2] MTA-ELTE Lendület Lingual Articulation Research Group, Budapest, Hungary
{malradhi, csapot, nemeth}@tmit.bme.hu

Abstract

Voice conversion (VC) aims to modify the speech signal of a source speaker to make it sound like being uttered by a target speaker, while keeping the linguistic contents unchanged. The traditional VC techniques are usually categorized as parallel VC, which learns a mapping using the training data of parallel utterance-pairs from the source and target speakers. Most of these methods rely on a time alignment procedure, which occasionally fails to estimate natural voices because the required alignment procedures introduce artifacts that affect all time samples in a given frame and takes up a lot of processing power. To mitigate this difficulty, this paper proposes a method that allows a non-parallel VC, in which the linguistic features are not shared between source and target speakers, by using a novel generative adversarial network (GAN). Moreover, it conducts the first study to accurately measure the impact of the sinusoidal model in non-parallel many-to-many VC, which does not involve any kind of parallel utterances or time alignment. Experimental results demonstrate that the proposed model successfully improve the speaker similarity of the converted speech. We also found objectively that the proposed method has a better capability for converting the source speaker to the target one than the state-of-the-art system.

Introduction

The technology of voice conversion (VC) has great potential in the development of various speech tasks such as text-to-speech, speaking assistance, and speech enhancement [1]. Towards the practical use of these VC applications, it is necessary to expand fundamental VC approaches. Various statistical methods have been proposed for VC, such as Gaussian mixture models [2], kernel partial least squares regression [3], frequency warping [4], and also non-linear neural networks have been studied [5]. Developing a conversion framework for a particular speaker typically requires considerable parallel data between the source and target speakers, but this may degrade the VC performance due to the inaccurate alignment.

In contrast, building a VC system from non-parallel speech dataset is complicated but highly valuable in real-time applications. A number of non-parallel VC techniques such as feature alignment [6], speaker adaptation [7], and variational autoencoder [8] have been proposed. Although these techniques make it possible to convert the speaker individuality, the conversion accuracy and converted voice quality are naturally reduced compared with those of the target voices. Still, this problem is very challenging and has room for improvement. This paper focuses on performing VC with completely nonparallel data using GAN together with a sinusoidal model to synthesize more realistic converted speech. It allows simultaneous training of multiple domains, i.e. many-to-many mapping within a single network. The converted speech has acceptable quality and similarity compared with the target utterances.

Proposed Methodology

To convert one voice to another, we first fed the source speech into the sinusoidal analyzer to extract the fundamental frequency, maximum voiced frequency, and spectral envelope [9]. Second, we construct an acoustic feature vector by stacking those features. Then, we normalize the acoustic features over all the speakers in the training dataset to have zero-mean and unit-variance. In the training time, we train a single generator G to convert an input source x into an output signal y conditioned on the target label c, $G(x,c) \rightarrow y$. We also introduce a single discriminator to produce probability distributions over both sources and domain labels. Both the generator and the discriminator networks are iteratively updated, where one module is being optimized, whereas the module parameters of another are corrected. Inspired by [10], three main losses (adversarial, classification, and cycle consistency) are calculated and minimized to optimized both G and D. Finally, a sinusoidal synthesis approach is used for reconstructing speech from converted features. Figure 1 depicts the structure of the proposed approach.

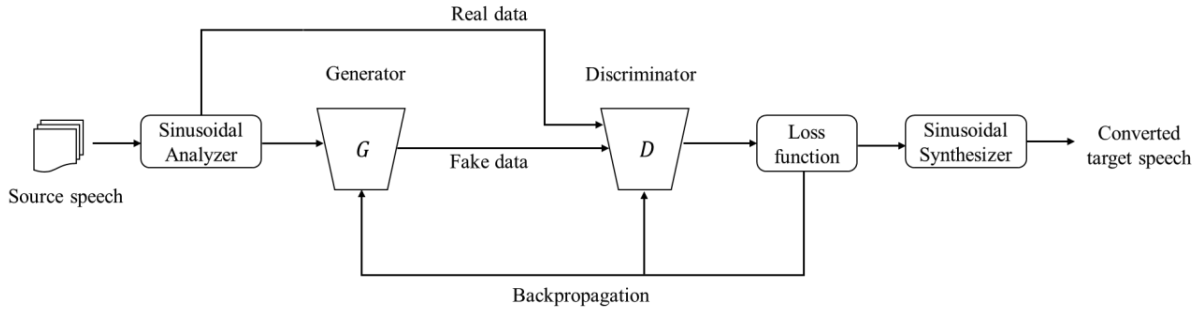


Figure 1: Overview of the developed VC based on sinusoidal model.

Experimental Results and Discussion

Our proposed system was evaluated on the CSTR VCTK Corpus of 8 English speakers (4 males and 4 females). We run both inter-gender (female-to-male and male-to-female) and intra-gender (male-to-male and female-to-female) conversions. Hence, 12 different combinations of source and target speakers. A log spectral distance (LSD) metric is considered to evaluate the quality of the proposed model. The state-of-the-art system followed the framework of StarGAN-VC model that achieved a good level of naturalness and speaker similarity of the converted speech in non-parallel VC [11]; therefore, we compared our system with it. From the demonstration samples in Figure 2, it was objectively confirmed that the proposed method has a better capability for converting the source speaker to the target one than the state-of-the-art system.

To confirm the performance of our proposed method, we conducted subjective evaluation experiments involving a nonparallel many-to-many speaker conversion task. Fifteen listeners participated in our listening tests. Audio samples are provided at https://malradhi.github.io/contSM-VC/. The results are shown in Figure 3. As the results reveal, the proposed method is found to be comparable to the StarGAN-VC model; and the proposed framework validates the effectiveness of the sinusoidal model with continuous features in the conversion pipeline. We can conclude that our framework is able to generate a voice similar to the target speaker in comparison to the system with more sophisticated discontinues parameters.

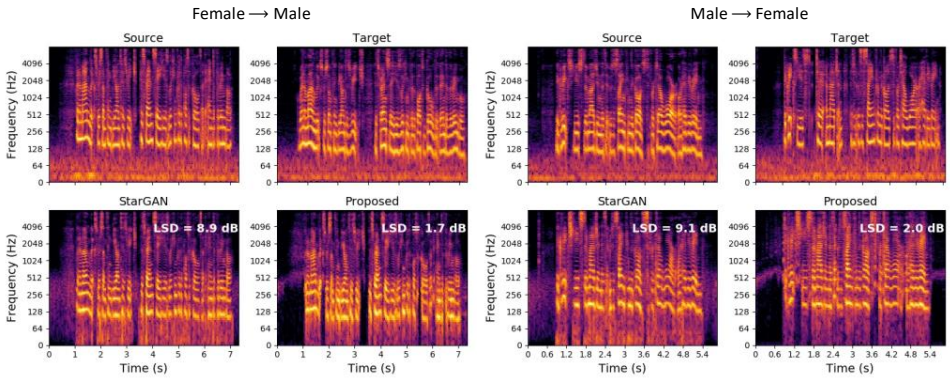


Figure 2: Spectrograms of the source, target, and converted speeches.

