

Book of abstracts

SpeakVar Workshop:

A workshop on intraspeaker and interspeaker variability



The workshop is organized within the project FK128814 sponsored by the National Research, Development and Innovation Office, Hungary.

October 2–4, 2023, Budapest

Editor: Gráci, Etelka Tekla
DOI: <https://doi.org/10.18135/SpeakVar.2023>

Contents

Cécile Fougeron: <i>Looking at intra-speaker variation on several dimensions in French</i>	5
Kirsty McDougall: <i>Probing the phonetic bases of voice similarity: Implications for voice paradises</i>	6
Andrea Deme, Kornélia Juhász, Zsuzsa Szánthó, Szabina Zsoldos and Reinhold Greisbach: <i>Duration of consonants and long and short vowels as a function of speech rate in Hungarian</i> ...	11
Andrea Balázs, Bence Kas, Ildikó Tóth and Krisztina Lakatos: <i>Early language development is related to infant temperament, gender and gestational age</i>	13
Tamás Gábor Csapó: <i>Intraspeaker and interspeaker variability of ultrasound tongue images during speech production</i>	15
Ákos Gocsál: <i>A longitudinal study of a Hungarian newsreel announcer's articulation rate</i>	18
Anneliese Kelterer, Dina El Zarka, Michele Gubian and Barbara Schuppler: <i>An analysis of speaker variation in F0 and intensity contours with the combination of FPCA and LMER</i>	21
László Fejes: <i>Variability in Erzya conjugation</i>	24
Judit Bóna: <i>Intraspeaker and interspeaker variability of temporal variables in young and elderly speakers' speech across four speaking tasks</i>	27
Tekla Etelka Grácsi, Anna Kohári: <i>Age-related and between-speaker variability of the spectral features of voiceless sibilants in Hungarian</i>	29
Sarah Melker: <i>Ageing, identity, and cross-linguistic voice comparison</i>	32
Zoltán G. Kiss and Zsuzsanna Bárkányi: <i>Multilevel variation in multilingual speech</i>	35
Veronika Harmati-Pap, Bence Kas, Ildikó Tóth and Noémi Vadász: <i>Lexical and morphosyntactic variance and its background factors in maternal narrative infant-directed speech</i>	39
Heather Weston: <i>Physical activity as a source of intraspeaker variability</i>	42
Ágnes Hámori and Judit Bóna: <i>Age patterns and individual variability in turn taking: timing and strategies in the conversation of 5- and 9-year-old children</i>	45
Katalin Mády: <i>Does a more complex phoneme system bias phonological decisions? – dialectal variation in Hungarian vowel harmony</i>	49

Tekla Etelka Grácsi, Anna Huszár, Alexandra Markó: <i>Filled pauses as individual speech characteristics?</i>	52
Varga Mónika: <i>Beszélőn belüli és beszélők közötti variáció a múltban? – Jelenségcsoportok és módszerek lehetőséges szerepéről a 18. századi megnyilatkozók azonosításában</i>	55
Hunyadi-Kóbor Zita és Tar Éva: <i>A szóvariabilitás sajátosságai 4;6-4;11 éves, tipikus beszédfejlődésű és beszédhanghibák tüneteit mutató gyermekek beszédprodukcójában</i>	57
Jankovics Julianna: <i>Időzítés a spontán beszédben – Tanulásban akadályozott fiatalok közötti eltérések</i>	60
Deme Andrea és Juhász Kornélia: <i>A hiátustöltő és fonemikus [j]-variánsok akusztikai megvalósulása valódi szavakban a nyelvállás akusztikai vetületének és a vokalikus szakaszuk időtartamának szempontjából</i>	64
Főző Eszter: <i>A beszédsajátosságok változása 10 év elteltével – A beszédleíratok lexikai vizsgálata és az eredmények felhasználhatósága a beszélőazonosítás műveletében</i>	66
Tatár Zoltán: <i>Automatikus beszélőprofil készítése</i>	69
Markó Alexandra: <i>A zöngésségi hasonulás a nyugati nyelvjárásterületen – egy esettanulmány tanulságai</i>	71

LOOKING AT INTRA-SPEAKER VARIATION ON SEVERAL DIMENSIONS IN FRENCH

Cécile Fougeron

Laboratoire de Phonétique et Phonologie (CNRS & U. Sorbonne Nouvelle)

cecile.fougeron@sorbonne-nouvelle.fr

Our knowledge on speech production is mainly based on group data, where speakers are recorded a limited number of times and grouped together to obtain a global ‘averaged’ pattern. Although, there is a general recognition that speech varies from speaker to speaker and from token to token, variations between speakers and within a speaker’s speech are often treated as random variables and are poorly understood.

In this presentation, I will focus on variation in a speaker’s speech, i.e. intra-speaker variation. A better understanding of the sources and the extent of intra-speaker variation is particularly important when we use speech to infer information about an individual, usually from a single recording, such as speaker identity in forensic phonetics or speaker condition in clinical phonetics. It is also important for our fundamental knowledge of the speech production mechanism when we consider the range of adaptation a speaker is able - or not - to make in the different contexts (situational, emotional, interactional) in which speech is used in real life. In this talk, I will present ongoing studies in French that aim to document how and to what extent speakers' speech varies within a single session or between different sessions with various delays. Intra-speaker variation is investigated on several speech dimensions and is compared to inter-speaker variation.

PROBING THE PHONETIC BASES OF VOICE SIMILARITY: IMPLICATIONS FOR VOICE PARADES

Kirsty McDougall

University of Cambridge

kem37@cam.ac.uk

Introduction

A voice parade is similar to a visual identity parade, only rather than viewing a line-up of faces, the ‘earwitness’ is played a series of speech recordings including the voice of the suspect and a number of ‘foil voices’ from which they attempt to pick out the voice that they heard at the crime scene. In England and Wales, voice parades are currently conducted according to guidelines published in 2003 (Home Office 2003). The procedure has been successfully implemented in a number of cases (e.g. de Jong-Lendle *et al.* 2015), although its use is relatively rare (Robson 2017). According to the guidelines, a voice parade should consist of nine voice samples, each containing 60 seconds of short extracts (whole utterances) of spontaneous speech edited together to give an overall impression of the speaker’s voice. The procedure is continuing to evolve with developments in technology and research understanding (e.g. McDougall 2013). Understanding voice similarity — that is, why listeners perceive some voices to sound more similar to each other than others — is essential when it comes to choosing suitable speakers to act as foils in a voice parade. In order to provide a fair comparison against a suspect’s voice, foil voices should be perceived by lay listeners as sounding relatively similar to the suspect and each other with none of the line-up of speakers markedly standing out from the rest (Nolan and Grabe 1996). Research into listeners’ perception of voice similarity is limited and the phonetic correlates of voice similarity are under-explored (cf. Nolan *et al.* 2011 and references therein). The present paper will outline some research into the phonetic underpinnings of perceived voice similarity conducted on accents of British English.¹

Voice Similarity: An Experiment on Standard Southern British English

When listeners perceive voices as sounding similar to each other, there may be two sources of similarity contributing to this percept: linguistic factors and personal factors. Linguistic factors concern the language or dialect and accent(s) used by the speakers (Nolan *et al.* 2011). Personal factors are related to speakers’ anatomies and the individual choices they make to execute the movements for speech. As a first step towards understanding these two sources of speaker similarity and their interrelationship, the approach of the work presented here is to hold linguistic factors constant and examine personal factors, by considering groups of speakers with the same linguistic and demographic background.

The *DyViS* database (Nolan *et al.* 2009) of recordings of 100 speakers provides one such pool of speakers of the same sex (male), age (18-25 years), education level (university), and accent background (Standard Southern British English, SSBE). By asking a group of listeners to judge the extent of (dis)similarity between pairs of speakers within such a matched-demographic group, the analyst can assess relationships between the phonetic characteristics of the individual speakers and patterns of perceived voice similarity. The extent of voice similarity among 15 *DyViS* speakers from such an experiment (Nolan *et al.* 2011, McDougall 2013) is shown in

¹ The author is sincerely grateful to the colleagues who have collaborated with her at various stages of the studies described here: Anil Alexander, Nathan Atkinson, Gea de Jong-Lendle, Peter French, Linda Gerlach, Erica Gold, Philip Harrison, Toby Hudson, Finnian Kelly, Christin Kirchhübel, Nikolas Pautz, Alice Paver, Elaine Schmidt, Harriet Smith, Louisa Stevens, and especially Francis Nolan. The most recent work was conducted within the IVIP project ‘Improving Voice Identification Procedures’ which is funded by the UK Economic and Social Research Council, reference ES/S015965/1, <https://www.phonetics.mml.cam.ac.uk/ivip/>.

Figure 1A, a plot resulting from applying multidimensional scaling (MDS) to the pairwise judgements of 20 British English-speaking listeners. Listeners had been asked to judge the similarity of all pairings of the 15 speakers on a nine-point Likert scale from 1 (very similar) to 9 (very different). The stimuli were two 3s samples of spontaneous speech per speaker. The first two dimensions from the MDS are shown on the graph: the closer the datapoints for any two speakers, the more similar-sounding they were judged by the listeners.

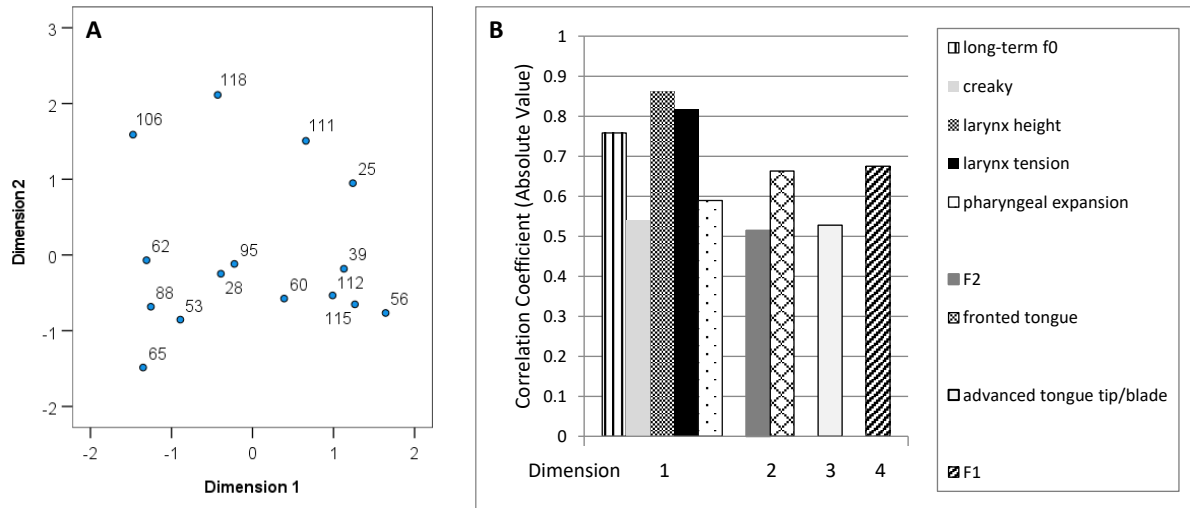


Figure 1. A: Plot of the first two of five dimensions produced by the MDS, showing perceived distances among the 15 speakers (labelled with DyViS speaker numbers). B: Significant correlations ($p < 0.05$) between the five MDS dimensions and the phonetic properties tested.

In order to determine whether the voice similarity judgements had any underlying phonetic basis, the correlation between the MDS dimensions and a number of phonetic properties (f0, long-term average formant frequencies (LTF) F1-F4, articulation rate, and voice quality features using Vocal Profiling Analysis [VPA, Beck 2005]) was tested. Each phonetic feature that yielded a significant correlation with a given MDS dimension is shown in Figure 1B. The phonetic properties most important for perceived voice similarity are those correlating with Dimension 1, that is, mean f0 and the VPA features creaky voice, larynx height, larynx tension and pharyngeal expansion. These features are all linked to the larynx and pharynx. Dimension 2 is correlated with LTF F2 and fronted tongue, Dimension 3 with advanced tongue tip/blade, and Dimension 4 with LTF F1. So for this group of SSBE male speakers, one might propose that a key element of perceived voice similarity relates to the larynx and pharynx (Dimension 1), while factors relating to supralaryngeal configuration are also important (Dimensions 2/3/4). Articulation rate did not yield a significant correlation in this experiment.

Voice Similarity Within and Between Different Accents

The above study of 15 SSBE speakers (henceforth ‘SSBE1’) has been extended to examine multiple groups of speakers, each homogeneous for accent and demographic within-group, but with different accents between the groups (McDougall 2021). Two further groups from DyViS (henceforth ‘SSBE2’ and ‘SSBE3’) were added to the SSBE1 group to enable a within-accent comparison, as well as three further groups, one each from York, Bradford and Wakefield in Yorkshire, England. There were 15 speakers in each group, also all male and aged 18-25 years (SSBE, York) or 18-30 years (Bradford, Wakefield). The speech was taken from YorViS (McDougall *et al.* 2015) for York, and WYRED (Gold *et al.* 2018) for Bradford and Wakefield, both databases using the same speaking tasks as DyViS. The format of the experiment for the five new speaker groups was the same as the previous experiment. 100 different listeners (20

per speaker group) judged the similarity of the voices within one of the five new groups. In the acoustic analysis, long-term f0, LTF F1-F4, and articulation rate were measured for each speaker.

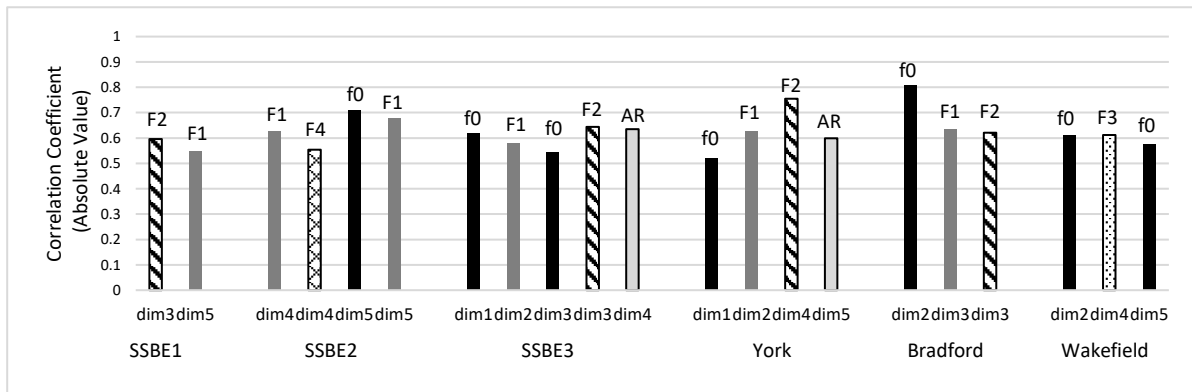


Figure 2. Significant correlations (absolute value; $p < 0.05$) between the five MDS dimensions (dim1-5) and the acoustic-phonetic properties tested for the six speaker groups.

For each group of speakers, the acoustic-phonetic measures which yielded significant correlations with an MDS dimension are shown in Figure 2.² Long-term f0 plays an important role in listeners' judgement of voice similarity, frequently occurring as the feature correlating with the lowest (i.e. Dimension 1) or one of the lowest dimensions in each analysis. LTF measures produce significant correlations in different ways across the six groups. LTF1 reveals a significant result for at least one dimension for all groups except Wakefield, while LTF2 appears for all but SSBE2 and Wakefield. LTF3 only correlates with a higher dimension (4) for Wakefield and likewise for LTF4 in SSBE2. Articulation rate is significant for SSBE3 and York only, and only in higher dimensions (4 and 5, respectively). In overview, within the SSBE accent groups, relationships between acoustic features and perceived voice similarity are similar, but not identical, with f0 and F1 featuring predominantly. The different Yorkshire groups show a consistent role for f0, with variable results for LTFs and articulation rate.

The Impact of Sample Duration on Voice Similarity Judgements

A further question with respect to perceived voice similarity in the study above is whether, at 3s, the brevity of the stimuli allows sufficient time for the long-term features analysed to become established for the listeners making each judgement. To follow this up, McDougall *et al.* (2022) conducted an experiment using 10s speech stimuli, focusing on the 15 speakers in the SSBE2 group. Since the longer stimuli increased the size of the task, the 120 speaker pairings requiring judgement were randomly divided into four subsets, and 20 listeners were allocated to each subset. Normalisation processes were applied both to the existing judgements of the 3s stimuli for SSBE2 and the new judgements. Correlations between the newly determined MDS dimensions and the speakers' long-term f0, LTF F1-F4 and articulation rate were calculated. For both 3s and 10s stimuli, f0 played a key role in voice similarity judgements. LTF measures gave relatively consistent correlation patterns in higher MDS dimensions. Articulation rate did not show any significant correlation for either 3s or 10s samples. Although perceived voice similarity appears to be somewhat variable, the new analysis did not highlight different phonetic features emerging as important when longer stimuli were used.