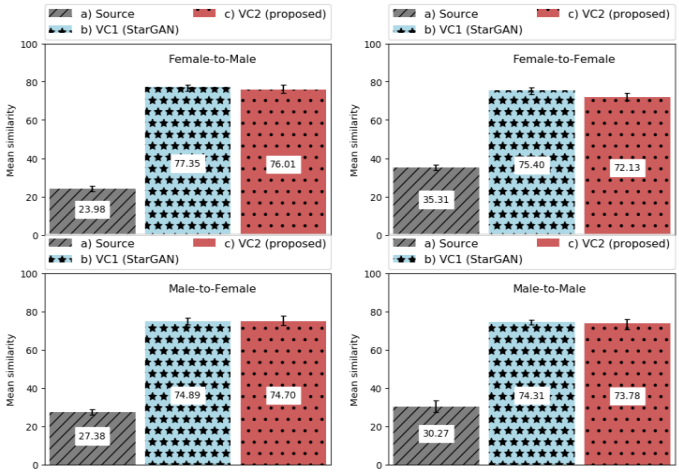


Figure 3: Average preference scores on speaker similarity.

References

[1] Mohammadi, S.H. & Kain, A. (2017). 'An overview of voice conversion systems'. Speech Communication, 8, pp. 65-82.

[2] Toda, T., Black, A. W. & Tokuda, K. (2007). 'Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory'. IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 8, pp. 2222-2235.

[3] Helander, E., Sil, H., Virtanen, T. & Gabbouj, M. (2012). 'Voice conversion using dynamic kernel partial least squares regression'. IEEE transactions on audio, speech, and language processing, vol. 20, no. 3, p. 806–817.

[4] Erro, D., Moreno, A. & Bonafonte, A. (2010). 'Voice conversion based on weighted frequency warping' IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, no. 5, pp. 922-931.

[5] Al-Radhi, M.S., Csapó, T.G. & Németh, G. (2019). 'Continuous vocoder applied in deep neural network based voice conversion'. Multimed Tools and Applications, vol. 78, p. 33549–33572.

[6] Hashimoto, T., Saito, D. & Minematsu, N. (2016). 'Arbitrary speaker conversion based on speaker space bases constructed by deep neural networks'. In: *Proceeding of APSIPA* , pp. 1-4.

[7] Lee, C.-H. & Wu, C. (2006). 'Map-based adaptation for speech conversion using adaptation data selection and non-parallel training'. In: Interspeech, pp. 17-21.

[8] Hsu, C., Hwang, H., Wu, Y., Tsao, Y. & Wang, H. (2017). 'Voice conversion from unaligned corpora using variational autoencoding Wasserstein generative adversarial networks'. In: Interspeech, pp. 3364-3368.

[9] Al-Radhi, M.S., Csapó, T.G., & Németh, G. (2018). 'A Continuous Vocoder Using Sinusoidal Model for Statistical Parametric Speech Synthesis'. In: Proceedings of the Speech and Computer, Lecture Notes, vol. 1109, pp. 11-20.

[10] Choi, Y., Choi, M., Kim, M., Ha, J., Kim, S. & Choo, J. (2018). 'Stargan: Unified generative adversarial networks for multi-domain image-to-image translation'. In: Proceedings of the Computer Vision and Pattern Recognition, pp. 8789–8797.

[11] Kameoka, H., Kaneko, T., Tanaka, K. & Hojo, N. (2018). 'StarGAN-VC: non-parallel many-to-many Voice Conversion Using Star Generative Adversarial Networks," In: proceedings of Spoken Language Technology Workshop, pp. 266-273.

Kódváltás magyar-angol kétnyelvűek beszédprodukciójában

Salamon Attila
Pannon Egyetem, Veszprém

Mivel napjainkban az internet és a közösségi média egyre nagyobb teret biztosít az angol nyelv terjedésének, egynyelvű nyelvi környezetben is várhatóvá vált annak használata, különösképpen az angol nyelvet magas fokon ismerők körében. Ez a tény szolgált érdeklődésem alapjául és ösztönzött egy újszerű kutatás elvégzésére.

Kutatásom célja az egynyelvű nyelvterületen élő kétnyelvűek beszédprodukciójában fellelhető kódváltások vagy kódkeverések kielemzése és megjelenésük okainak feltárása. Mivel a résztvevők a kétnyelvű nyelvi módban (Grosjean, 2013)[1] vannak, mindkét nyelvük aktív, vagy legalábbis egy bizonyos mértékig az. A redukált gátló hatás miatt kódváltásokra/kódkeverésekre lehet számítani. A kódkeverés definíciója alapján – miszerint a mátrix és a beágyazott nyelv meghatározása nehézkessé válik a beszédprodukció során – kódkeverésnek tekintettem a beágyazott nyelvi elemek megjelenését, mikor azok egyértelműen elkülöníthetőek voltak a mátrix nyelvtől. Navracsics (2010)[3] szerint a kódváltást diskurzusjelölő, prozódiai jel vagy valamilyen megakadásjelenség előzi meg általában, viszont példáim alapján ennek ellenkezőjére is számos példát találtam. A szakirodalom segítségével meghatározott definíciók és jelenségek magyarázata után a hipotéziseim a következők:

1) Nyelvváltás számos esete várható pragmatikai okokból fakadóan. 2) A beszélgetések során az idő előrehaladtával valószínűleg gyengül a kontroll a nyelvek felett, így gyakoribb kódváltás, sőt kódkeverés várható a beszélgetések vége felé. 3) A beszélgetések témája befolyással bír a kódváltások gyakoriságára.

A párbeszédek és a megvizsgált példák egy podcast sorozatból lettek feldolgozva, amit a magyar anyanyelvű résztvevők (4 angol szakos egyetemi hallgató) szándékosan vettek fel. A résztvevők mind kései kétnyelvűek, 9-10 éves koruktól kezdtek ismerkedni az angol nyelvvel. Bár egynyelvű nyelvi környezetben élnek, napi rendszerességgel használják az angolt. Mivel a hangrögzítés önként történt, közönségnek szánt, így a megfigyelői paradoxonból (Labov, 1972)[2] származó problémák nagyrészt elkerülhetővé váltak; a beszédprodukció feltételezhetően nagyon közel áll a természeteshez. Az epizódok online elérhetőek bárki számára a YouTube videó megosztó weboldalon. Összesen 10 epizód volt kielemezve, melyekben egy-egy különböző témát dolgoznak fel. A beszélgetések bázisnyelve a beszélgető felek döntése szerint magyar.

Az adatelemzéshez kvalitatív módszereket alkalmaztam, kivéve a nyelvváltások gyakoriságának vizsgálata során. A kvalitatív elemzés az első, míg a kvantitatív vizsgálat az azt követő hipotézisek validitását volt hivatott ellenőrizni.

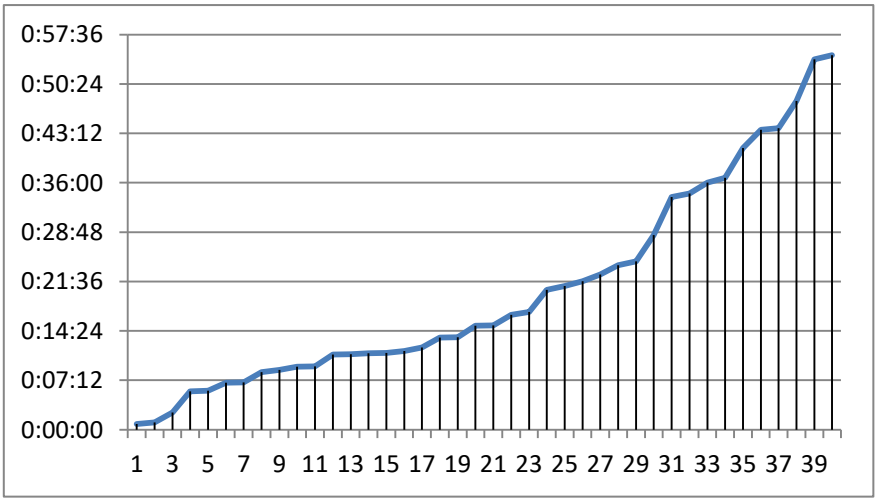
A kvalitatív elemzés részben szubjektív eszközökkel történt. Először a nyelvváltások típusát határoztam meg a szakirodalom segítségével. Ezek után a kódváltások pragmatikai okát próbáltam megtalálni – ha volt egyáltalán – a hangfelvételek alapján, illetve azt vizsgáltam, hogy megelőzte-e azokat megakadásjelenség vagy sem.
A kódváltások kvantitatív elemzése grafikonok segítségével történt, melyeken jól vizsgálható volt azoknak száma és megjelenésük pontos ideje a beszélgetésekben.

Első hipotézisem alátámasztást nyert a vizsgálatok eredményei alapján, mivel a kódváltások többsége pragmatikai erő által jött létre, habár számos esetben nem találtam lehetséges magyarázatot a kódváltás megjelenésére. A mátrix nyelv és a beágyazott nyelv elkülönítése alapján a tanulmányban kódkeverésre nem volt példa, míg – a szakirodalomtól eltérő eredmények alapján – kódváltásból négy kategóriát különböztettem meg: a) egyértelmű pragmatikai okból történő kódváltás megakadásjelenséget követően, b) egyértelmű ok nélküli kódváltás megakadásjelenséget követően, c) pragmatikai okból történő kódváltás megakadásjelenség nélkül, d) egyértelmű ok, valamint megakadásjelenség nélkül megjelenő kódváltás. Egy lehetséges válasz a nyelvi fúzió gyakori megjelenésére, hogy egy bizonyos szemantikai vagy pragmatikai szükséglet merül fel a tartalmat hordozó szó beillesztésére a mátrix nyelvbe, illetve grammatikai (szintaktikai) szükséglet áll fenn, hogy az a szó a mátrix nyelv mondatába illeszthetővé váljon. Ez okozza az angol szó magyar toldalékokkal való kiegészítését (pl.: le-spoiler-ez-het-jük).

A második hipotézis ellenőrzése nem volt egyszerű, mivel az epizódok majdnem felében a kódváltási gyakoriság eloszlása alátámasztja, míg a többi rész elveti azt. Nincs egyértelmű válasz arra, hogy a résztvevők az idő előrehaladtával egyre kevésbé tudták elnyomni a második nyelvüket vagy sem. Az bizonyos, hogy más tényezők is részt vehettek a kódváltások gyakoriságának eloszlásában, mint például a mentális és fizikai állapot, kontextus, vagy előfeszítés (priming).

A harmadik hipotézist alátámasztották a tanulmány eredményei, mivel bizonyos témák valóban jóval több kódváltást idéztek elő, mint például a technológia (40) vagy a politikai korrektség/tabuk (41), ellentétben a halál (9) és szexualitás (11) témákkal.

A tanulmány eredményei alapján általános következtetéseket elhamarkodott lenne meghozni, mivel egy esettanulmányról van szó, viszont az kijelenthető, hogy a korábbi általános állítások, melyek igaznak bizonyultak más, egymáshoz közelebbi nyelvcsaládok nyelveit összevetve, nem feltétlenül állják meg helyüket a magyar nyelv esetében.



1. ábra. Kódváltások a ’Technológia I.’ c. beszélgetésből

Irodalom

[1] Grosjean, F. & Li, P. (2013). *The Psycholinguistics of Bilingualism*. Malden, MA & Oxford: Wiley-Blackwell, p. 5-16.

[2] Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia. University of Pennsylvania. p. 209.