² Note that the results for SSBE group 1 differ from those presented earlier due to use of different MDS procedures and correlation formulae: McDougall (2013) used ALSCAL MDS and Spearman correlation (due to the non-continuous scales of VPA variables); McDougall (2021) used INDSCAL MDS and Pearson correlation.

Automatic Assessment of Perceived Voice Similarity?

A further direction for this work on perceived voice similarity has been to examine relationships between listeners' judgements and ASR-based (Automatic Speaker Recognition) assessments of the similarity of speakers (Gerlach *et al.* 2020, 2023). In Gerlach *et al.* (2023), the listener judgements of the similarity of speakers in the six groups described in the McDougall (2021) experiment above were tested for their correlation with ASR assessments produced by a number of 'sessions' available in the software VOCALISE (Kelly *et al.* 2019). VOCALISE provides automatic approaches to speaker modelling using different combinations of x-vectors, i-vectors and automatically extracted phonetic features. In this study, comparison scores using two samples (~4 mins) per speaker were generated by VOCALISE and used to indicate the similarity of pairs of speakers. Correlations between ASR comparison scores and listener judgements were highly significant for all six speaker groups. Considerable further work on the detail of relationships between ASR output and listener perception is needed, but these initial results raise the encouraging prospect that automatic techniques could be a useful resource to draw on in predicting voice similarity relationships among speakers.

Implications for Voice Parades and Future Directions

The findings outlined here have important implications for the construction and execution of voice parades. Understanding listeners' perception of similarity relationships between voices is crucial to the forensic phonetician in selecting the foil voices for a voice parade. A knowledge of which voices sound more typical or more distinctive to listeners will also be helpful in predicting what kind of voice is likely to be identified well or poorly in a parade. It is clear that long-term f_0 plays a central role in perceived voice similarity and that formants also make a key contribution, but a lot of further work is needed to unpack the subtleties of the relationships between these and other phonetic features and the many other factors at play. The study of voice similarity in six accent-controlled speaker groups highlighted some differences across the four accents examined, as well as showing some variation within the main patterns for the three same-accent SSBE groups. It is possible that some of this variation is due to other factors such as the choice of speakers and utterances, given that the experiment involved only small numbers of each. The finding that the phonetic features showing significant correlations with MDS perceptual dimensions were much the same for 3s and 10s samples indicates that 3s appears sufficient for listeners to get hold of speakers' characteristics, and that analysts should proceed with using 3s samples for the experimental 'fairness' check in voice parade construction to keep the task size manageable (cf. McDougall 2013).

Amassing a body of results from experiments such as those described above will provide the basis for a model of perceived voice similarity which will lead to more efficient and more systematic decisions in the choice of foil voices. Development of a model which can use phonetic properties to predict listeners' perception of voice similarity requires considerable further research, exploring larger speaker groups, more languages and accents, and a wide range of demographic profiles. Analysis of further phonetic variables such as segmental features, disfluencies and more aspects of voice quality is also needed, alongside the long-term features focused on in the work above. The specific accent background of the listeners making the judgements is a further relevant factor in need of investigation.

The work by Gerlach *et al.* finding high correlations between voice similarity judgements and automatic assessments shows the potential for partial automatization of foil selection. While the input of an expert phonetician in maintaining control over demographic considerations, and in overseeing and checking the process, would remain essential, the opportunity for increasing efficiencies in foil selection this way is a very promising direction for this work.

References

- Beck, J. M. 2005. Perceptual analysis of voice quality: the place of Vocal Profile Analysis. In: Hardcastle, W. J. – Beck, J. M. (eds.), *A Figure of Speech: A Festschrift for John Laver* New Jersey: Lawrence Erlbaum. 285–322.
- G. de Jong-Lendle – Nolan, F. – McDougall, K. – Hudson, T. 2015. Voice lineups: a practical guide. In: *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, Paper number 0598. <http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0598.pdf>
- Gerlach, L. – McDougall, K. – Kelly, F. – Alexander, A. – Nolan, F. 2020. Exploring the relationship between voice similarity estimates by listeners and by an automatic speaker recognition system incorporating phonetic features. *Speech Communication* 124. 85-95.
- Gold, E. – Ross, S. – Earnshaw, K. 2018. The ‘West Yorkshire Regional English Database’: investigations into the generalizability of reference populations for forensic speaker comparison casework. In: *Proceedings of Interspeech 2018*, Hyderabad. 2748-52.
- Home Office (2003). Advice on the use of voice identification parades. UK Home Office Circular 057/2003 from the Crime Reduction and Community Safety Group, Police Leadership and Powers Unit.
- Kelly, F. – Forth, O. – Kent, S. – Gerlach, L. – Alexander, A. 2019. Deep neural network based forensic automatic speaker recognition in VOCALISE using x-vectors. In: *Proceedings of the International Conference on Audio Forensics 2019*, Portugal. <http://www.aes.org/e-lib/browse.cfm?elib=20477>
- McDougall, K. 2013a. Assessing perceived voice similarity using multidimensional scaling for the construction of voice parades. *International Journal of Speech, Language and the Law* 20/2. 163-72.
- McDougall, K. 2013b. Earwitness evidence and the question of voice similarity. *British Academy Review* 21: 18-21. https://www.thebritishacademy.ac.uk/documents/805/BAR_21-06-McDougall.pdf
- McDougall, K. – Duckworth, M. – Hudson, T. 2015. Individual and group variation in disfluency features: a cross-accent investigation. In: *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, Paper number 0308. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0308.pdf>
- McDougall, K. 2021. Ear-catching versus eye-catching? Some developments and current challenges in earwitness identification evidence. In: *Proceedings of XVII AISV (Associazione Italiana Scienze della Voce) Conference*, Zürich. https://www.aisv.it/StudiAISV/2021/vol_8/studiAISV_8.pdf
- Nolan, F. – McDougall, K. – de Jong, G. – Hudson, T. 2009. The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law* 16/1. 31-57.
- McDougall, K. – Paver, A. – Nolan, F. 2022. The impact of duration of speech sample on listeners’ judgements of voice similarity. Poster presented at the British Association of Academic Phoneticians Colloquium, York, 4-8 April 2022.
- Nolan, F. – Grabe, E. 1996. Preparing a voice lineup. *Forensic Linguistics* 3/1. 74-94.
- Nolan, F. – McDougall, K. – Hudson, T. 2011. Some acoustic correlates of perceived (dis)similarity between same-accent voices. In: *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong. 1506-09. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Nolan/Nolan.pdf>.
- Robson, J. 2017. A fair hearing? The use of voice identification parades in criminal investigations in England and Wales. *Criminal Law Review* 1. 36-50.

Duration of consonants and long and short vowels as a function of speech rate in Hungarian

**Andrea Deme¹ – Kornélia Juhász^{1,2} – Zsuzsa Szánthó¹ -- Szabina Zsoldos¹ --
Reinhold Geisbach³**

¹Eötvös Loránd University, Hungary

²HUN-REN Hungarian Research Centre for Linguistics, Hungary

³University of Cologne, Germany

deme.andrea@btk.elte.hu, juhasz.kornelia@nytud.hun-ren.hu, zsuzsa.szantho@gmail.com,
szamboc@gmail.com, reinhold.greisbach@uni-koeln.de

Speech rate or rate of articulation corresponds to the duration of segments: faster articulation is the result of speech sounds produced shorter. However, one might expect that not each segment may be reduced to the same extent in their duration in fast speech. Due to their homogenous structure throughout the total segmental duration, and the lack of an obstruction in the oral cavity, vowels are expected to be more flexible in this sense than (prototypical) consonants (i.e., obstruents) (Kozhevnikov & Chistovich 1965, Wood 1973, Gósy 2004), which feature some kind of an obstruction in the mouth and can have a complex inner structure (as is the case with stops). Furthermore, differences are also expected according to phonemic length: it was shown that in Japanese, long vowels are affected more by speech rate than short vowels (i.e., they are reduced or lengthened in fast and/or slow speech, respectively) (Hirata 2004), while Korean showed a similar effect of speech rate in both vowel groups, and no asymmetries in this sense (Magen & Blumstein 1993). In these languages, phonemic length contrast is expressed primarily by durational differences.

In Hungarian, vowel length is also phonologically distinctive. Traditionally, it is assumed that the phonological length opposition is implemented phonetically as a durational and spectral difference in low vowels (/ɛ/ vs. /e:/ and /ɒ/ vs. /a:/), but in higher (more close) vowels, only a duration difference can be found (Gósy 2004). With respect to the effect of speech rate on vowel and consonant durations, and the duration of phonologically long and short vowels at fast and normal speech rates in Hungarian, we find no replicable and systematic analyses on a greater number of speakers that are also reported in detail. However, we have some evidence that the above outlined durational asymmetries are at work, that is, i) consonants are more resistant to speech rate effects (i.e., vowels reduce more in fast speech than consonants) (Magdics 1969), and ii) long vowels are affected more by speech rate than short vowels (i.e., long vowels reduce to a higher degree than short vowels) (Magdics 1969, Mády 2008). In the present study, we investigate these two hypotheses in acoustic data obtained in real words, and we involve a greater number of speakers. Additionally, we also examine if the ratio of long and short vowel's duration is maintained across different speech rates, and test if phonological length opposition is invariant as a function of tempo.

We analyze CVC shaped real words in the production of 15 Hungarian speaking females. In these sequences, V was one of the following 6 vowels that constitute long-short vowel pairs in Hungarian: /u/, /u:/, /i/, /i:/, /ɒ/, or /a:/ (in the last pair, durational difference is also accompanied by quality distinction). In the onset, we placed laryngeal or alveolar consonants: /h z s t r/. In the coda, alveolar consonants were positioned: /z t d k n r/. As a result, target sequences did not constitute minimal pairs, hence probably did not facilitate exaggeration of contrastive features of segments (e.g., vowel length). Speakers produced target words in carrier sentences, where the target word bore sentence level accent: *Legyen <target word>! Let it be <target word>!* We recorded samples in two speech rate conditions: at i) comfortable speech rate ("normal" speech), and ii) maximum speech rate ("fast" speech). Maximum speech rate was achieved by the method of Greisbach (1992). Speakers repeated each target sentence several times. They

always started with a comfortable tempo (marked as normal speech later on in the analysis), and then, followed by a silent pause, they started to repeat the same item several times. Speakers were instructed to try to speak faster and faster at each repetition, until articulation broke down or speakers ran out of air. Each participant produced 6 of these sets (i.e., one normal rate variant followed by fast repetition variants) for each target word resulting in 72 sets per speaker in total. We labeled all sets manually in Praat (Boersma & Weenink 2022): we segmented each word, checked their durations, and labeled the shortest repetition as fast speech (while normal speech was always the first item produced at a comfortable speech rate, as mentioned). We segmented speech sounds in all of the normal and fast variants in each set. We analyze and compare duration of vowels and consonants in the two speech rate conditions. We also analyze the difference, and ratio of long and short vowel pairs in the different conditions.

We expect that vowels reduce in their duration more than consonants, and that long vowels show more shortening effect of increased speech rate than short vowels. Results are hoped to contribute to our better understanding of how phonological features are implemented in the phonetic realization of speech, and how reduction of segmental features takes place.

Acknowledgements

The research was supported by the TKA–DAAD grant No. 177375., the NKFIH grant No. FK128814, and the ÚNKP-23-3-I-ELTE-335 grant (Zs. Sz.).

References

- Boersma, P. & Weenink, D. 2022. *Praat: doing phonetics by computer* [Computer program]. Version 6.3. <http://www.praat.org/>
- Gósy, M. 2004. *Fonetika, a beszéd tudománya*. Osiris Kiadó, Budapest
- Greisbach, R. 1992. Reading aloud at maximal speed. *Speech Communication* 11. 469–473.
- Hirata, Y. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32, 565–589.
- Kozhevnikov V.A. & Chistovitch L. A. 1965. *Speech articulation and perception*. Joint Publications Research Service, Washington.
- Mády, K. 2008. Magyar magánhangzók vizsgálata elektromágneses artikulográffal normál és gyors beszédben. *Beszédkutatás*. 52–66.
- Magdics K. 1969. A magyar beszédhangok időtartama nyugodt és gyors beszédben. *Nyelvtudományi Értekezések* 67. 45–63.
- Magen, H. S. & Blumstein, S. E. 1993. Effects of speaking rate on the vowel length distinction in Korean. *Journal of Phonetics* 21. 387–409.
- Wood, S. 1973. What happens to vowels and consonants when we speak faster? *Working papers Lund University, Department of Linguistics and Phonetics* 9. 8–39.

Early language development is related to infant temperament, gender and gestational age.

Andrea Balázs(1)(3), Bence Kas(1)(2), Ildikó Tóth(3), Krisztina Lakatos(3)

(1) HUN-REN Hungarian Research Centre for Linguistics, Institute for General and Hungarian Linguistics, Budapest, Hungary; (2) Eötvös Loránd University Bárczi Gusztáv Faculty of Special Needs Education, Budapest, Hungary; (3) Research Centre for Natural Sciences, Cognitive Psychology and Neuroscience Institute, Budapest, Hungary

An important question of recent research on early language development is the role of biological and cognitive factors that regulate and influence infant and toddler behavior.

Early language development shows great individual differences both in terms of the extension and expansion of receptive and expressive vocabulary. Research showed that the range of a 12-month-old, typically-developing child's receptive vocabulary may span from 25 to more than 200 words. Similar variation can also be observed in the development of expressive vocabulary. (Fenson et al., 1994). Some children say their first words by the age of 12 months while others only after 18 months. The rate of development begins to balance out by the third year of life. (Fenson et al., 1994, Kovács et al., 2018).

In order to tap into biological factors influencing language development, we studied infant temperament which is assumed to be constitutionally based, can be measured in early childhood, and shows a relatively stable pattern extending over a lifetime. These temperamental features contribute to the regulation of behavior including, affect regulation, attention direction, and motor activity (Rothbart, 1981, 2007). Infants' attentional control and the capacity for self-regulation have been reported to be associated with language development in infancy and early childhood as well. It was previously found that these temperament traits correlate with the efficiency of language acquisition, including the time of appearance of first words and the time and speed of vocabulary expansion during the first 2 years of life. (Canfield and Saudino, 2016, Dixon and Shore, 1997; Dixon and Smith, 2000). Others found a positive correlation between inhibition and expressive language (Smith, et al. 2014). Larson et al. (2020) found that inhibition skills may help to improve subsequent morphological comprehension in typically developing children and children with specific language impairment.

We conducted a longitudinal study from birth to 36 months to investigate the longitudinal relationship between infant temperament and early language development. We add further evidence to the contribution of infant temperament to early language development by identifying the aspects of temperament, combining with other biological factors present in infancy such as gender or gestational age at birth that may play a prominent role in early language development.

Data of 94 full-term infants were included in the present analyses. Temperament and early non-verbal communication skills, receptive and expressive vocabulary were measured from 6 to 36 months. Language competence was evaluated with the Hungarian version of the MacArthur-Bates Communicative Development Inventory. Temperament was assessed by the Very Short Form of Infant Behavior Questionnaire and the Very Short Form of Early Childhood Behavior Questionnaire.

Surgency is associated with more extensive use of communicative gestures and greater receptive vocabulary based on repeated language measurements between the ages of 9 and 18 months both concurrently and across ages. Effortful control is also related to better gestural communication in infancy. Surgency measured at the ages of 18 and 24 months shows a pattern of

consistent albeit weak correlations with expressive language skills repeatedly measured every two months between 18 and 30 months.

To predict language development, regression models including infant sex, gestational age, and temperament were tested. At 18 months, receptive vocabulary is predicted by infant surgency, gestural communication by infant gender, surgency and effortful control. Expressive vocabulary at 18 months was only predicted by gestational age at birth, at 24 months by infant gender and gestational age, and at 30 months by gestational age.