[3] Navracsics, J. (2010). *Egyéni kétnyelvűség*. Szegedi Egyetemi Kiadó

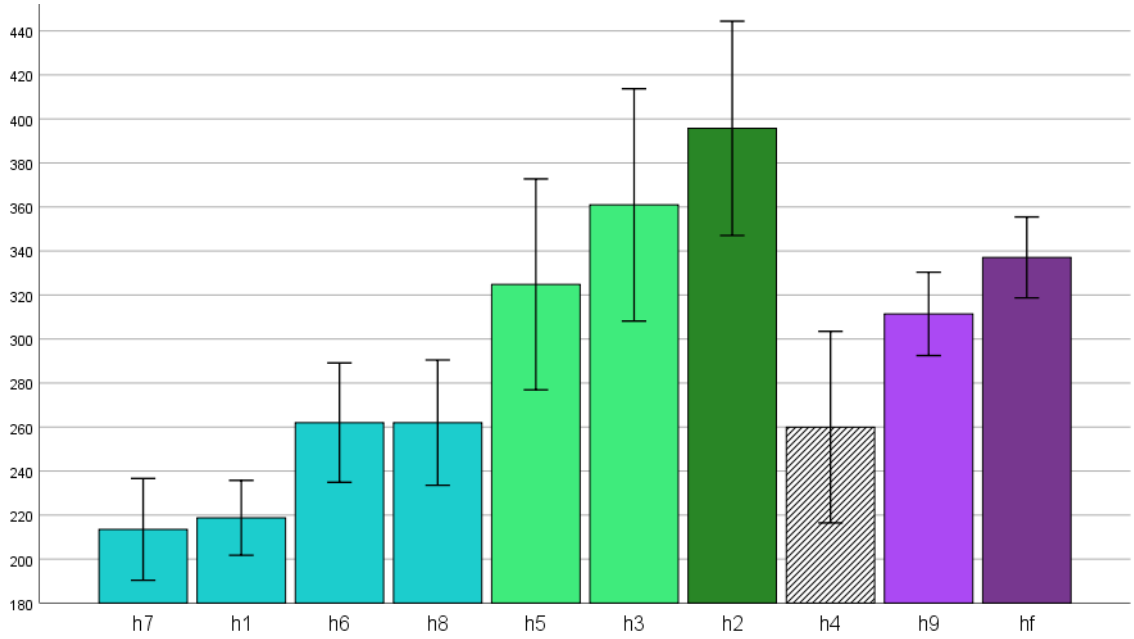
A *hát* diskurzusjelölő prozódiai megvalósulásának vizsgálata felolvasásokban

Szeteli Anna[1], Gocsál Ákos[1,2], Szente Gábor[1] és Alberti Gábor[1]
[1] PTE, Pécs
[2] MTA NYTI, Budapest

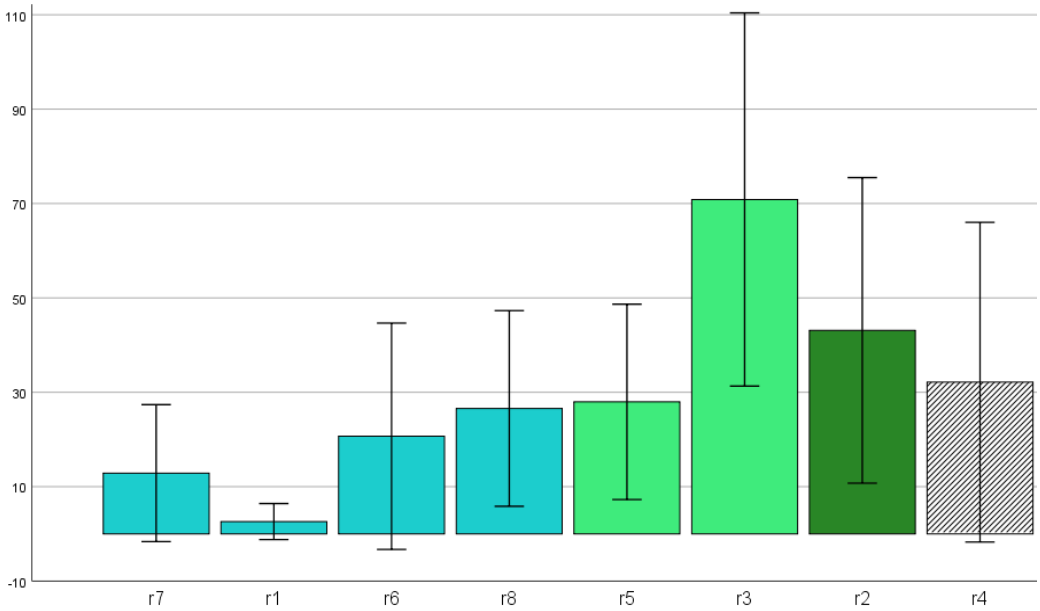
A spontán beszédet szervező egyik leggyakoribb elem, a *hát* diskurzusjelölő [1], az utóbbi években a pragmatika mellett (pl. [2, 3, 4]) a beszédkutatás irányából is jelentős figyelmet kapott a hazai tudományos közösség felől [5, 6, 7]. Előadásunkban egy egyrészt a pragmatikai kategóriákat kiindulópontjául vevő, másrészt a korábbi beszédkutatási munkák által lefektetett kategóriákat éppúgy szem előtt tartó kísérlet lebonyolításáról és eredményeiről számolunk be. A fenti szembeállítás természetesen nem úgy értendő, hogy az ilyen munkák szükségszerűen eltérő eredményekre vezetnének, azonban a pragmatika, illetve a beszédkutatás nézőpontjából kiinduló kutatások gyakran – más szempontok alapján – eltérő kategóriákat állítanak fel.

A kísérlet 63 adatközlő bevonásával zajlott és leginkább a felolvasás műfajába sorolható, bár az adatközlők kaptak felkészülési időt. A szövegek többfordulós párbeszédeket tartalmaztak, melyeket az adatközlők a kísérletvezetővel bonyolítottak le. A kilenc párbeszéd egy közös kerettörténetbe volt beágyazva, és a különböző párbeszédek a tíz vizsgált funkciójú *hát* diskurzusjelölőt tartalmazták. A kísérletben 21 év körüli egyetemisták vettek részt, az adatközlők életkora ezáltal homogénnek mondható, a férfiak és a nők aránya pedig kiegyenlített (25:28). Az első négy párbeszéd [3] pragmaszemantikai kategóriáit követve tartalmazott egy határozott (h1), egy bizonytalan (h2), egy aggodalmas (h3) és egy incselkedő (h4) diskurzusjelölőt, melyek mondatkezdők voltak, továbbá egy mondatvégi nyomatékosító funkciójú *hát* is bekerült (hf, mint 'final'). Az iménti kategóriákat már [7] is tesztelte kis mintán, jelen kísérlet tehát a fenti kategóriákat egészítette ki egy a beszélő attitűdjét erősen hordozó lemondó típussal (h5), illetve az [5] által vizsgált kategóriákkal, melyeket inkább formai, mint funkcióbeli szempontok figyelembevételével különítettek el. Így adódott egy mondandó indító (h6), egy összegző (h7), illetve egy értékelő típus (h8). Utolsó kategóriaként egy a másik mondatvégi jelölőtől (hf) eltérő funkciójú, konkluzív záró *hát* került bevezetésre (h9), hogy a nyomatékosító *hát*tal (hf) való prozódiai kapcsolatára fény derülhessen. A fenti párbeszédek elemzése a Praat és az SPSS programok segítségével történt. A diskurzusjelölők hosszadatai alapján több kategória szignifikánsan elkülönült egymástól (ld. 1. ábra), mint például a határozott (h1) és a bizonytalan (h2), melyek esetében utóbbi kézenfekvő módon hosszabban valósul meg. A kísérletes módszer egyik fő hozadékának tarjuk, hogy ezekben az esetekben a hosszúságban megjelenő különbségeket tisztán pragmatikai elemnek könyvelhetjük el, mivel nem merül fel a beszédképzésből adódó egyéb jelenség – mint például a hezitáció – háttértényezőként, hanem az adatközlő a szótag elnyújtásával egyértelműen és szándékosan jelöl bizonytalanságot, aggodalmat stb. Előadásunkban a hosszadatokon kívül figyelmünket a diskurzusjelölő magánhangzójából három ponton kivett alapfrekvencia-adatok felé fordítjuk (3. ábra), melyeknek a statisztikai értékeléshez szükséges kinyerését a spontánbeszéd-korpuszokban sok esetben egyidejű beszéd, illetve irregularitás lehetetleníti el [5]. Az alapfrekvencia-adatokból származó legmarkánsabb eredményünk a nyomatékosító *hát* irregularitása a mondatvégi párjához a következtető típusú diskurzusjelölőhöz képest. Bíztatóak továbbá a diskurzusjelölő után tartott szünet tekintetében megfigyelhető tendenciák (2. ábra).

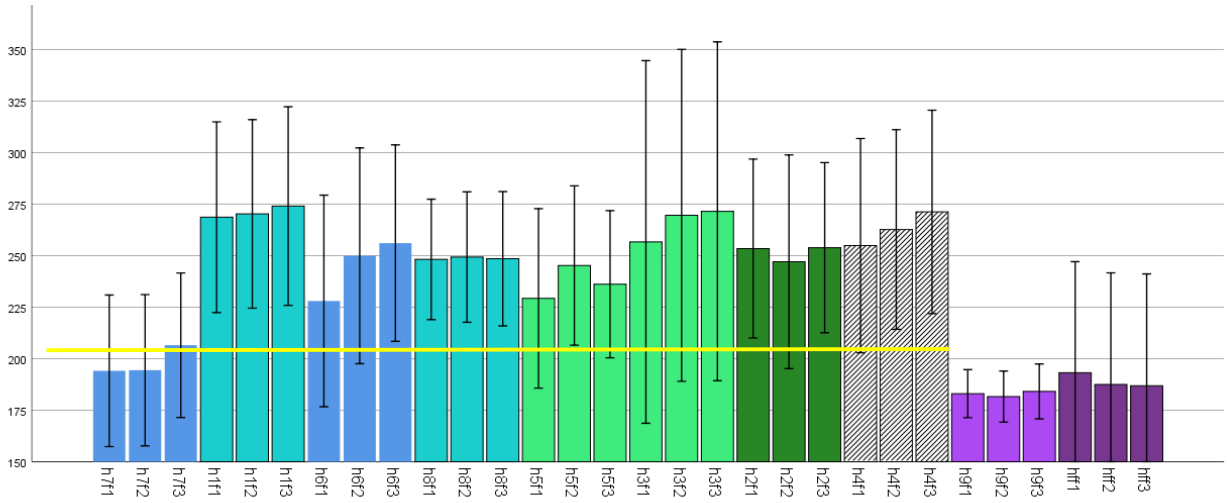
Előadásunkban a fenti kísérlet eredményeit mutatjuk be részletesen, ami [7] pilot kísérletének kategóriáit nagyszámú (n=63) adatközlőre alkalmazza. Emellett a diskurzusjelölőre jellemző hangtani tulajdonságokat öt további pragmatikai funkciót kijelölve vizsgáljuk, melyek megválasztása során a magyarországi beszédkutatás kurrens publikációit is figyelembe vettük. Így bár kísérletes eljárást alkalmaztunk spontán beszéd adatbázis vizsgálata helyett, kutatásunk [5] konstruktív replikációjának is tekinthető.



1. ábra. Hosszkülönbségek a tíz vizsgált kategóriában (95%-os konfidenciaintervallumokkal)



2. ábra. A diskurzusjelölő után tartott szünetek a tíz vizsgált kategóriában (95%-os konfidenciaintervallumokkal)



3. ábra. Három ponton mért frekvenciaadatok a tíz vizsgált kategóriában (95%-os konfidenciaintervallumokkal)

Irodalom

[1] Dér Cs. I. & Markó A. (2007). 'A magyar diskurzusjelölők szupraszegmentális jelöltsége.' *Nyelvelmélet nyelvhasználat*. Szerk. Gecső T. & Sárdi Cs. Székesfehérvár – Budapest: Kodolányi János Főiskola – Tinta Könyvkiadó. 61–67. old.

[2] Schirm A. (2011). *A diskurzusjelölők funkciói: a* hát, *az* -e *és a* vajon *elemek története és jelenkori szinkrón státusa alapján*. Doktori disszertáció, Szeged.

[3] Alberti G. (2016). Hát a *meg* meg a *hát*. *"Szavadd ne feledd!" Tanulmányok Bánréti Zoltán tiszteletére.* Szerk. Kas B. Budapest: MTA NytI. 17–27. old.

[4] Szeteli A. & Alberti G. (2018). Hát igen, más hát! *Mesterek és Tanítványok 2: Tanulmányok a bölcsészet- és társadalomtudományok területéről*. Szerk. Czeferner D., Böhm G. & Fedeles T. Pécs: PTE BTK KTDT. 27–53. old.

[5] Dér Cs. I. & Markó A. (2017). A *hát* funkciói a prozódiai megvalósulás függvényében. *Beszédkutatás* 25. 105–117. old.

[6] Schirm A., Szabó V. & Gocsál Á. (2017). *Beszélt nyelvi elemek a tanári megnyilatkozásokban*. Előadás: Újdonságok a szemantikai és pragmatikai kutatásokban konferencia. Szeged, 2017. április 28.

[7] Szeteli, A., Gocsál Á. & Alberti G. (2019). Szemafor hát! *Jelentés és Nyelvhasználat* 6:1. 33–63. old.