According to our results, different dimensions of temperament, infant gender and gestational age are related to aspects of communicative development in infancy. Infants and toddlers with higher surgency might enter into communicative situations more readily and show more engagement with adult social partners which is favorable for communication development. Effortful control may aid the learning process by supporting attentional control. In toddlerhood gestational age seems to be the predictor for expressive vocabulary. Gestational age was identified as a predictor for language in preterm infants previously. Our results extend this association to the smaller time window in the case of full term infants and describe a shift in predictors according to age and aspect of language development. Gestational age may mark prenatal developmental processes that may exert influence on the development of verbal communication at later, more developed stages.

References

- Canfield, C. F. – Saudino, K. J. 2016. The influence of infant characteristics and attention to social cues on early vocabulary. *Journal of Experimental Child Psychology*, 150, 112-129.
- Dixon Jr, W. E. – Shore, C. 1997. Temperamental predictors of linguistic style during multiword acquisition. *Infant Behavior and Development*, 20(1), 99-103.
- Dixon Jr, W. E. – Smith, P. H. 2000. Links between early temperament and language acquisition. *Merrill-Palmer Quarterly* (1982-), 417-440.
- Fenson, L. – Dale, P. S. – Reznick, J. S. – Bates, E. – Thal, D. J. – Pethick, S. J. – Stiles, J. 1994. Variability in early communicative development. *Monographs of the society for research in child development*, pp. 1–185.
- Kovács, Z. – Kas, B. – Pintye, M. 2018. Szempontok a nyelv-és beszédfejlődési zavarok szűréséhez és állapotmegismeréséhez.
- Larson, C. – Kaplan, D. – Kaushanskaya, M. – Weismer, S. E. 2020. Language and inhibition: predictive relationships in children with language impairment relative to typically developing peers. *Journal of Speech, Language, and Hearing Research*, 63(4), 1115-1127.
- Rothbart, M. K. 1981. Measurement of temperament in infancy. *Child development*, 569-578.
- Rothbart, M. K. 2007. Temperament, development, and personality. *Current directions in psychological science*, 16(4), 207-212.
- Smith Watts, A. K. – Patel, D. – Corley, R. P. – Friedman, N. P. – Hewitt, J. K. – Robinson, J. L. – Rhee, S. H. 2014. Testing alternative hypotheses regarding the association between behavioral inhibition and language development in toddlerhood. *Child development*, 85(4), 1569-1585.

Intraspeaker and interspeaker variability of ultrasound tongue images during speech production

Tamás Gábor Csapó¹

¹Budapest University of Technology and Economics
csapot@tmit.bme.hu

Introduction

When investigating articulation during speech production, all of the articulatory tracking devices are highly dependent on the session and on the speaker. As an example, ultrasound tongue imaging recordings from two separate sessions (by ‘session’ we mean that the probe fixing headset is dismounted and mounted again to the speaker) are nearly always misaligned and such recordings are not directly comparable. However, speech technology research in this field (articulatory-to-acoustic mapping and acoustic-to-articulatory inversion) would need large amounts of aligned data. Another problem is that the imaging quality of ultrasound is highly dependent on the speaker (see Fig. 1), i.e. for some speakers longer part is visible from the tongue contour, or the tongue surface is less blurry than with other speakers. The quality of the images is influenced by many factors, such as the anatomy of the speaker or the condition of the tissues of the articulatory organs (e.g., hydration). The variation between speakers may also be due to the fact that the ultrasound transducer is positioned differently (in different orientations) for different head sizes. The recording software usually provides the possibility to adjust the ultrasound hardware parameters (e.g., transducer frequency, field of view, depth, dynamic range, line density, etc.), but this may not be a sufficient solution for all speakers. Overall, the possibility of comparing speakers is limited due to potentially different orientations. Another difficulty for the comparison might be the varying speed of articulation of different speakers, which is the main focus of the current study.

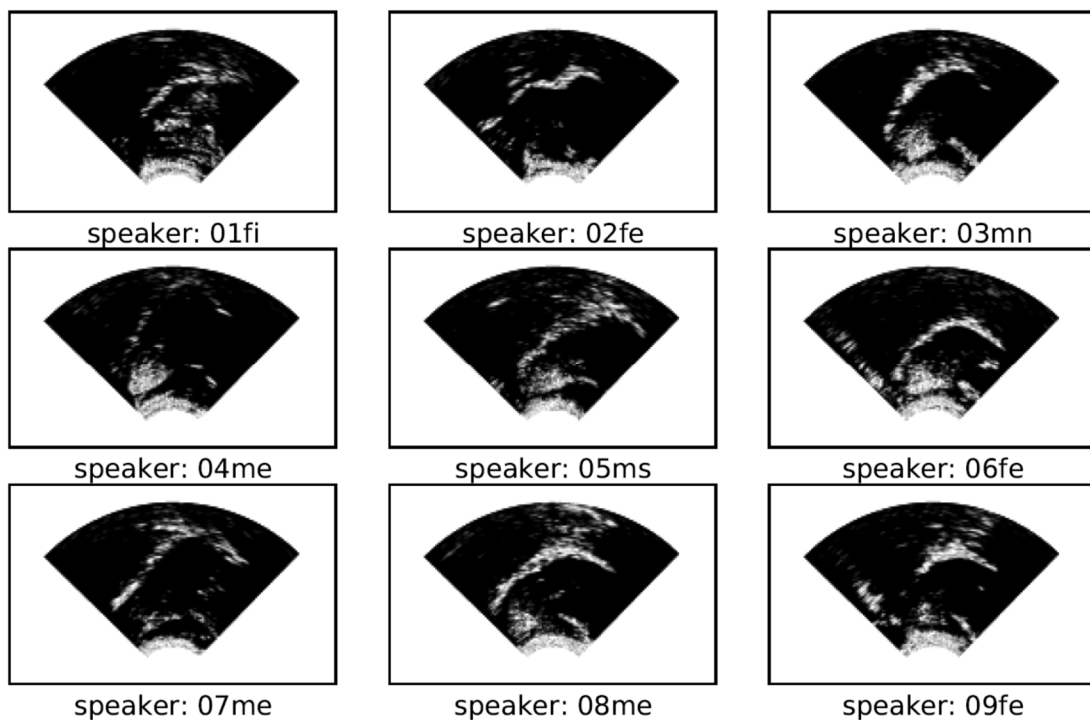


Figure 1. Examples of the differences in the quality of ultrasound tongue images between speakers from the UltraSuite-Tal80 database.

In the international research field, there is a lack of methods for the alignment of ultrasound tongue imaging data recorded in different sessions or with different speakers, or to time-align the articulatory data.

Methods

In this research, we investigate the above questions and analyze the session and speaker dependency of articulatory movement using ultrasound tongue imaging. Especially, we focus on the inter- and intraspeaker differences in time, and try to compare multiple data of different lengths.

Existing databases were used to investigate the ultrasound tongue images; recordings were selected from native speakers of Hungarian and English. The Hungarian recordings were created for the previous research on articulatory-to-acoustic mapping (Csapó et al. 2019). For English data, the UltraSuite-TaL80 database was used (Ribeiro et al. 2021; https://ultrasuite.github.io/data/tal_corpus/).

The difficulty with ultrasound tongue image sequences is that as they are videos, they are relatively difficult to visualize in 2D images. One way of showing articulatory information is a 'kymogram' (Lulich et al., 2018) : a kind of 'articulatory signal over time': the middle slices of the ultrasound tongue images were cut (approximately corresponding to the middle of the tongue) and plotted as a function of time, similarly to a spectrogram. From two speakers of the Hungarian dataset, we selected one sentence for demonstration purposes, which occurred in both speakers' recordings. Similarly, we selected multiple speakers from the UltraSuite-TaL dataset. Based on the speech signal, we computed MFCC spectrogram. Based on the ultrasound signal, we plotted kymograms. Dynamic Time Warping (DTW) is a long-established method for comparing speech samples of different lengths – here we calculate DTW paths on the speech signal, and use that information to time-align articulatory data of multiple sessions or speakers.

Results

Example spectrograms and kymograms of two speakers of the Hungarian dataset are shown in Fig. 2. In the spectrograms (1st and 2nd subfigure), it can be seen that the speech sample took about 6.2 seconds to be uttered by one speaker and about 4.8 seconds by the other speaker, i.e. their articulatory speeds are different. In the kymograms (3rd and 4th subfigure), the articulatory landmarks / inflexion points appear at different locations for the two speakers. For example, the sentence starts with the Hungarian back vowel 'ɔ' and continues with a front vowel 'e', followed by 'ɔ' again. Therefore, back-front movement of the tongue is visible on the 3rd and 4th subfigures roughly around 0.6-1.2 second. The 5th subfigure shows the result when the speech of the two speakers is DTW-aligned and the articulatory data is stretched for the second (higher articulatory speed) speaker to match the first (lower articulatory speed) speaker. Thus, articulatory data of the two speakers were successfully time-aligned, and the landmarks / inflection points of tongue movement are at similar locations in the DTW-aligned kymogram. Earlier, Jayanthi et al. (2017) were applying DTW to time-align speech and articulatory data, using electromagnetic articulography. Following that work, in this study, we are using kymograms and DTW to compare and time-align multiple ultrasound tongue image recordings from the same speaker, and across speakers. In the talk, we will include more examples of such intra- and interspeaker articulatory differences and similarities.

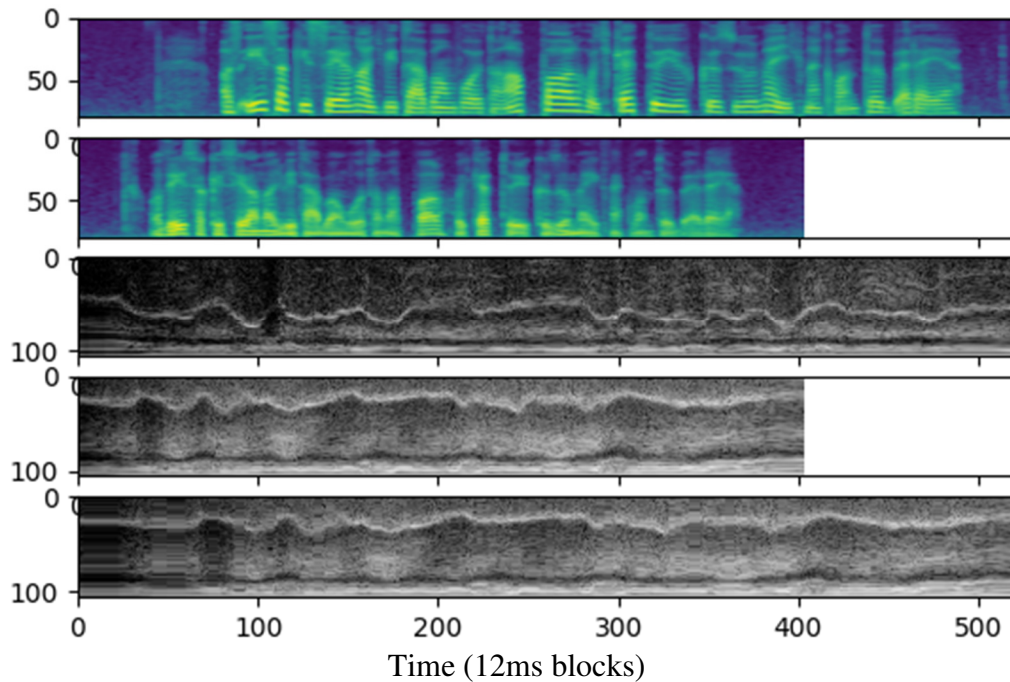


Figure 2. Results: spectrogram and ‘kymogram’ (temporal change of the midline of the ultrasound tongue images), from two speakers of the Hungarian dataset. The sentence is: “Az északi szél nagy vitában volt a Nappal, hogy kettőjük közül melyiknek van több ereje.”

Future plans

Regarding the potential application of the above results, we plan to use the results for speech-based brain-computer interfaces to supplement the brain signal (measured with EEG, ECoG or sEEG) with ultrasound tongue image based articulatory information.

References

- Csapó, Tamás Gábor, Mohammed Salah Al-Radhi, Géza Németh, Gábor Gosztolya, Tamás Grósz, László Tóth, and Alexandra Markó. 2019. “Ultrasound-Based Silent Speech Interface Built on a Continuous Vocoder.” In Proc. Interspeech, 894–98. Graz, Austria.
- Jayanthi, Sai Muralidhar, Lucie Menard, and Catherine Laporte. 2017. “Divide-and-Warp Temporal Alignment of Speech Signals between Speakers: Validation Using Articulatory Data.” In Proc. ICASSP, 5465–69. New Orleans, LA, USA.
- Lulich, Steven M, Kelly H Berkson, and Kenneth de Jong. 2018. “Acquiring and Visualizing 3D/4D Ultrasound Recordings of Tongue Motion.” *Journal of Phonetics* 71: 410–24.
- Ribeiro, Manuel Sam, Jennifer Sanger, Jing-Xuan Xuan Zhang, Aciel Eshky, Alan Wrench, Korin Richmond, and Steve Renals. 2021. “TaL: A Synchronised Multi-Speaker Corpus of Ultrasound Tongue Imaging, Audio, and Lip Videos.” In 2021 IEEE Spoken Language Technology Workshop (SLT), 1109–16. Shenzhen, China.

A LONGITUDINAL STUDY OF A HUNGARIAN NEWSREEL ANNOUNCER'S ARTICULATION RATE

Gocsál, Ákos

University of Pécs, Faculty of Music and Visual Arts
HUN-REN Hungarian Research Centre for Linguistics

gocsal.akos@pte.hu

1. Introduction

There is a growing interest among speech researchers in studying the temporal changes of the acoustic structure of speech as speakers grow older. Literature suggests that there are two main approaches to researching this topic. The first approach uses recordings that were not made for research purposes, i.e. media recordings, such as radio presentations. One prominent example of this approach is provided by Reubold, Harrington & Kleber (2010) who studied the voices of Queen Elizabeth II and other well-known British personalities. In the Hungarian literature, Balázs (1993) studied the voices of a Hungarian actor and an actress, using radio recordings. The other approach uses speech databases in which utterances by the same speakers were recorded at least at twice, several years apart. Examples include a study by Fuchs, Koenig & Gerstenberg (2021), who used a database of French speakers' interviews, or Grácz et al. (2020) who described a database for longitudinal speech research. Several results have already been published using this database (Grácz & Krepsz, 2020; Markó et al., 2021). Both approaches have their own advantages and disadvantages. Previous recordings may be available for a wide time range, but due to the lack of a researcher's control, recording conditions may be very different, which may distort the behaviour of the observed parameters. The professionally designed databases may eliminate this problem, but it is a major limitation that a considerable amount of time is required between the two or more recordings. The present study belongs to the first type of longitudinal speech research. The speech material of the Hungarian newsreels offers opportunities, so far unexploited, for longitudinal research. Of the great variety of possible speakers, here we choose the news announcer, whose voice appeared in the widest time range in the newsreels and examine how this announcer's articulation rate changed over time. It is hypothesized that the speaker's articulation rate got slower as the speaker grew older.

2. Methods

2.1. Selection of the speaker

All Hungarian newsreels are available online at www.filmhiradokonline.hu and they can be freely used for educational and research purposes. For the present research, we define "newsreel" as a complete news program created on a weekly basis. The newsreels usually included several individual news and were not longer than 10-12 minutes. In this study, only sound newsreels are relevant, which were produced between 1931 and 1989. Although the website has a query function and the newsreel database has an extensive set of keywords, the announcers' names were not entered there, thus identification of the speakers and finding their first and last readings required extra research. Furthermore, no publication was found that could have supported the identification of the speakers, thus, the present author randomly selected newsreels and attempted to identify the speakers, based on his previous research experience with newsreels and familiarity with the announcers' voices. The author was only able to identify announcers in newsreels produced after 1948, such as György Bán, László Körmeny, and Piroska P. Debrenti. Until the mid-1960s, only a limited number of announcers appeared in the newsreels but from the second half of the 1960s, a relatively large number of other announcers' voices were used, such as Ferenc Bözsöny, Gábor Dömök, Ferenc Péter, József P. Kovács, Pál Szalóczy, András Ulbrich, and István Thoma, who were also identified. Then, using a trial-and-error approach, the author attempted to find the first and the last appearance of each speaker, thus finding the speaker whose voice appeared in the longest period of time in the newsreels. This procedure revealed that the György Bán (1918-1985) appeared between 1949 and 1970, representing the longer time span in the newsreels, so his voice was selected for the present research.