UH- és MRI-nyelvkontúrok optimalizációja

Trencsényi Réka[1] és Czap László[2]
[1] Debreceni Egyetem, Villamosmérnöki Tanszék
[2] Miskolci Egyetem, Automatizálási és Infokommunikációs Intézet

Az emberi beszédprodukció tanulmányozásának legalapvetőbb eszközei az ultrahang (UH) [2] és mágneses rezonanciás képalkotó (MRI) [5] technikával készült dinamikus felvételek. Az emberi testet oldalnézetben ábrázoló, ún. szagittális síkban létrehozott kétdimenziós UH- és MRI-metszetek vizsgálatával és feldolgozásával releváns kvalitatív és kvantitatív információkat nyerhetünk az artikuláció legfontosabb jellemzőiről. A kvalitatív megállapítások főként a nyelv és a szájpad relatív pozíciójára vonatkoznak különböző beszédhangok és hangátmenetek esetén, míg a kvantitatív leírások a geometriai paraméterek azonosítására és a köztük lévő összefüggésekre fókuszálnak, melyeknek fontos szerepe van a beszéd artikulációs és akusztikai jellemzői közötti kapcsolatok megértésében. A kvantitatív elemzések igen sokféle és változatos módon valósíthatók meg [1,3,4]. Jelenlegi vizsgálataink kiindulópontjai a jelzett UH- és MRI-keretekre [6,8] illesztett nyelvkontúrok, amiket saját fejlesztésű automatikus algoritmusaink segítségével állítottunk elő. A felhasznált UH- és MRI-források részleteiket tekintve sok eltérést mutatnak, ami a beszélők nemét és nemzetiségét, a felvételek geometriáját, felbontását és skáláját, illetve a vokális traktus vizuálisan értékelhető anatómiai szegmenseit érinti. Kutatómunkánk célja az UH- és MRI-felvételek kölcsönösen egyértelmű megfeleltetése az UH- és MRI-nyelvkontúrok között megvalósított megfelelő geometriai transzformációk kidolgozásával és a transzformáció paramétereinek optimalizálásával.

A transzformáció egzakt matematikai alakjának felírásakor a rendelkezésre álló UH-felvételek speciális geometriájára támaszkodtunk. A képalkotó UH-fej ugyanis a szájüregi régiónak egy olyan radiális tartományát szondázza, amely egy rögzített *C* középpontból mérve 90°-os szög alatt látszik. Ebből adódóan kézenfekvő az UH-képek és a hozzájuk tartozó nyelvkontúrok pontjait egy olyan *C* origójú polárkoordináta-rendszerben kezelni, ahol az egyes képpontok helyzetét egyértelműen megadja a *C* pontból mért *r* sugár, illetve a kép függőleges középtengelyétől mért előjeles φ szög. A transzformáció célja az UH-keretek radiális geometriájának beágyazása az MRI-felvételek síkbeli descartes-i koordinátákkal jellemzett négyszöges geometriájába úgy, hogy az ugyanazon hanghoz rendelt UH- és MRI-nyelvkontúrok a lehető legnagyobb mértékű fedésbe kerüljenek egymással. Az UH-nyelvkontúrok transzformációja három alapvető műveletet foglalhat magába: a sugártartomány skálázását, a szögtartomány skálázását, illetve a szögtartomány elforgatását. A három operáció matematikailag az

$$\begin{aligned} r' &= R \cdot r \\ \varphi' &= FI \cdot \varphi \\ \varphi_0{}' &= \varphi_0 + FI_0 \end{aligned} \qquad (1)$$

formulák segítségével valósítható meg, ahol az *R* és *FI* skálafaktorok a sugár- és szögtartomány normálását teszik lehetővé, a FI_0 tag pedig a szögtartomány kezdőszögének eltolását végzi. Az (1) összefüggések tehát az UH-nyelvkontúrt a megfelelő MRI-keretre illesztik. Az (1)

transzformációk inverzének alkalmazásával azonban a fordított irányú konverzió is végrehajtható, azaz az

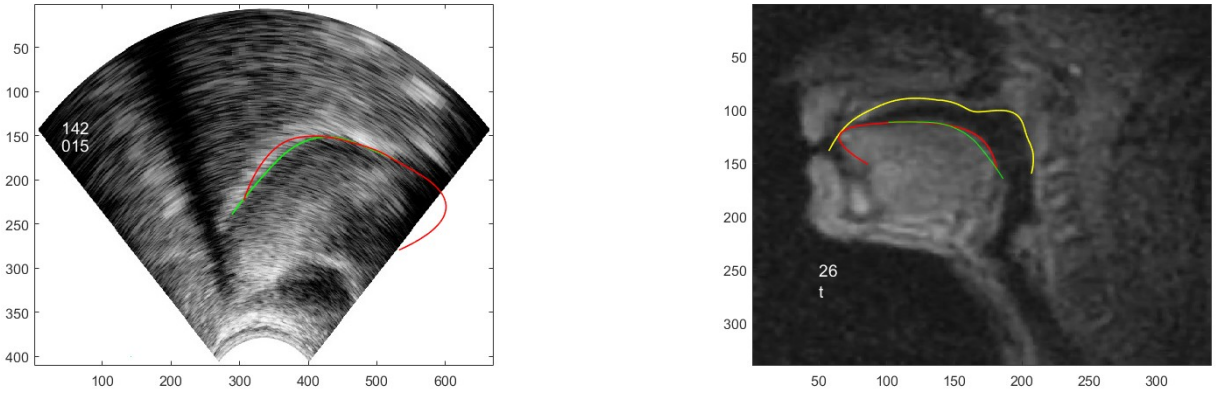
$$r = r' / R$$
$$\varphi = \varphi' / FI$$
$$\varphi_0 = \varphi_0' - FI_0 \qquad (2)$$

alakú inverz operációk segítségével az MRI-nyelvkontúr rávetíthető a megfelelő UH-keretre. Az UH-MRI, illetve MRI-UH irányban elvégzett transzformációk $\{R, FI, FI_0\}$ paraméterhalmazának szükségszerűen meg kell egyeznie, hiszen ezáltal biztosítható az UH- és MRI-környezet relatív skálaarányának megtartása a konverzió irányától függetlenül. A vizsgálatok során a *FI* faktor értékét mindvégig *FI*=1 szerint rögzítettük, ami azt jelenti, hogy a transzformáció szögtartó.