2.2. Selection of the speech samples

Starting with 1949, four years with six-year intervals between them were selected, i.e. 1949, 1955, 1961 and 1967. In these years, the speaker was 31, 37, 43 and 49 years old. From each year, seven news from different newsreels were chosen, resulting in a total of 28 speech samples. A preliminary inspection of the news revealed that there were great differences in length, short texts were thus excluded, only readings longer than six sentences were selected. Furthermore, the news stories were selected so that they represent a great variety of topics, such as industry, agriculture, public holidays, cultural life, lifestyle, natural disasters, foreign news etc., while texts including exclamations, dialogue-like utterances, or an ironic tone were excluded.

2.3. Acoustic analyses and calculations

Following the instructions of Neuberger et al. (2014), first level transcriptions (i.e. pause-to-pause labels) were made using Praat 6.5.8 (Boersma & Weenik, 2019). The duration of each pause-to-pause segment was measured,

and their total sum was calculated for each speech sample. Previous observations, however, indicate that the measured duration values need to be corrected, because films were shot at 24 frames per second, while digitalization converted them to a rate of 25 frames per second (Gocsál, 2022). This means that the digitalized version of the newsreels is somewhat faster than the speed at which they were originally shot and displayed. Duration values were therefore multiplied by 25/24 to obtain the original rate. Also, all pronounced speech sounds were counted. Articulation rate was calculated by dividing number of all speech sounds in the given speech sample by the total sum of the section durations in seconds. This resulted in seven articulation rate values per year. Pearson's correlation coefficient was used to determine if correlation exists between year and articulation rate. Friedman test and pairwise comparisons between the four sets of data were also carried out using SPSS 27.

3. Results

3.1. Descriptive statistics

The total length of the 28 speech samples with pauses excluded was 1375.36 s and 18444 speech sounds were counted. The mean length of the speech samples with pauses excluded was 49.1 s (SD = 11.03), and length ranged between 27.32 and 73.2 s. The corrected duration values were as follows. Total duration: 1432,67 s, mean length: 51.16 s (SD = 11,49), range: 28.46–76.26 s. The mean number of speech sounds was 658.71 (SD = 155.99), ranging between 368 and 1036. For calculating articulation rates, the corrected duration values were used. Mean articulation rate was 12.86 sounds/s (SD = .55), ranging from 11.98 to 13.94 sounds/s.

3.2 Association between speaker age and articulation rate

Articulation rates broken down by year are demonstrated in Figure 1. An inspection of Figure 1 suggests that there is no linear association between speaker age and articulation rate, which is confirmed by the Pearson's correlation coefficient ($r = -.154, p = .435$). Since the distribution of the data suggests that there may be a significant difference between the medians of the datasets, a Friedman test was performed. The Friedman test confirmed that such a difference exists between at least two sets of data, $\chi^2(3) = 11.229, p = 0.011$. Dunn-Bonferroni post hoc tests were conducted to compare the four sets of articulation rate values. The results of these pairwise comparisons are shown in Table 1. The results suggest that the mean articulation rate of the speech samples recorded in 1955 is significantly lower than that recorded in 1949, and the difference between the data of 1955 and 1967 approaches statistical significance with the 1955 data suggesting slower articulation.

4. Conclusions

The results do not confirm our hypothesis that the newsreel announcer spoke slower as he grew older. Moreover, the unexpectedly slow articulation rate found with the 1955 data highlight that factors other than speaker age may have significantly influenced his articulation rate. At present, only speaker age and acoustic data are available, which cannot explain the observed radical decrease in articulation rate. Temporary health issues, different recording conditions, or even producer's instructions may have influenced articulation rate. But even if the 1955 data are left out, the comparisons of the 1949, 1961 and 1967 data do not show a significant trend towards slower articulation, at least regarding the period between 31 and 49 years of age of the speaker. The results may also indicate that that this type of oral reading had an "ideal" rate of articulation of around 13 sounds/s, and the speaker, regardless of his age, tried to use this rate with some variation, except for the 1955 readings.

It was a limitation in this study that a low number of speech samples were used. Using more speech samples per each year would significantly improve the results. Furthermore, choosing speech samples from years with smaller intervals between them, or even from each year would particularly help understand if his slower speech in 1955 was only an exception for that year or representative of a tendency which lasted for years.

It also needs to be mentioned that within the same year, every effort was made to select news stories of diverse topics, but it was not possible to consistently select the same topics across the years. In 1949, the majority of the news presented was about construction and industrial production. In 1961, the focus was on agriculture, and by 1967, an increase of news about lifestyle or international politics was observed. It is possible that the topics of the news influenced tempo, but due to the inconsistency of topics across the years and the low number of news, it was not possible to confirm this. Further studies should address these issues and it is believed that sources of variation of articulation rate, presently unknown, can be thus revealed.

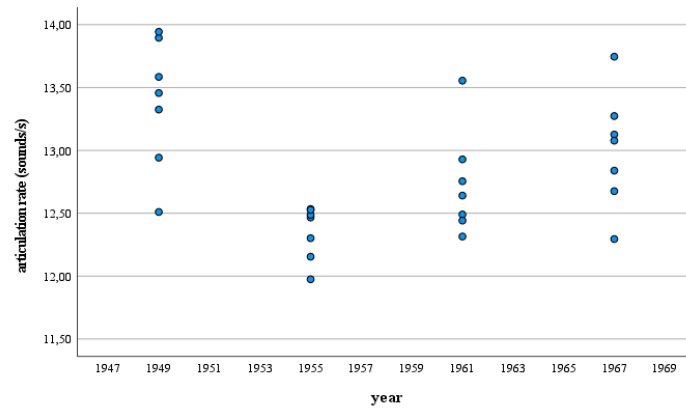


Figure 1. Articulation rates by year.
Speaker age was 31, 37, 43 and 49 years in 1949, 1955, 1961 and 1967 respectively

pair	test statistics	std. error	std. test stat.	sig.	adj. sig.
Articulation rate 1949 Articulation rate 1955	2.000	.690	2.898	.004	.023
Articulation rate 1949 Articulation rate 1961	1.429	.690	2.070	.038	.231
Articulation rate 1949 Articulation rate 1967	.286	.690	.414	.679	1.000
Articulation rate 1955 Articulation rate 1961	-.571	.690	-.828	.408	1.000
Articulation rate 1955 Articulation rate 1967	-1.714	.690	-2.484	.013	.078
Articulation rate 1961 Articulation rate 1967	-1.143	.690	-1.656	.098	.586

Table 1. Pairwise comparison of articulation rates

References

- Balázs, B. 1993. Az időskori hangképzés jellemzői. *Beszédkutatás '93.* 156–165.
- Boersma, P., & Weenik, D. 2019. *Praat: Doing phonetics by computer.* www.praat.org
- Fuchs, S. – Koenig, L. L. – Gerstenberg, A. 2021. A longitudinal study of speech acoustics in older french females: analysis of the filler particle *eah* across utterance positions. *Languages* 6/4. 211. <https://doi.org/10.3390/languages6040211>
- Gocsál, Á. (2022). Beszédtípusok és artikulációs tempóik a korai hangos filmhíradókban. In: Csárdás L. – Bóna J. (eds.): *Sokszínű beszédtudomány.* Budapest: Akadémiai Kiadó. https://mersz.hu/hivatkozas/m970sb_4/#m970sb_4
- Grácsi, T. E. – Huszár, A. – Krepesz, V. – Száraz, B. – Damásdi, N. – Markó, A. (2020). Longitudinális korpusz magyar felnőtt adatközlőkről. In: *XVI. Magyar Számítógépes Nyelvészeti Konferencia.* Szeged: JATEPress. 103–114. http://acta.bibl.u-szeged.hu/67677/1/msznykonf_016_103-114.pdf
- Grácsi, T. E. – Krepesz, V. (2020). Évek múltán a zöngé. Egyes zöngéjellelmzők változása 11 év alatt 6 férfi beszélő beszédében. In: Fóris Á. – Bölcskei A. – Bóna J. – Grácsi T. E. – Markó A. (eds.): *Nyelv, kultúra, identitás. Alkalmazott nyelvészeti kutatások a 21.századi információs térben. III. Fonetika.* Budapest: Akadémiai Kiadó. https://mersz.hu/dokumentum/m675nyki3f_32/
- Markó, A. – Huszár, A. – Krepesz, V. – Grácsi, T. E. (2021). Az alaphérfvéncia jellemzőinek longitudinális összevetése felnőtt beszélők felolvasásában. *Beszédtudomány – Speech Science* 2/1. 99–134.
- Neuberger, T. – Gyarmathy, D. – Grácsi, T. E. – Horváth, V. – Gósy, M. – Beke, A. (2014). Development of a Large Spontaneous Speech Database of Agglutinative Hungarian Language. In: Sojka, P. – Horák, A. – Kopeček, I. – Pala, K. (eds.): *Text, Speech and Dialogue.* Cham: Springer International Publishing. 424–431. https://doi.org/10.1007/978-3-319-10816-2_51
- Reubold, U. – Harrington, J. – Kleber, F. (2010). Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication* 52. 638–651.

This research was supported by the Hungarian National Research, Development and Innovation Office of Hungary, project No. FK-128814.

An analysis of speaker variation in F0 and intensity contours with the combination of FPCA and LMER

Anneliese Kelterer¹ – Dina El Zarka¹ – Michele Gubian² – Barbara Schuppler³

¹ Department of Linguistics, University of Graz

² Institute of Phonetics and Speech Processing, Ludwig Maximilian University of Munich

³ Signal Processing and Speech Communication Laboratory, Graz University of Technology

In this talk, we will present a method combining Functional Principal Component Analysis (FPCA; Ramsay & Silverman, 2005; Gubian et al., 2015) with Linear Mixed Effects Regression models (LMER; Bates et al., 2015) to investigate speaker variation in holistic prosodic contours (F0 and intensity). The phenomenon examined here is the prosodic realization of three main types of information structures (IS) in Egyptian Arabic (EA): focus-background, topic-comment and all-new. For an SVO sentence like *ħaliima najjimit amaani* “Halima put Amani to bed”, for instance, these IS types were elicited by means of a question-answer paradigm where the focus of the answer corresponds to the wh-word of the question: a.) “Amani’s fast asleep. Who put her to bed?” (focus-background; FB); b.) “What did Halima do?” (topic-comment; TC); c.) “How come it’s so quiet in here today?” (all-new; AN).

While EA does not behave like English, that is, it does not deaccent post-focus material, previous studies found that focus prominence was expressed by an expanded F0 range, and backgrounding was expressed by post-focus compression (Hellmuth, 2010, 2011; El Zarka et al., 2020). A qualitative investigation of the present data, as well as previous studies on EA (Hellmuth, 2010, 2011; Cangemi et al., 2016) also found considerable inter-speaker variation in the prosodic realisation of different IS conditions.

In the present study, we examine the speaker-specific strategies to express prominence relations between focus and background in some detail. The combination of FPCA and LMER allows us to investigate this variation in quantitative terms and on the basis of whole contours rather than individually measured features from different parts of the utterance. FPCA captures different modes of variation in the F0 and intensity curves that are represented by scores. These scores are numbers that can be input into a regression analysis to examine the relationship between the mode of variation and the IS conditions. In an LMER analysis, we can not only include speakers as random effects in addition to fixed effects, but also examine how much speaker variation contributes to the overall variation explained by the models. Moreover, LMERS provide us with individual speakers’ estimates, which can, in turn, be used to reconstruct estimated F0 and intensity curves of the three IS conditions for individual speakers. This workflow gives us a quantitative means to characterise different speakers.

The data investigated here consists of 367 SVO sentences with similar segmental structure that were elicited in the three IS types with the above-mentioned question paradigm. 18 speakers (11f, 7m) were recorded and F0 and intensity contours of the whole sentence were extracted, speaker normalized and time-normalized with respect to word and stressed syllable boundaries.

Figure 1 shows the results of the FPCA. PC1 (left) captures more (negative scores; blue) or less (positive scores; red) downstep after the first accent accompanied by a corresponding intensity. PC2 (right) captures a higher first and second peak with lower values at the end of the phrase at positive scores (red), and an overall flatter contour (lower first and second peak, less

downstep in the last word) at negative scores (blue). These changes are accompanied by a corresponding change in intensity only in the first and second word, but F0 and intensity are decoupled at the end of the phrase, i.e., changes in F0 in the third word are not accompanied by any change in intensity.

To investigate the relationship between the F0 and intensity variation captured by PC1 and PC2, respectively, with the three IS conditions, we built two LMER models predicting scores for PC1 and PC2 with Category (FB, TC, AN) as fixed effect and Speaker ($n = 18$) as random intercept with Category as random slope. Including a random slope allows us to capture not only absolute differences between speakers, but whether some speakers make a larger difference between the categories, or a difference in another direction, than others.

PC1 mostly represents Category variation: 26% of variation in this model is explained by IS Category, and FB had significantly lower values (cf., blue curves in Figure 1) than TC and AN. Moreover, 16% of variation was explained by Speaker and 14% by the random slope, indicating some speaker variation along this mode of variation (more vs. less downstep after the first accent with correlating intensity). A look at individual speakers' estimates shows that this variation is due to 4 of 18 speakers not following the general trend of more downstep and lower intensity after the first accent in FB, while the 14 speakers that follow the general trend do so to different degrees. In the model for PC2, only 7% of variation was explained by Category, 46% by the Speaker intercept and 14% by the random slope. This indicates that flatter vs. steeper F0 curves with intensity decoupling in the last word are mostly related to speaker variation. An overall significantly higher estimate for AN than FB and TC, however, indicates that AN is the category in which F0 tends to end low while intensity stays high.

From these two models, we can extract category estimates for individual speakers. With these estimated speaker scores and the information about which score corresponds to which change in the F0 and intensity curves, we can draw estimated curves for different speakers. Two speakers are presented in Figure 2.

F09 had the largest estimated difference between FB and TC in PC1, and the highest estimated score for AN in PC2. A reconstruction of her estimated F0 and intensity curves based on PC1 and PC2 is shown in Figure 2 (left). In FB (yellow), the second accent is downstepped and the F0 curve is compressed after the first accent with a lower intensity in these parts of the utterance. In TC (green), F0 and intensity stay higher after the first peak than in FB. In AN, F0 and intensity are much higher in the first two words, but F0 descends to values as low as FB while intensity stays high at the end of the phrase. This corresponds to the auditory impression of an overall decrease of prominence after focus in FB in contrast to a prominent downstepped final accent in AN. By contrast, F08 had very small differences between category estimates for PC1 and PC2 scores. These small estimate differences correlate with hardly any difference in the estimated F0 and intensity curves (Figure 2, right). This fits with the auditory impression of this speaker who does not seem to make a prosodic difference between the three IS types most of the time. In general, the study showed that the majority of speakers showed differences between TC/AN and FB. However, this distinction is realized by different strategies for individual speakers.

Bibliography

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01

- Cangemi, F., Zarka, D. E., Wehrle, S., Baumann, S., & Grice, M. (2016). Speaker-specific intonational marking of narrow focus in Egyptian Arabic. *Proceedings of the 8th International Conference on Speech Prosody*, (pp. 1-5). doi:10.21437/SpeechProsody.2016-69
- El Zarka, D., Kelterer, A., & Schuppler, B. (2020). An analysis of prosodic prominence cues to information structure in Egyptian Arabic. *Proceedings of Interspeech 2020*, (pp. 1883-1887).
- Gubian, M., Torreira, F., & Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, 49, 16–40. doi:10.1016/j.wocn.2014.10.001
- Hellmuth, S. (2010). The (absence of) prosodic reflexes of given/new information status in Egyptian Arabic. *Information Structure in Spoken Arabic*, 165-188. (J. Owens, & A. Elgibali, Eds.) Routledge.
- Hellmuth, S. (2011). Acoustic cues to focus and givenness in Egyptian Arabic. *Instrumental Studies in Arabic Phonetics*, 299-324. (Z. M. Hassan, & B. Heselwood, Eds.) Benjamins.
- Ramsay, J., & Silverman, B. (2005). *Functional Data Analysis*. Springer Series in Statistics.