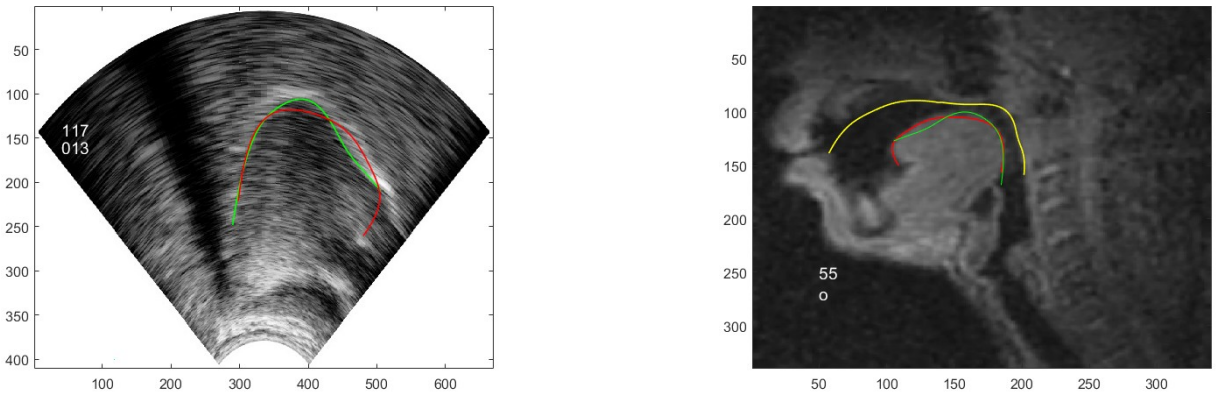
Az (1)-(2) transzformációk az *R* és FI_0 paraméterek számszerű meghatározásával válnak érvényessé, amihez egy lehetséges utat kínál a paraméterek értékeinek optimalizálása. Az optimalizációs eljárás során egy általunk kidolgozott algoritmus felhasználásával megkeressük azt a paraméterhalmazt, amely esetén a transzformált UH-nyelvkontúr és a referenciagörbeként szolgáló MRI-nyelvkontúr közötti távolság minimális. A távolság számítása a két görbe összes lehetséges pontpárjára megtörténik, majd képezzük az UH-nyelvkontúr egyes pontjaihoz rendelt legkisebb távolságok átlagát [7]. A sikeres transzformációhoz azonban nemcsak az *R* és FI_0 paraméterek pontos értékeire van szükségünk, hanem ismernünk kell az MRI-kereten kijelölt C' középpontot is, ami az UH-felvétel *C* középpontjának a transzformáltja. Ezeken túlmenően az optimalizációs algoritmus konstrukciójakor jó támpontként szolgálhat a gégefedő *G* csúcspontja. Ennek folytán az optimalizációs algoritmust olyan matematikai formulák mentén alkottuk meg, melyek lehetővé teszik az $\{R, FI_0, C', G\}$ paraméterek egyidejű optimalizációját. Az *R* skálafaktort az UH- és MRI-kereten mérhető, a középpont és a gégefedő csúcspontja közötti R_{UH} és R_{MRI} radiális távolságok arányaként értelmeztük. A FI_0 szögelfordulás az R_{UH} és R_{MRI} szakasz, illetve a kép függőleges középtengelye által bezárt előjeles szögek összegeként adódik. Az MRI-keretek C' középpontját a (c_1,c_2) descartes-i koordinátapár jellemzi, ahol c_1 a függőleges, c_2 pedig a vízszintes tengely mentén mért pozíció. A *G* paraméter a gégefedő csúcspontját lokalizálja az UH-kereteken a (g_1,g_2) descartes-i adatpár kijelölésével, ahol g_1 a függőleges, g_2 pedig a vízszintes tengely koordinátája. Az $\{R, FI_0, C', G\}$ paraméterek optimalizációját egy öt beszédhangból álló halmazra végeztük el, mely a H = $\{a, e, k, t, s\}$ hangokat tartalmazta. Az eredmények verifikálása céljából olyan UH-MRI kontúrpárok egymásra vetítését is ellenőriztük az optimalizáció által adott paraméterek beállítása mellett, melyek nem szerepeltek a H halmazban. Az eredmények további validálása jelenleg is folyamatban van. A nyelvkontúrok optimalizációját az 1. és 2. ábrák példázzák az $R = 0.41$, $FI_0 = 0.3$ rad, $C' = (250,140)$, $G = (242, 273)$ értékek esetén.

Köszönetnyilvánítás

Köszönjük az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoportjának, hogy rendelkezésünkre bocsátották a SonoSpeech rendszerrel készült ultrahang felvételeket.



1. ábra. Az optimalizáció eredménye a *t* hang esetén az UH- (zöld) és MRI-nyelvkontúr (piros) együttes ábrázolásával



2. ábra. Az optimalizáció eredménye az *o* hang esetén az UH- (zöld) és MRI-nyelvkontúr (piros) együttes ábrázolásával

Irodalom

[1] Danner, S. G., Barbosa, A. V., Goldstein, L.: *Quantitative analysis of multimodal speech data.* Journal of Phonetics, 71, 268-283 (2018)

[2] Denby, B., Stone, M.: *Speech synthesis from real time ultrasound images of the tongue.* In 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1, I-685 (2004)

[3] Fulcher, L., Lodermeyer, A., Kähler, G., Becker, S., Kniesburges, S.: *Geometry of the vocal tract and properties of phonation near threshold: calculations and measurements.* Applied Sciences, 9(13), 2755 (2019)

[4] Ojalammi, A., Malinen, J.: *Automated segmentation of upper airways from MRI-vocal tract geometry extraction.* In International Conference on Bioimaging, 3, 77-84 (2017)

[5] Scott, A. D., Wylezinska, M., Birch, M. J., Miquel, M. E.: *Speech MRI: morphology and function.* Physica Medica, 30(6), 604-618 (2014)

[6] Xu, K., Csapó, T. G., Roussel, P., Denby, B.: *A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization.* The Journal of the Acoustical Society of America, 139(5), EL154-EL160 (2016)

[7] Zharkova, N., Hewlett, N.: *Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/and /a/.* Journal of Phonetics, 37(2), 248-256 (2009)

[8] https://sail.usc.edu/span/rtmri_ipa/je_2015.html

Czech vowel quantity in Polish speakers as perceived by Moravian-Silesian listeners

Jitka Veroňková, and Tomáš Bořil
Charles University, Prague, Czech Republic

Introduction

The quantity of vowels in Czech is difficult for non-native speakers, including Polish speakers. The Polish vowel system contains phonologically only short vowels, however, according to some studies, there may be a phonetic difference in the duration of stressed and unstressed vowels [5]. Unlike Polish, in Czech the vowel length is phonological; it distinguishes lexemes or grammatical forms. In this paper, we present the experiment focusing on L2 Czech speakers' vowel length with Polish mother tongues as perceived by L1 Czech listeners. It is based on sentences with ambiguous context.

Method

The set contained pairs of sentences differed only in the target words. The words were disyllabic and varied in a combination of short (S) / long (L) vowels. Other conditions included, for example, position of a target word in the middle of a sentence, only short vowels in the remaining words of the sentence etc. Example: *Tyhle* ***valy*** (SS) */* ***vály*** **(**LS) *ze dřeva budily pozornost.* (Eng. These wooden *pastry boards / mounds* attracted attention.) The set consisted of 13 sentences, including, except pairs, also one triad. In the text for recording, the sentences were mixed among others to mask the target sound phenomena.

Five Polish speakers recorded the set. Three speakers studied Czech at university as a master's degree (two in Poland, one in the Czech Republic); they lived in the Czech Republic for 4–8 years (proficient users according to CEFR [2]). Two speakers had good language skills, studying humanities (but not Czech) in Poland; their stay in the Czech Republic was shorter (max. 2 years) and they attended Czech language courses there (independent users according to CEFR [2]).