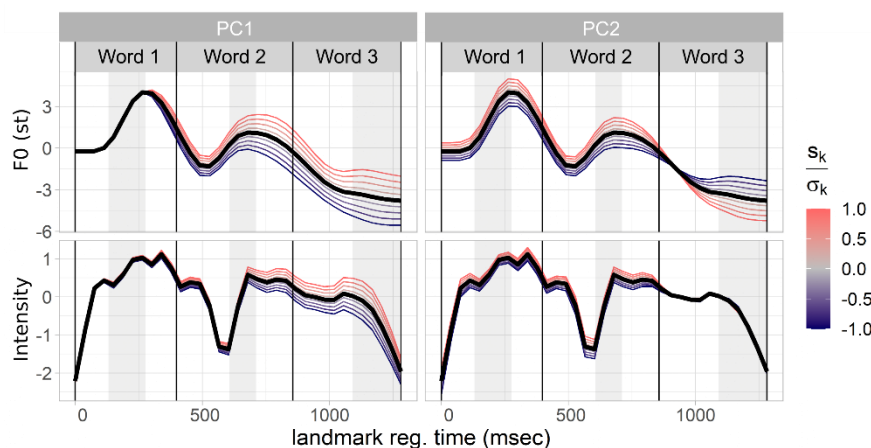


Figure 1. FPCA results for PC1 (left) and PC2 (right). Vertical lines indicate word boundaries, shaded areas indicate stressed syllables.

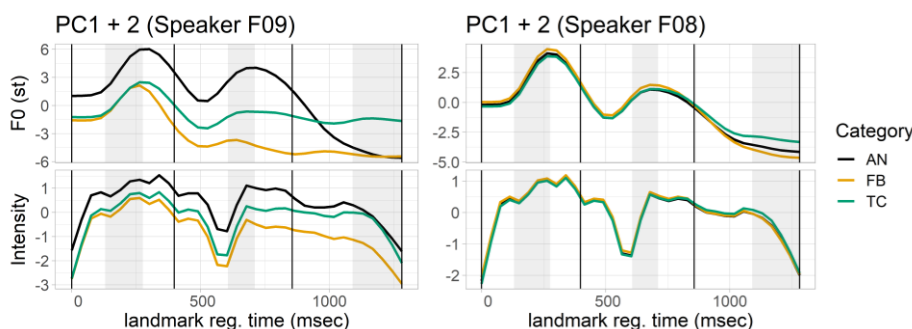


Figure 2. Predicted F0 and intensity curves for speakers F09 (left) and F08 (right). Vertical lines indicate word boundaries, shaded areas indicate stressed syllables.

Variability in Erzya conjugation

Fejes, László

HUN-REN Hungarian Research Centre for Linguistics

fejes.laszlo@nytud.hun-ren.hu

ORCID 0000-0002-2756-5189

The descriptions of the Erzya verbal stem types (Bartens 1999: 122, Keresztes 1990: 39, Keresztes 2011: 76, Mészáros 1998: 33, Pall 1996: 20) are usually quite vague and even contradict each other. Based on these descriptions and the example paradigms, the following generalisations can be made:

1. Every verb has two stems: a *vowel-final* and a *consonant-final*. The only difference between them is the presence or the absence of the vowel at the end of the stem.
2. The stem-final vowel can be /a/, /o/ or /e/, so the stem types can be labeled as *A stems*, *O stems* and *E stems*. However, the difference between O stems and E stems depends on harmony: /e/ occurs after palatalized consonants or /j/ and after front vowels in the previous syllable, /o/ occurs otherwise. Consequently, based on the aperture of the stem-final vowel, one can speak about *low stems* (identical with *A stems*) and *mid stems* (an umbrella term for *O stems* and *E stems*). Notice that these labels refer to paradigm types, not stems in actual forms.
3. There is another category of paradigm types which refers to the distribution of the vowel-final and consonant-final stems inside the paradigm. In general, consonant-final stems are used with vowel-initial suffixes (suffixes can begin with /a/ or /i/) and vowel-final stems occur before consonant-initial suffixes. However, the consonant-final stem is used before certain consonant-initial suffixes with many mid stems. These are labelled here as *unstable stems*. Mid stems, the consonant-final allomorph of which is used before vowel-initial suffixes, are labelled as *stable stems*.

Consequently, five stem types can be distinguished (cf. Table 1): one for low (A) stems (always stable); and four for mid stems, which can be differentiated based on two intersecting criteria, the quality of the stem-final vowel (/o/ or /e/) and stability. A comparison of the paradigm types shows that there are four paradigm slot types from the point of view of the differences between the five stem types. In the first slot type (e.g. INF), the vowel typical for the stem type is always present at the boundary of the stem and the suffix; therefore, it should be considered belonging to the stem. In the second slot type (e.g. IND.PRS.1SG), there is always /a/ at the stem boundary, which is reasonable to analyse as belonging to the suffix. There is always /i/ at the stem boundary in the third slot type (e.g. IND.PRS.3SG), which suggests a similar segmentation. Finally, the vowel typical for the stem type is present in the fourth slot (e.g. IND.PST.3SG), but not with all the stems with a mid vowel: it is present with the stable stems but missing with the unstable ones. Therefore, it is reasonable to analyse the form with a consonant-initial suffix.

form	low stem		mid stem		
	/a/ stem	/o/ stem	/e/ stem		
	(stable)	unstable	stable	unstable	stable
	'know'	'be born'	'sleep'	'do'	'climb'
1 INF	sodams	tʃatʃoms	udoms	tiejems	etsiems
2 IND.PRS.1SG	sodan	tʃatʃan	udan	tjejan	etsjan
3 IND.PRS.3SG	sodi	tʃatʃi	udi	tjeji	etsi
4 IND.PST.3SG	sodasi	tʃatʃsi	udosi	tjejsi	etsiesi

Table 1. Erzya paradigm types with key paradigm slots to distinguish them

Although grammars suggest that stems with a mid vowel are either stable or unstable (with a few exceptions), corpora (Arkhangelskiy 2019) show that vacillation between the forms with and without the vowel are quite widespread, although most verbs show a strong preference. Additionally, grammars do not give any clues about how to determine whether a stem is stable or unstable, implying that stability is a completely lexicalized property of the verb. Nonetheless, corpus research shows that the stability of the stem – or, at least, its clear preference – is quite well predictable from the phonemic structure of the stem, at least in most of the cases. However, while one would expect that the presence or absence of the mid vowel depends on syllable structure, that is, the vowel-final stem is used when otherwise the syllables would be ill-formed, it is not the case. While consonant-final stems ending in an /n^(j)/-final clusters are strongly dispreferred and consonant-final stems ending in a /t^(j)/ or /d^(j)/-final clusters are strongly preferred in paradigm slots type 4, consonant-final stems ending in a single consonant behave quite unpredictably: some of them clearly prefer the consonant-final or the vowel-final allomorphs in paradigm slots type four, while some of them do not show a definite preference.

The presentation discusses the further minor tendencies which influence whether a stem will be (dominantly) stable or unstable. In cases when it is possible, that is, when the very same stem shows a greater variability in stability, it is attempted to determine whether the variability is an interspeaker (interdialectal) one or an intraspeaker variation.

References

- Arkhangelskiy, Timofey. 2019. Corpora of social media in minority uralic languages. In *Proceedings of the fifth Workshop on Computational Linguistics for Uralic Languages, Tartu, Estonia, January 7 - January 8, 2019*, 125–140. Association for Computational Linguistics.
- Bartens, Raija. 1999. *Mordvalaiskielten rakenne ja kehitys*. Suomalais-Ugrilainen Seura, Helsinki.
- Keresztes László. 1990. *Chrestomathia morduinica*. Tankönyvkiadó, Budapest.
- Keresztes László. 2011. *Bevezetés a mordvin nyelvészetbe*. Debreceni Egyetemi Kiadó, Debrecen
- Mészáros Edit. 1998. *Erza-mordvin nyelvkönyv kezdőknek és haladóknak*. JATEPress, Szeged.
- Pall, Valdek. 1996. *Ersa keel. Õpiku konspekt ja sõnaloend*. [Valdek Pall], Tallinn.

INTRASPEAKER AND INTERSPEAKER VARIABILITY OF TEMPORAL VARIABLES IN YOUNG AND ELDERLY SPEAKERS' SPEECH ACROSS FOUR SPEAKING TASKS

Judit Bóna¹

¹ELTE Eötvös Loránd University, Budapest, Hungary

bona.judit@btk.elte.hu

Introduction

A large number of studies have investigated the temporal variables in the speech of speakers of different ages and speech tasks. Studies agree that 1) speech becomes slower in old age (Hartman-Danhauer 1976; Ramig 1983; Duchin-Mysak 1987; Smith et al. 1987; Gocsál 2000; Bóna 2014), and 2) speech tasks that require higher cognitive load are produced slower and with more frequent pauses and disfluencies (e.g. Ramig 1983; Duchin-Mysak 1987; Jacewicz et al. 2010). However, the majority of studies do not analyse the inter- and intra-speaker variability and how this is reflected in different speech tasks. The aim of this presentation is to investigate the intraspeaker and interspeaker variability of suprasegmental temporal parameters in the speech of young and old speakers in four different speech tasks with different cognitive loads. It is hypothesised that there will be large individual differences in both age groups, with some elderly individuals producing similar temporal parameters to young people. There will also be differences between speakers in how the speech tasks affect their speech.

Methods

For this study, 20 young (20-30-year-old) and 20 elderly (66-90-year-old) speakers were selected from the BEA database (Gósy et al. 2012). All participants were monolingual, native Hungarian speakers, without known mental or hearing problems, and with similar levels of education in both groups (all participants had at least a secondary education). The following 4 speech tasks were examined from each speaker: 1. interview (about 5 minutes from each interviewee), 2. narrative recall (the duration of which was speaker-dependent), 3. three-part conversation (4.5-5 minutes per interviewee, consisting of several turns), and 4. oral reading (the duration of which was also speaker-dependent). In total, 10.1 hours of speech was analysed. The parameters analysed by Praat were the following: speaking rate, average articulation rates, variability of articulation rates per speech segment, frequency of pauses, duration, average duration and variability of pause durations, frequency of filled pauses, average duration and variability of filled pauses. Results were compared by age and speech tasks and individual differences were also examined. Statistical analyses (mixed model ANOVA) were carried out by SPSS 20.

Results

Results confirmed the hypotheses. There were large individual differences at both ages, and the effects of different speech tasks on the temporal variables were different for each participant. It means that e.g. there were participants whose oral reading was the fastest, while for others, conversation. Figure 1 shows the distribution of speaking rates, Figure 2 and 3 show how speaking rates changed in each individual speaker.

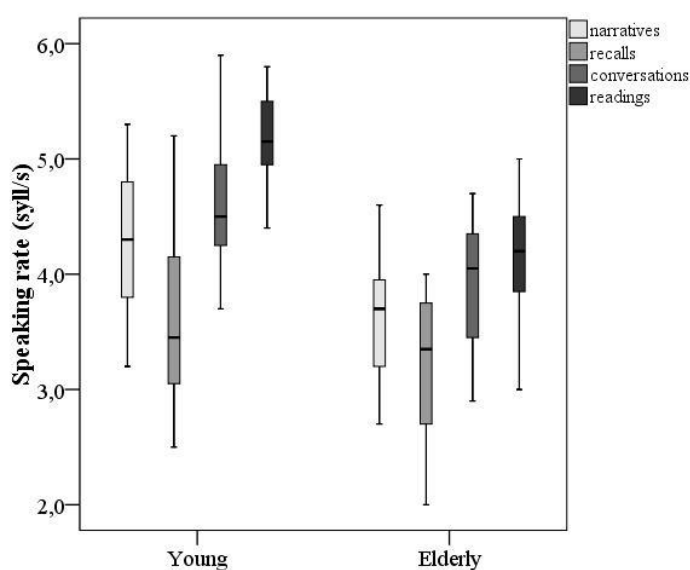


Figure 1: Speaking rates in the two age groups and four tasks

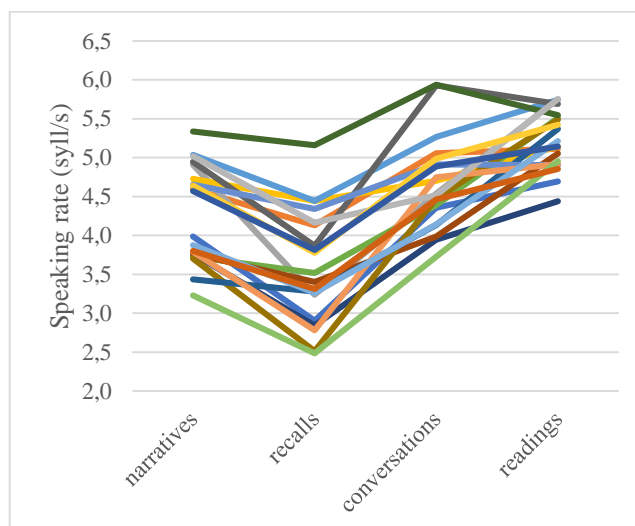


Figure 2: Individual differences in speaking rates (young speakers)

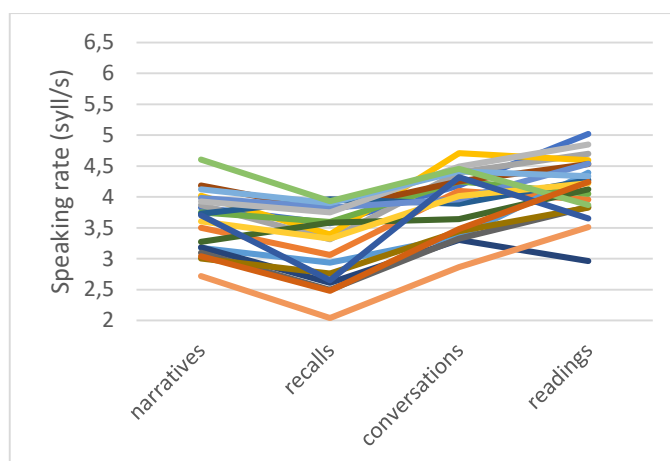


Figure 3: Individual differences in speaking rates (elderly speakers)

Conclusion

The results show that there are specific features of elderly speech compared to young speakers' speech that are present in all speech tasks (such as slower articulation rate), while other speech characteristics differ from young speakers only in certain speech tasks, and are not true for elderly speech in general. The results point out the importance of the speech task which is used to compare groups of speakers and to examine individual differences. The practical relevance of these results will be discussed at the end of the presentation.

References

- Duchin, Sandra W. – Mysak, Edward D. 1987. Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders* 20. 245–257.
- Gocsál Ákos 2000. A beszéd időviszonyai különböző életkorú személyeknél. *Beszéd kutatás 2000*. 39–50.
- Gósy Mária – Gyarmathy Dorottya – Horváth Viktória – Gráczki Tekla Etelka – Beke András – Neuberger Tilda – Nikléczy Péter 2012. BEA: Beszéltnyelvi adatbázis. In Gósy Mária (szerk.): *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó, Budapest. 9–24.
- Hartman, David E. – Danhauer, Jeffrey L. 1976. Perceptual features of speech for males in four perceived age decades. *The Journal of the Acoustical Society of America* 59. 713–715.
- Jacewicz, Ewa – Fox, Robert Allen – Wei, Lai 2010. Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America* 128. 839.
- Ramig, Lorraine A. 1983. Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders* 16. 217–226.
- Smith, Bruce L. – Wasowicz, Jan – Preston, Judy 1987. Temporal characteristics of the speech of normal elderly adults. *Journal of Speech and Hearing Research* 30. 522–529.

Age-related and between-speaker variability of the spectral features of voiceless sibilants in Hungarian

Tekla Etelka Grácz^{1,2} – Anna Kohári¹

¹HUN-REN Hungarian Research Centre for Linguistics

²Eötvös Loránd University

graczi.tekla.etelka@nytud.hun-ren.hu, kohari.anna@nytud.hun-ren.hu

Introduction

The present study aims to reveal the within- and between-speaker variability, and the possible age group differences in the realisations of Hungarian intervocalic /s/ and /ʃ/. The latest results of forensic phonetics also found sibilants worthy of study. Kavanagh (2012) reported that (the realisations of) /m/ and /s/ are highly speaker-specific. Furthermore, Fernandez (2022) observed high speaker-specificity of CoG, SD for /s, ʃ, z/. The spectral features of the sibilants vary with the exact place and length of the constriction that vary per speaker depending on the size and shape of the oral cavity (Kavanagh 2012), but also with speaker-specific/group features, like age, (e.g. Bóna – Beke 2013), gender (e.g. Romeo et al. 2013, Weinrich – Simpson 2015, Grácz – Kohári 2011.), age and gender (e.g. Kong – Kang 2021). It is important to understand within the study of the variability of sibilants that its differences not merely occur due to vocal tract size and shape differences but also serve as resources for the negotiation of identity (see for summary: Zimman 2017, Carralero-Fernandez 2022). Therefore, the between-speaker variability might be worth studying in further languages in search for forensic applicability.

Methods

40 native male speakers of Hungarian were selected from the BEA database (Neuberger et al. 2014). 10 subjects were chosen for each age groups (mean age and SD in years: 22.4±1.28, 32.9±1.32, 41.7±1.44, 53.9±2.06) Recordings of the reading tasks (altogether 38 sentences) were automatically segmented by WebMAUS (Schiel 1999, Kislér et al. 2017) based on the available transcriptions (Gyarmathy et al. 2014, Mády et al., submitted), and manually corrected for the selected /s/ and /ʃ/ realisations. The reading tasks include 19 intervocalic word-medial /ʃ/ but only 9 intervocalic word-medial /s/ segments, thus 9 further intervocalic but word-initial /s/ consonants were also selected for the analysis. The segment boundaries of the final labels were determined based on the second formants (F2) of the neighbouring vowels.

The duration and the first four spectral moments, and the peak frequency were extracted at 15 equidistant time points, and the duration of the entire consonant was measured. The measurements were conducted in Praat (Boersma – Weenink 2023) automatically based on Wikse-Barrow and her colleagues' work and script (2022), using the same settings. The statistical analyses were carried out in R (R Core Team 2022), separated for the two fricatives. Three male speakers had to be eliminated from the statistical analyses. One due to an injured file (20 ys group), and two due to having extremely outlying values (30, 40 ys group). Linear mixed models (LMMs, Bates et al. 2015, Kuznetsova et al. 2017, Bartoń 2022) were used to test the within-group differences of durations for age. The spectral moment trajectories of these occurrences were analysed by generalised additive mixed models (= GAMM, Wood 2017, van Rij et al. 2022). The possible age group differences and the possible effect of the vowel context were tested by separate GAMMs. The within-speaker vs. between-speaker variability was tested by comparing within-speaker variance to the between-speaker variance and Euclidean distances.

Results

No difference was found among the age groups in the duration (LMM). According to the generalised additive mixed models (Fig. 1), the highest estimated peak curve for /s/ is expected for the 30-year-old speakers, while the lowest for the 50-year-olds, with a significant difference between the 30s and 50s group. CoG of /s/ showed a similar pattern to the peak frequency. The spectral standard deviation and the skewness of /s/ did not, but the kurtosis did also show a difference between these two age groups. As for /ʃ/, the estimated slopes were somewhat different among the age groups but did not exhibit a significant difference in any of the measures. The vowel context had the following effects according to the GAMMs: The preceding vowel's height influenced the transition phase of /ʃ/, and the backness and height features of both the preceding and following vowels had an effect on the transition phases of /s/. The evaluation of the variance data is in progress.

Conclusions

The 30- and 50-year-old male speakers' tend to have significant differences in the peak frequency and CoG of the /s/ in roughly its "steady state" with higher values for the younger group (estimated difference curve: Fig. 2. left). Also, the spectral kurtosis was higher for the 50-year-old group at the start of the consonant, roughly for the transition phase (estimated difference curve: Fig. 2. right). We did not find an age-related difference for /ʃ/ for male speakers.

The significant age-related difference might be the consequence of the between-group variance as the highest values appear for the 30 year-old speakers, and the lowest values for the 50-year-old group. At the present point, we find two possible explanations: there is either a larger within-speaker variability in this age group, or a tendency of general age-related difference between 50+ years old speakers and the younger age groups. Both explanations have to be tested?) in future work.

The between- and within-speaker variability on the present corpus will be addressed in the final study.

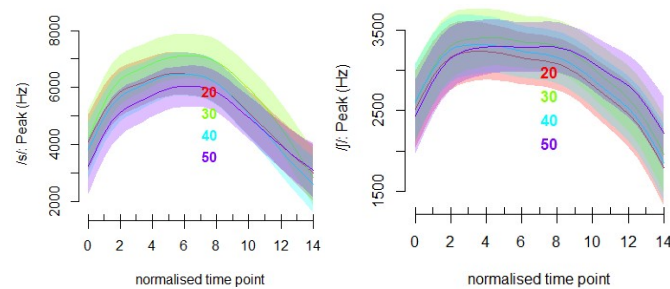


Figure 1. The estimated peak (Hz) trajectories (mean and 95% confidence intervals) of /s/ (left) and /ʃ/ (right) by the age group. (Male speakers).

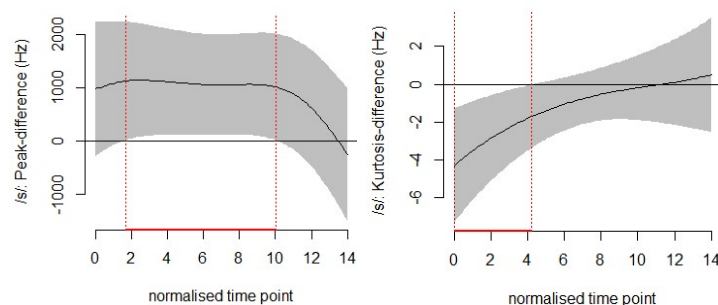


Figure 2. The estimated spectral peak (left) and kurtosis (right) difference between 30 and 50-year-old male speakers (mean and 95% confidence intervals).

Acknowledgement

The research was funded by the National Research, Development and Innovation Office (project number: FK128814). The first author was also supported the Bolyai János Research Scholarship of the Hungarian Academy of Sciences and by the ÚNKP 23-5.

References

- Bartoń, K. 2022. *MuMIn: Multi-Model Inference. R package version 1.47.1*, <<https://CRAN.R-project.org/package=MuMIn>>.
- Bates, D. – Maechler, M. – Bolker, B. – Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67/1. 1–48. doi:10.18637/jss.v067.i01.
- Boersma, P. – Weenink, D. 2022. Praat: doing phonetics by computer. <http://www.praat.org/>
- Bóna, J. – Beke, A. 2013. A zöngétlen réshangok akusztikai szerkezete fiatal és idős korban. In Gecső Tamás – Sárdi Csilla (szerk.): *Az interkulturális kommunikáció elmélete és gyakorlata*. Székesfehérvár–Budapest: Kodolányi János Főiskola–Tinta Könyvkiadó. 38–43.
- Fernandez, S. C. 2022. Speaker Specific Information in the Acoustic Characteristics of English Fricatives. *International Journal of Forensic Linguistic* 3. 105–115.
- Grácz, T. E. – Kohári, A. 2021. Spiránsok mássalhangzó-kapcsolatokban és intervokális helyzetben. *Beszéstudomány—Speech Science* 2/1. 24–46.
- Gyarmathy, D. – Neuberger, T. – Grácz, T. E. 2014. Lejegyzési útmutató a BEA Spontánbeszéd-adatbázis háromszintű annotálásához. *Alkalmazott Nyelvtudomány* 14/1. 35–44.
- Kavanagh, C. M. 2012. *New consonantal acoustic parameters for forensic speaker comparison*. PhD thesis. The University of York.
- Kisler, T. – Reichel U. D. – Schiel, F. 2017. Multilingual processing of speech via web services. *Computer Speech & Language*. 45, 326–347.
- Kuznetsova, A. – Brockhoff, P. B. – Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82/13. 1–26. doi:10.18637/jss.v082.i13 <<https://doi.org/10.18637/jss.v082.i13>>.
- Mády, K. – Grácz, T. E. – Kohári, A. – Mihajlik, P. (submitted): Revised annotation conventions in Hungarian speech corpora.
- R Core Team 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Romeo, R., Hazan, V., Pettinato, M. 2013. Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of the Acoustic Society of America* 134/5. 3781–3792.
- van Rij, J. – Wieling, M. – Baayen, R. – van Rij, H. 2022. *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. R package version 2.4.1.
- Schiel, F. 1999. Automatic Phonetic Transcription of Non-Prompted Speech. In: *Proc. of the ICPHS*. 607–610.
- Weinrich, M. – Simpson, A. P. 2015. Gender-specific differences in sibilant contrast realizations in English and German. In: *Proceedings of the 18th International Congress of Sciences*. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0261.pdf>
- Wood, S. N. 2017. *Generalized Additive Models: An Introduction with R (2nd edition)*. Chapman and Hall/CRC.

AGEING, IDENTITY, AND CROSS-LINGUISTIC VOICE COMPARISON

Sarah Melker

Institute for English Studies, Karl-Franzens-Universität, Graz, Austria

sarah.melker@uni-graz.at

1. Introduction

Ohala's frequency code (1994) postulates universals linking pitch (F_0) to biological factors and social meaning. In humans, and across languages, the frequency code predicts that higher pitch is a salient cue that signals smallness, politeness, youth, femininity, and submissiveness. In the past two decades much fruitful work has been done to unravel the complexities of the biological ageing process on the human voice (Fougeron, Guitard-Ivent, & Delvaux, 2021; Harrington, Palethorpe, & Watson, 2007; Reubold, Harrington, & Kleber, 2010), including in the field of forensic phonetics (Künzel, 2007; Rhodes, 2017). At the same time, these findings have raised further questions about the universality of age-related changes in speech production and its perception, as well as discrepancies between individuals in the ageing process imputable to health or social network reasons. Hejrná and Jespersen (2021) have recently drawn attention to this gap along with potential methods to remedy them by measuring both physiological and psychological factors. A previously unexplored angle which may give insight into speaker design in choosing an age-identity is that of proficient LX adult immigrants. Studies on large immigrant groups have shown that LX speakers actively choose degrees of belonging based on indexical features (Kozminska, 2021).

2. Methods

This study focusses on a single adult LX speaker of English, with Hungarian as a native language: the actress Zsa Zsa Gabor. Her English-language production over time is compared to her production of her native language (Hungarian) and additional languages (French, German), as well as to productions of her sister (also an adult LX speaker of English) and L1 speakers of English with similar social roles (actresses) and in similar speaking styles (TV interviews). The study follows previous studies in measuring fundamental frequency and vowel formants to track change over time, as well as rhythm metrics (Pellegrino, 2019) which have shown varying results as an indicator of ageing, and are useful in gauging LX proficiency and transfer (White & Mattys, 2007). The results presented here deal with pitch (measured as F_0 in Hz) and pitch span (maximum-minimum F_0) over time. Female speakers were chosen for the comparison material because as listeners they have been found to be more accurate in age judgements (Kelly & Harte, 2015). If this is due to heightened awareness of salient ageing cues, then it might be expected that a female LX could also be attuned to relevant features and may attempt to modulate these in producing her own speech. Celebrities were chosen due to the ease of accessing material as well as the factors that pressure associated with their public profile might have on portraying their age identity. Hungarian was chosen due to its dissimilarity from English in several prosodic parameters, e.g. Rácz and Papp (2016) report unexpected indexicalisation of pitch in Hungarian male speech, results which diverge from previous findings in other languages, which may inform findings in the present study.

The study asks the following questions: (1) how does the ageing process affects LX speech; (2) how might an LX's speech incorporate features that are salient to her; (3) what is the extent to which an LX speaker displays features identifying her with an age group if her exposure to different categories differs as compared to a typical L1.

3. Results

Biological ageing processes are mostly confirmed: median F_0 drops with age, and pitch excursion increases.

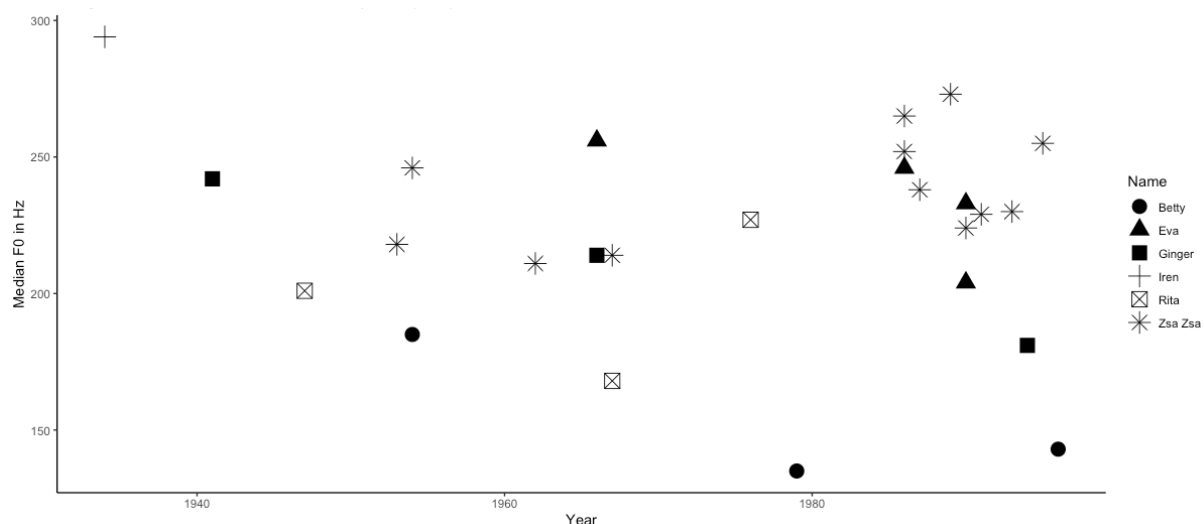


Figure 1. Longitudinal F_0 of Women Celebrities Corpus

However, there is one major outlier in the group: Zsa Zsa continues to maintain or increase her F_0 as she ages. This seems to occur independently from the language she uses, showing a single indexical category.

4. Conclusion

This study intended to examine the longitudinal F_0 and F_0 excursion of an LX speaker of English. These results add nuance to what has previously been found, and can be interpreted in the light of Pellegrino (2019) which took into account known health factors. It further supports recommendations in Hejná and Jespersen (2021) on considering social factors in ageing. It is planned to carry out perception testing among L1 English and L1 Hungarian listeners to determine whether potentially age-identity related features are salient to different groups and if so which ones. Further research may involve voice quality; investigation of pitch excursion as it correlates to health or social factors.

References

- Fougeron, C., Guitard-Ivent, F., & Delvaux, V. (2021). Multi-dimensional variation in adult speech as a function of age. *Languages*, 6(4), 176.
- Harrington, J., Palethorpe, S., & Watson, C. I. (2007). Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. *Interspeech*, 2753-2756.
- Hejná, M., & Jespersen, A. (2021). The coming of age: How do linguists tease apart chronological, biological and social age?. *Language and Linguistics Compass*, 15(1), e12404.
- Kozminska, K. (2021). Scaling diasporic soundings in the globalised world: A study of Polish stops in the UK. *Language & Communication*, 77, 17-34.
- Kelly, F., & Harte, N. (2015). Forensic comparison of ageing voices from automatic and auditory perspectives. *International Journal of Speech, Language & the Law*, 22(2), 167-202.
- Künzel, H. J. (2007). Non-contemporary speech samples: auditory detectability of an 11-year delay and its effects on automatic speaker identification. *International Journal of Speech, Language and the Law*, 14(1), 109-136.
- Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. *Sound symbolism*, 2, 325-347.
- Pellegrino, E. (2019). The effect of healthy aging on within-speaker rhythmic variability: A case study on Noam Chomsky. *Loquens*, 6(1), e060.

- RÁCZ, P., & PAPP, V. (2016). Percepts of Hungarian pitch-shifted male speech. In E. Levon & R. B. Mendes (Eds.), *Language, Sexuality, and Power: Studies in Intersectional Sociolinguistics* (pp. 151-167). Oxford University Press.
- Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F0 and the first formant: a longitudinal analysis in adult speakers. *Speech Communication, 52*(7-8), 638-651.
- Rhodes, R. (2017). Aging effects on voice features used in forensic speaker comparison. *International Journal of Speech, Language & the Law, 24*(2), 177-199.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501-522.

MULTILINGUAL VARIATION IN MULTILINGUAL SPEECH

Zoltán G. Kiss¹ – Zsuzsanna Bárkányi^{2,3}

¹Eötvös Loránd University, Budapest

²HUN-REN Hungarian Research Centre for Linguistics

³The Open University, UK

gkiss.zoltan@btk.elte.hu, bzs@nytud.hu

The handling of inter- and intraspeaker variability in speech has been a thorny issue in phonetic and laboratory phonological studies, both from a methodological and statistical point of view. These issues are multiplied in multilingual speech research as there are various “external” speaker-dependent variables that the learning of additional languages (L2, L3, Ln) introduce, which are often absent in native language (L1) research. They include not only the potential effect of age and gender but also that of the length of study, previous language experience, the context of study (e.g. immersion or heritage language), type of instruction (or lack of it), linguistic and metalinguistic awareness, etc. (e.g. Ortega 2014; Kopečková et al. 2023). All these speaker-related effects can have an influence on linguistic transfer, be it positive/facilitative or negative/inhibitory.