63 sentences in total were used to create a perceptual test (13 sentences x 5 speakers; 2 sentences were excluded because of slips of tongue) using Praat MFC (multiple forced-choice) environment [1]. Listeners determined which variant of the target word they heard. The purpose was to test the segmental intelligibility, not the overall understanding of the content [3: 76, 6], therefore all four S/L combinations of a target word were at listeners' disposal. On the screen, they were presented with all the variants written in phonological transcription in order to diminish the influence of orthography, e.g. /vali/, /vali:/, /va:li/, /va:li:/; usually two forms fitted in the carrier sentence, two did not (lexically or grammatically).

The test was perceived by 25 native Czech listeners (students of Czech studies from University of Ostrava) from Moravian-Silesian region of the Czech Republic around Ostrava city. This region lies on the border with Poland and is characterized by some sound features similar to those in Polish. For example, vowel shortening is typical. That is why we administered perception tests in this area – to test, how sensitive those listeners are to vowel length. The focus of the following research will lie in comparing these speakers with listeners from the Bohemian part of the Czech Republic. Differences in perception of i-vowels between listeners' groups from Bohemian and Moravian parts have already been detected. [4]

The listeners' agreement with the original text has been counted and the types of substitutions have been analysed. In addition to perceptual analysis, we manually labelled the items to obtain vowel duration and then computed normalized vowel duration [1].

Results

Regarding the segments, 81.0% of vowels were perceived in accordance with the original text, and 19.0% incorrectly, of which 46.7% were originally short and 53.3% were originally long. Regarding the target words, 64.9% of perceived items corresponded to the original text.

The agreement with the speaker's intent (see Table 1) prevailed in words originally containing just one long vowel, i.e. in patterns SL (87.2%) and LS (82.2%). Compared to that, the pattern with two long vowels (LL) caused difficulties. In about half of the cases (46.4%), the pattern was perceived in accordance with the original text, but in the same volume (46.0%), the listeners indicated the LS pattern with the second vowel short. For three carrier sentences, a perceived LS variant fitted into the sentence instead of original LL pattern (pl form → sg form, e.g. *lánů* → *lánu*). In one of the sentences, the listeners indicated LS variant as well, even though it did not suit the sentence (verb → noun: *mýlí* → *míli*). We therefore believe that, based on the task, listeners focused on the intelligibility of the target words, not on the content of the sentence. This conviction is supported by the assessment of original SS pattern as well. It was recognized with higher success, but no longer so certain (62.2%), and in case of mismatch, listeners' choice most often vacillated between LS or SL variants regardless of their appropriate use in the sentence.

These findings were also confirmed acoustically (see Fig. 1). Distribution of a-vowel duration indicates two visible peaks clearly distinguishing S and L vowels (e.g. *lanu – lánu*), unlike i-vowels and u-vowels, where only the duration of S shows a clear peak, whereas values of L are extended as a band over the entire range (e.g. *viru – virů – víru – vírů/výrů*).

Regarding the speakers, intelligibility based on target words ranged from 0.5–0.7, except for one speaker who achieved the level of 0.8. In the aggregate of success score, no difference between proficient speakers and independent speakers was noted apparent. This was in accordance with the authors' subjective impression of the style of the recordings. Some speakers, especially the less advanced, consciously tried to produce length according to the canonical form. On the contrary, it was clear in case of some speakers that they did not make any special effort.

Conclusion

According to the perception test, the Polish speakers were relatively successful in realizing the Czech vowel length. The volume of errors was equal for both short and long vowels. Of the analyzed patterns, a pattern with two long vowels caused most difficulties. Conscious effort may contribute to correct pronunciation. In the following research, we plan to administer the perception test to the Bohemian listeners as well, i.e. those with a different dialectological background. Furthermore, the relationship between vowel length and foreign accent may be useful to examine.

Acknowledgement

This research was supported by the Czech Science Foundation project No. 18-18300S "Phonetic properties of Czech in non-native and native speakers' communication". We would like to thank Radek Čech (University of Ostrava) for the opportunity to administer the perception test.

Table 1: Intelligibility of target words regarding the patterns. Columns: perceived variants – number of assessments. In %. S – short vowel, L – long vowel.

Original/Perception	**SS**	**SL**	**LS**	**LL**	Sum
SS	**62.2**	19.2	15.8	2.8	100
SL	11.2	**87.2**	0.0	1.6	100
LS	1.3	5.8	**82.2**	10.7	100
LL	1.0	6.6	**46.0**	**46.4**	100

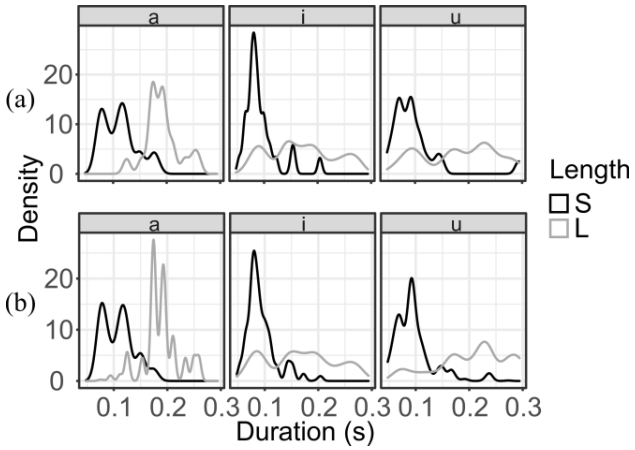


Figure 1: Distribution of normalised vowel durations split into S/L groups produced by L2 speakers (a) as written in the original text and (b) as perceived by L1 Czech listeners.

References

[1] Boersma, P. & D. Weenink (2020). '*Praat: doing phonetics by computer*' [Computer program]. Version 6.1.10 (2020). http://www.praat.org/.

[2] Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR), https://www.coe.int/en/web/common-european-framework-reference-languages, last accessed 2020/06/10.

[3] Munro, M.J. & T. M. Derwing (1995). 'Foreign accent, comprehensibility, and intelligibility in the speech of second language learners.' In *Language Learning* 49(1), pp. 73–97.

[4] Podlipský, V.J., R. Skarnitzl & J. Volín (2009). 'High front vowels in Czech: a contrast in quantity or quality?' In: Proceedings of Interspeech, vol. 2009, pp. 132–135.

[5] Rojczyk, A. (2019). 'Quality and duration of unstressed vowels in Polish.' In *Lingua* 217, pp. 80–89.

[6] Thomson, R. (2018). 'Measurement of accentedness, intelligibility, and comprehensibility'. In: *Assessment in second language pronunciation.* Ed. by O. Kang & A. Ginther, pp. 11–29. London and New York: Routledge.