However, there is no agreement in the literature which theoretical account explains best the attested transfer or the lack of it. Most research on multilingual acquisition studies morphosyntactic phenomena. Hermas (2015) claims that L1 has a privileged role and thus serves as the exclusive source of transfer, while according to the L2 Status Factor model (Bardel and Sánchez 2017), the acquisition of L3 and subsequent languages is cognitively more similar to that of L2 and thus it is the L2 that is more likely to serve as source of transfer. According to the Typological Primacy Model (Rothman 2015), it is the more similar language that is transferred wholesale; while Berkes and Flynn (2012) claim that all previously acquired languages are available for transfer but only in case the influence is facilitative. On the other hand, Westergard et al. (2017) (Linguistic Proximity Model) and Slabakova (2017) (Scalpel Model) advocate for property-by-property rather than wholesale transfer, depending on which aspects of L1 or L2 are perceived to be more similar.

The quantification of “similarity” is, however, not straightforward. One variable that has been used to investigate the role of similarity in crosslinguistic influence is cognateness, namely whether the translation forms are phonologically and semantically similar in two languages or not. Studies in this area have shown that cognateness can make target-language production and perception faster as the lexical access is more accurate in the case of cognate words (e.g. Costa, Santesteban & Caño, 2005). However, cognates can also cause phonetic interference and thus might be produced differently from non-cognates (Amengual 2011). Data on the nature of the driving forces behind phonological transfer processes in multilingual speech are relatively scarce and similarly to other linguistic levels, no prevalent model has been identified. Neither has the link between perception and production been clearly defined.

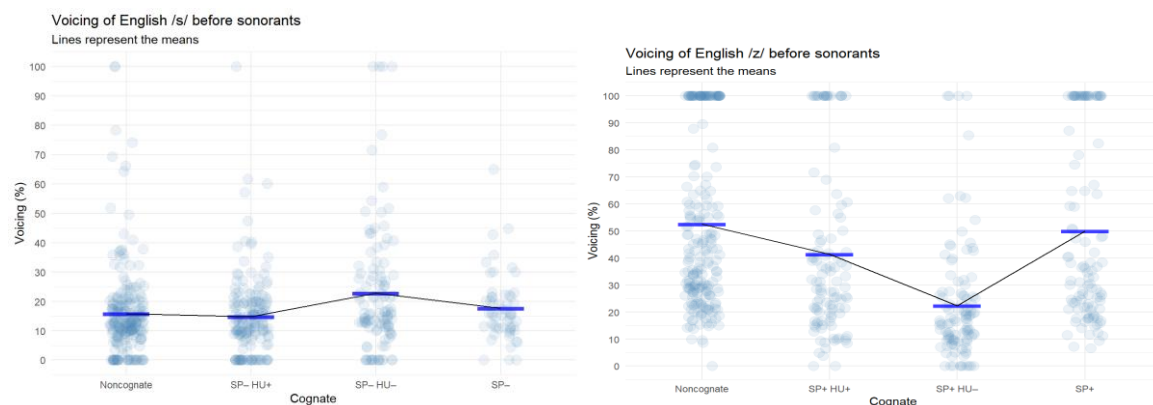
The aim of this research is therefore to investigate inter- and intraspeaker variability as well as the role of cognateness in the speech of L2/L3 learners of English and Spanish, whose L1 is Hungarian. Our focus will be on a particular dynamic phonological process, regressive voicing assimilation in which the target consonants are the alveolar sibilants /s/ and /z/ and the trigger consonant is either a plosive (“regular” regressive voicing assimilation, RVA) or a sonorant (pre-sonorant voicing, PSV). 14 young adult participants (L1 Hungarian) were recruited who started learning the other two languages later and in non-immersion context. As there is no clear

sequential or dominance difference between English and Spanish for these participants, we refer to both English and Spanish as L2 or L3.

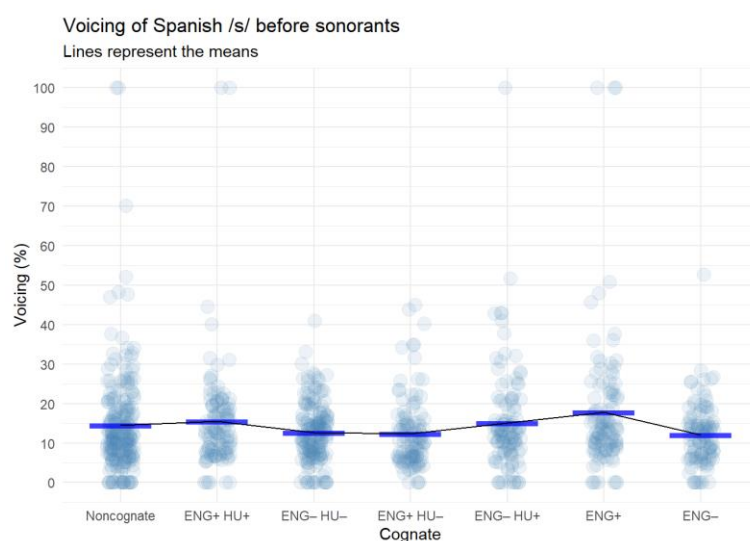
PSV in Spanish is very similar, though not identical to RVA in Hungarian. It is a RVA but also triggered by sonorant consonants, and due to the phonotactic restrictions of Spanish, it mostly applies to the sibilant fricative /s/ as in *prisma* ‘prism’, but there is no phonologically voiced sibilant in the inventory of Spanish. PSV is a relatively infrequent laryngeal pattern typologically, it is nonexistent in Hungarian, and it does not create a novel segment for L1 Hungarian learners as Hungarian has both /s/ and /z/ in its phoneme inventory. The question arises whether learners are able to notice the phonetic variability in the realisation of this segment in Spanish, learn PSV, and produce the alveolar fricative accordingly. In learners’ English interlanguage, on the other hand, we examine whether they are able to “unlearn” RVA, as English does not have this process (and it neither has PSV), although similarly to Hungarian, it has a /s/–/z/ contrast.

Our hypotheses to be tested are as follows: (1) The voice target of L3 (voicelessness or voicedness) is less likely to be realised if voicing in both L1 and L2 cognates contradicts it. (2) The L1 pattern dominates over the L2 pattern. To test these, we carried out a production and a perception experiment with the same 14 L1 Hungarian speakers of L2/L3 English and Spanish. The production experiment was a time-limited reading task and measured the amount of voicing within the fricative target (within words and between them), and how this was affected by the cognateness of the words, taking into account speaker and test word variability, too. The perception experiment examined the perceived salience of RVA and PSV, and the link between production and perception.

Our preliminary results concerning PSV support the above hypotheses for English. Pre-sonorant /s/ was produced with increased voicing on average compared to noncognate words if the test word was a cognate with negative facilitation in both Spanish and Hungarian (i.e., if the word is pronounced with /z/ in both languages; Fig. 1). Also, English words with pre-sonorant /z/ had much less voicing on average in cognates with inhibitory effect in Hungarian than in noncognates (i.e., when in Hungarian the word has a pre-sonorant /s/ rather than /z/, Fig. 2). However, the results did not show a cognateness effect in Spanish: there was a steady absence of PSV across the board, regardless of cognateness or facilitation, with little speaker variation (Fig. 3). The difference between the two languages in this respect also manifested in perception: speakers evaluated the presence of RVA in English speech as much less native-like; however, the lack of PSV in Spanish speech was regarded just as native as when it was present, suggesting that partial/optional PSV in Spanish is not salient enough to be “noticed” and learned as suggested by Flege and Bohn’s (2021) Revised Speech Learning Model. It seems that L1 speakers mapped both the phonetically voiced and phonetically voiceless sibilant to /s/ in L2/L3 Spanish, that is they merged the contrast in Spanish while they maintained it in English. We will present further results from the pre-plosive position in our proposed talk. Overall, we aim to fill the gap of research on phonological transfer in particular, and in general, to shed further light on laryngeal processes in Hungarian, the nature of contrast maintenance and neutralisation in assimilatory contexts both articulatorily and perceptually, as well as to highlight the presence (or lack) of speaker-related variation in multilingual speech.



Figures 1 & 2 *Voicing of English /s/ and /z/ before sonorants* (+ = positive transfer, i.e. facilitation, – = negative transfer, i.e. inhibition)



Figures 3 *Voicing of Spanish /s/ before sonorants*

References

- Amengual, M. (2011). 'Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism.' *Bilingualism: Language and Cognition* 15 (3), 2012, 517–530
- Bardel C. & Sánchez L. (2017). 'The L2 status factor hypothesis revisited: The role of metalinguistic knowledge, working memory, attention and noticing in third language learning'. In Angelovska T. & Hahn A. (eds) *L3 syntactic transfer: Models, new developments and implications*. Amsterdam: John Benjamins, pp. 85–102.
- Berkes É. & Flynn S. (2012). 'Multilingualism: New perspectives on syntactic development.' In: Ritchie WC and Bhatia TK (eds) *The handbook of bilingualism and multilingualism*. 2nd edition. Chichester: John Wiley and Sons, pp. 137–67.
- Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, 94, 94–103
- Flege, J., & Bohn, O. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3-83). Cambridge: Cambridge University Press.
- Hermas A. (2015). 'The categorization of the relative complementizer phrase in third-language English: A feature re-assembly account'. *International Journal of Bilingualism* 19: 587– 607.
- Kopečková R., M. Wrembel, U. Gut & A. Balas. (2023) Differences in phonological awareness of young L3 learners: an accent mimicry study. *International Journal of Multilingualism* 20:2, pages 408-424.

- Rothman J. (2015). 'Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered'. *Bilingualism: Language and Cognition* 18(2): 179–90.
- Slabakova R. (2017). 'The scalpel model of third language acquisition'. *International Journal of Bilingualism* 21: 651–66.
- Westergaard M, Mitrofanova N, Mykhaylyk R. (2017) 'Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model'. *International Journal of Bilingualism* 21: 666–82.

Lexical and morphosyntactic variance and its background factors in maternal narrative infant-directed speech

Veronika Harmati-Pap¹ – Bence Kas^{1,2} – Ildikó Tóth³ – Noémi Vadász⁴

¹ HUN-REN Hungarian Research Centre for Linguistics, Institute for General and Hungarian Linguistics

² Eötvös Loránd University, Bárczi Gusztáv Faculty of Special Needs Education

³ HUN-REN Research Centre for Natural Sciences, Institute of Cognitive Neuroscience and Psychology

⁴ HUN-REN Hungarian Research Centre for Institute for Lexicology

harmati.pap.veronika@nytud.hu, kas.bence@nytud.hu, toth.ildiko@ttk.hu, vadasz.noemi@nytud.hu

Widely accepted features of maternal infant-directed speech (IDS) are increased speech quantity, shorter length of utterances, and lexical simplification (Snow, 1972; Phillips, 1973; Genovese et al., 2019). The lexical and syntactic properties of maternal IDS can be interpreted as a kind of instinctive strategic redundancy which could support the infants' language comprehension and acquisition. Our previous research has shown that the strategic redundancy and simplification are detectable in maternal infant-directed narrative speech compared to adult-directed speech (ADS), but it does not vary as a function of age of children between 6 and 24 months.

One possible explanation for this is the high variance between and within individuals. Lexical and syntactic features of IDS differ from ADS only in a subset of mothers, and they show different patterns of reducing their language at different stages of their infant's life (an example is shown in Figure 1, where the dots represent the mean length of utterance (MLU) of individual mothers recorded at each age points). Thus, lack of significant differences in IDS between ages on average may be explained by the diversity of speech modification tendencies within the group of mothers.

The aim of this study is to characterize the variance in the speech of the mothers and to identify some of the factors that are associated with this variance. In doing so, we seek to answer two research questions:

- Can a language development adaptation effect be detected? Mothers may use more complex or simpler language to adapt to the child's contemporaneous language ability, which suggests a correlation between children's language development and the degree of maternal language simplification.
- Can a gender effect be detected? Mothers may be influenced by the gender of their children (Tronick & Cohn, 1989), and hence there would be different levels of patterns of IDS in narratives for boys and girls.

The study involved 185 mother-child (90 girls) dyads. To collect narrative IDS samples, mothers told stories to their first-born children at age 0, 6, 18 and 24, 30, and 36 months in a standardized situation, based on the same picture book. For control samples, an adult-directed narrative (ADS) was also recorded at each age. The variables tested were factors of speech quantity (number of words, number of utterances), factors of syntactic complexity (average utterance length), and factors of lexical variation (type-token ratio). The narratives were analyzed using the algorithms of CLAN software (speech quantity and syntactic factors) on the one hand and the latest modular pipeline version of the *e-magyar* digital linguistic parsing system (Váradi et al, 2017), *emtsv* (Indig et al.; 2019, Simon et al., 2020) (lexical factors) on the other hand. To determine the linguistic ability of the infants, we used the total expressive and receptive vocabulary scores from the Hungarian version of the MacArthur-Bates

Communicative Developmental Inventory (CDI) (Kas et al, 2010). To answer the questions above, we examined a sub-sample: we used the results from the 18-month age point, with the data of 110 mother-child dyads (54 girls) from both data collection situations - narrative speech sample and CDI questionnaire (Kas et al., 2010).

A considerable variation was detected in IDS between mothers in their infant-directed narrative speech at 18 months of age, as well as in the receptive and expressive vocabulary acquired by the children at the same age. However, no significant correlation was seen between maternal IDS characteristics and children's expressive vocabulary (Table 1).

Sex distribution of children showed differences in expressive vocabulary. Expressive vocabulary measured by KOFA at 18 months of age was significantly different for boys and girls in favor of the latter ($\chi^2=4.244$, $p=0.039$).

Grouping by the sex of the children also revealed differences in the IDS of mothers at 18 months of age. In terms of speech quantity and lexical variation, mothers of girls told significantly shorter stories (word count $F=5.901$, $p=0.019$, Tukey: MD(1-2): -65.7, $p=0.014$), but with a larger vocabulary (type-token ratio: $F=4.548$ $p=0.038$, Tukey: Tukey: MD(1-2): 0.0246, $p=0.037$) (3-4. Figure 4). Thus, a tendency appeared for the IDS narrative in containing significantly more repetitions for male children than for female children. We checked whether this difference according to gender is related to the slightly better vocabularies of girls. However, analysis of covariance showed that children's expressive vocabulary does not influence the above mentioned differences in either word count ($F=0.254$, $p=0.616$) or type-token ratio ($F=0.516$, $p=0.475$). This also supports the conclusion that modification of maternal IDS is not driven by the infants' expressive language performance.

In the present study, 110 mother-child dyads participated to investigate the association of variance in IDS with the sex of the child and language development in narrative situations. High variability in syntactic and lexical features between speakers was revealed in the maternal infant-directed register, especially at 18 months of age. Contrary to our preliminary expectations, the language performance of the infants did not show a correlation with maternal IDS modification tendencies even though our data showed that mothers of male infants narrate with an increased number of repetitions than mothers of female infants. Although there are significant differences in expressive vocabulary of the infants by sex at this age, covariates do not play a role in maternal speech modification strategies, i.e., no data are shown that male infants' language skills caused greater language adaptation in maternal IDS in narrative situations. In sum, maternal IDS is more pronounced in the case of male children, but this is not caused by differences in boys' language performance but by some other external factors unrelated to language performance but related to the sex of the infants.

References

- Genovese, G. Spinelli, M., Romero L., L., Aureli, T., Castelletti, G., & Fasolo, M. 2019. Infant-directed speech as a simplified but not simple register: A longitudinal study of lexical and syntactic features. *Journal of Child Language*. 22-44.
- Kas B., Lőrík J., Szabóné Vékony A. & Komárominé Kasziba H. 2010. A korai nyelvi fejlődés új vizsgálóeszköze, a MacArthur-Bates Kommunikatív Fejlődési Adattár (KOFA) bemutatása és validitási vizsgálata. *Gyógypedagógiai Szemle XXXVIII/2*, 114–125.
- Phillips, J. R. 1973. Syntax and vocabulary of mothers' speech to young children: age and sex comparisons. *Child Development*, 44(1), 182–185.
- Snow, C. E. 1972. Mothers' speech to children learning language. *Child Development*, 43(2), 549–565.

Tronick, E. Z., & Cohn, J. F. 1989. Infant-Mother Face-to-Face Interaction: Age and Gender Differences in Coordination and the Occurrence of Miscoordination. *Child Development*, 60(1), 85–92.

Appendix

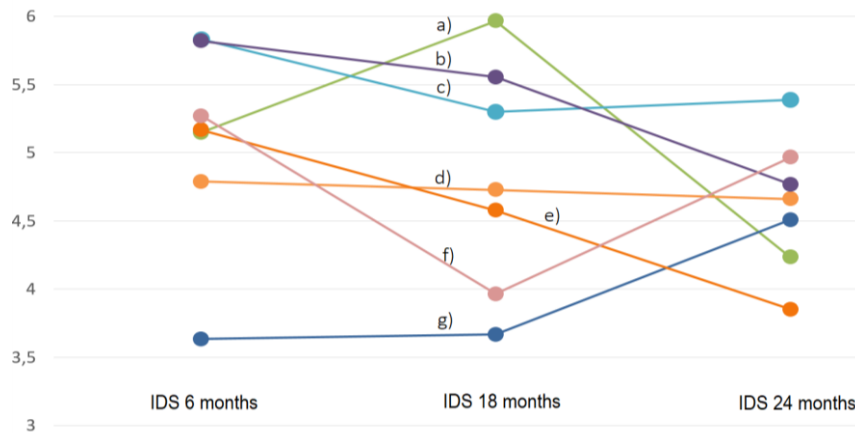


Figure 1: Examples of mean length of utterances in IDS, where each data set represents a single speaker.

Correlation Matrix

		MLU_18m_DIF	NRW_18m_DIF	NRU_18m_DIF	TTR_18m_DIF	VTTR_18m_DIF	PNTTR_18m_DIF
18m_exp-voc-total	Spearman's rho	0.102	-0.025	-0.074	0.061	-0.004	0.112
	p-value	0.292	0.795	0.446	0.531	0.964	0.246
receptive_18m	Spearman's rho	-0.028	-0.031	-0.016	0.005	-0.010	0.008
	p-value	0.829	0.809	0.902	0.972	0.937	0.951

Table 1: Correlative comparison between syntactic (MLU - mean utterance length), quantitative (NRW - number of words, NRU - number of utterances), and lexical (TTR - type-token ratio, VTTR - verb type-token ratio, PNTTR - noun type-token ratio) modifications of maternal speech and the child's receptive and expressive vocabulary measured by CDI at 18 months (Kas et al. 2010).

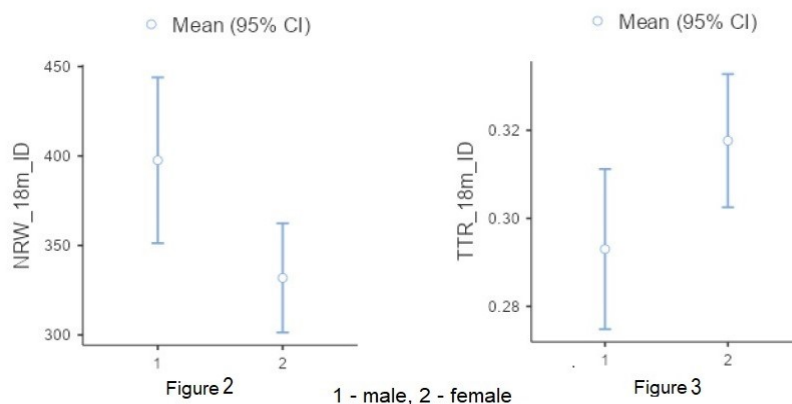


Figure 2-3: Gender differences in the number of words (NRW) and lexical variability (TTR) of maternal speech.

PHYSICAL ACTIVITY AS A SOURCE OF INTRASPEAKER VARIABILITY

Heather Weston

Humboldt-Universität zu Berlin, Germany
westonhe@hu-berlin.de

1. Introduction

People commonly speak during physical activity, but little is known about how such speech differs from speech at rest. Physiological adaptations to exercise affect parts of the body used for speech, particularly the respiratory system, which may affect the laryngeal mechanism and, in turn, the speech signal. Fundamental frequency (f_0) is known to increase during exercise (Van Puyvelde et al., 2018) but questions remain about the underlying mechanism and speaker variability (Godin and Hansen, 2008). The present study explores a possible link between exercise breathing and f_0 . Knowing more about how speech is affected by real-life situations can expand our knowledge on intraspeaker variability. It may also be relevant for automatic speech/speaker recognition, which works less well for non-typical voices – for example, due to respiratory pathologies, age, or panting while running for the bus. Speech under such conditions has received relatively little attention and more studies are needed (Hansen and Bořil, 2018).

2. Methods

Respiratory and speech recordings were made from 48 healthy female native speakers of standard German, who were recruited via a study database. The average age was 23.6 years ($SD = 3.8$). Mean body mass index was 22.0 ($SD = 3.5$), which falls within the healthy range.

A reading task elicited controlled speech samples across different levels of physical activity. To approximate a spoken register, a 126-word passage was generated from unscripted speech as follows: a female German speaker answered an open-ended question orally; her response was transcribed without disfluencies. Participants were told that the text was originally spoken and were instructed to read in whatever way felt natural. The text was read aloud once for familiarization prior to recording.

To approximate the physical effort (“workload”) required by daily activities (e.g., walking) and active occupations (e.g., fitness instruction), participants exercised at low and moderate workloads on a low-noise stationary bicycle. The workloads were defined using a relative measure of exercise intensity, percentage of maximal heart rate (HR_{max}), as 35% HR_{max} and 65% HR_{max} , respectively. The percentages were translated into target heart rates (in beats/minute, bpm) for each participant using a standard method in sports science, the Karvonen formula ($\text{target HR} = [(HR_{max} - HR_{rest}) \times \% \text{ intensity}] + HR_{rest}$).

A within-participant design was used with three workload conditions: (1) rest (sitting still), (2) cycling at low workload, (3) and cycling at moderate workload. Condition order was fixed (rest > low > moderate) because the cardiovascular effects of exercise may persist for several hours after exercise has ceased (Romero et al., 2017). Randomized order would thus make it difficult to ensure that the rest condition captured baseline speech.

Participants read the text three times while seated to obtain baseline speech and then cycled at a self-selected pace to transition to low workload. Once target heart rate stabilized, participants read the text again three times. Resistance on the bicycle was then increased to transition to moderate workload and participants read the text again three times. Heart rate was continuously monitored by the experimenters using a tablet readout. Speech was recorded at a sampling rate of 22,050 Hz using a head-mounted microphone. Breathing patterns were recorded using respiratory inductance plethysmography, which obtains breath depth and frequency by tracking chest wall kinematics.

The corner vowels /i a u/ in three words were delimited manually in PRAAT at the appearance of the second formant (Fischer-Jørgensen and Hutters, 1981). VoiceSauce (Shue

et al., 2009) was used to extract different glottal source parameters (f_0 range: 130–350 Hz) using speaker- and vowel-specific formant ceilings (Escudero et al., 2009). Only results for f_0 are reported here. Breathing depth and rate were obtained using a custom MATLAB–Praat pipeline.

Linear mixed-effects models were run in R and differences between workloads were compared using estimated marginal means (EMMs).

3. Results

At group level, f_0 was found to progressively increase with growing physical workload. Pairwise comparisons showed a significant increase in f_0 for both workloads ($ps < .001$). However, at individual level, speaker patterns varied considerably. Figure 1 contrasts the estimated marginal means obtained with the model with the different patterns shown by individual speakers, which include non-linear increases and decreases. This visual exploration of the data shows that speakers differed especially in their f_0 response to low workload.

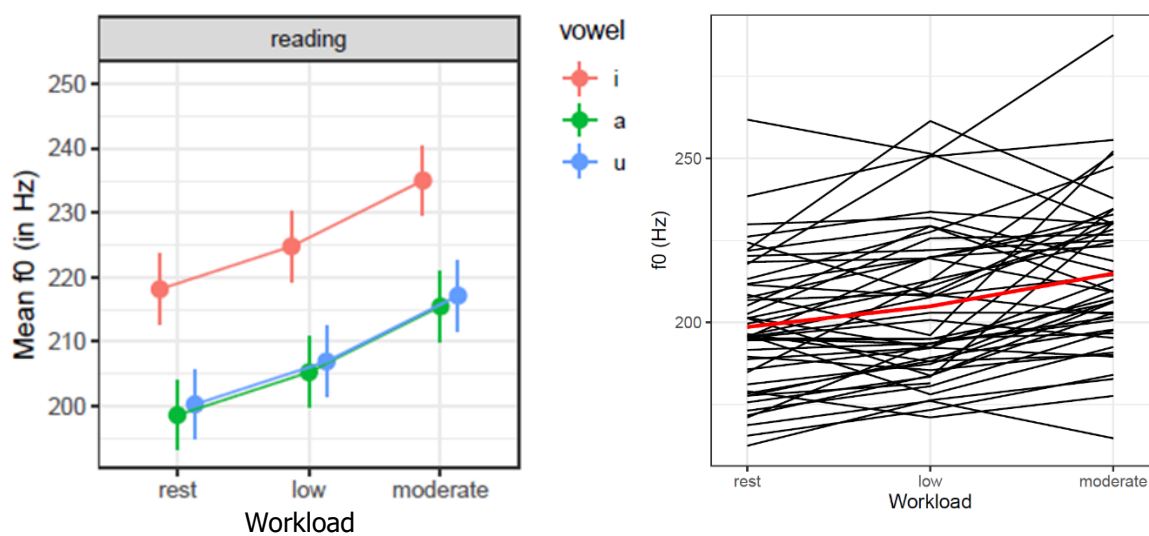


Figure 1. Estimated marginal means showing a linear increase in f_0 (left), while individual speakers (right) show diverse patterns with increasing physical workload; black lines depict mean for each speaker, red line shows group mean.

The analysis of the respiratory data showed a significant increase in breath depth for all workload comparisons ($ps < .001$), indicating that speakers took deeper breaths to speak during exercise than at rest. For breath rate, a significant increase was found from low to moderate and from rest to moderate ($ps < .001$), but not from rest to low workload ($p = .300$). A visual exploration of these data shows speaker differences in respiratory response (Figure 2): at low workload some speakers mainly increase breath depth (speaker jwi) while others tend to increase rate (speaker egb). These speakers also differ in the change in f_0 compared to speech at rest. However, it is not clear if the increase in f_0 is directly related to breath depth. Speaker cvq and zmq have similar breath depth/rate patterns, but different magnitudes of change in f_0 .

4. Conclusion

This study provided empirical evidence that both low and moderate physical workload can affect respiratory and laryngeal aspects of speech production. It extends previous group-level analyses by assessing individual responses to exercise, showing that participants differ considerably in their response to a low workload that is functionally equivalent to brisk walking. This finding may be relevant for automatic speech/speaker recognition applications because it demonstrates that everyday activity can give rise to intraspeaker variability in the acoustic signal. Further, the same level of activity can have different vocal effects for different speakers.

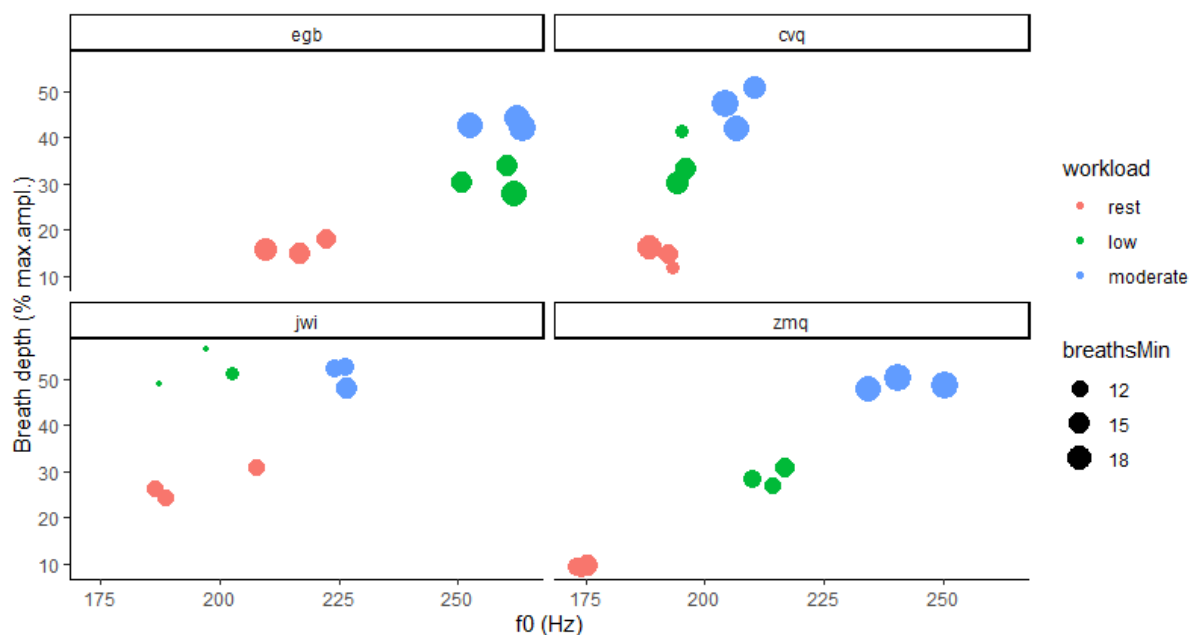


Figure 2. Relation between breath depth and f_0 with increasing physical workload for four speakers; dot size indicates breath rate (per minute)

References

- Escudero, P. – Boersma, P. – Rauber, A.S. – Bion, R.A.H. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America* 126. 1379–1393.
- Fischer-Jørgensen, E. – Hutter, B. 1981. Aspirated stop consonants before low vowels, a problem of delimitation – its causes and consequences. *Annual Report of the Institute of Phonetics University of Copenhagen* 15. 77–102.
- Godin, K.W. – Hansen, J.H.L. 2008. Analysis and perception of speech under physical task stress. *Proceedings of Interspeech 2008*. 1674–1677.
- Hansen, J.H.L. – Bořil, H. 2018. On the issues of intra-speaker variability and realism in speech, speaker, and language recognition tasks. *Speech Communication* 101. 94–108.
- Romero, S.A. – Minson, C.T. – Halliwill, J.R. 2017. The cardiovascular system after exercise. *Journal of Applied Physiology* 122. 925–932.
- Shue, Y.-L. – Keating, P. – Vicens, C. 2009. VOICESAUCE: A program for voice analysis. *Proceedings of the ICPHS XVII*. 1846–1849.
- Van Puyvelde, M. – Neyt, X. – McGlone, F. – Pattyn, N. 2018. Voice stress analysis: a new framework for voice and effort in human performance. *Frontiers in Psychology*. 9.

AGE PATTERNS AND INDIVIDUAL VARIABILITY IN TURN TAKING: TIMING AND STRATEGIES IN THE CONVERSATION OF 5- AND 9-YEAR-OLD CHILDREN

Hámori Ágnes¹ – Bóna Judit²

¹ HUN-REN Nyelvtudományi Kutatóközpont

²Eötvös Loránd Tudományegyetem

hamori.agnes@hun-ren.nytud.hu, bona.judit@btk.elte.hu

1. Introduction

In the last two decades, there has been a special interest in researching the timing of turn-taking in conversations (see e.g. de Ruiter et al. 2006, Magyari et al. 2014, Holler et al. 2015). In spontaneous conversations among adults, the transition between turns occurs rapidly and smoothly, with no long pauses or overlaps (Sacks et al. 1974). The average duration of speaker change is found to be between -250 and +250 ms (Stivers et al. 2009, de Ruiter et al. 2006, similar data in Hungarian cf. Markó–Gósy 2015, Horváth et al. 2021). The short gap between turns and the precise timing is remarkable because the time required to produce simple single-word utterances (minimum 600 ms, Indefrey and Levelt, 2004) far exceeds this average duration, suggesting a highly complex underlying mechanism involving earlier speech production planning and predictive processes (Levinson 2013, Holler et al. 2015). Taking this into account, the development of timing in children is particularly interesting; although relatively few studies have examined children's turn-taking timing (e.g., Casillas 2014), they point to the central importance of this phenomenon (Hilbrink et al. 2015). According to the results, children's turn-taking durations are strongly related to their age (also influenced locally by other factors, similar to the turn-takings of adults); however, this does not imply a linear decrease, but rather more complex changes (e.g., slowing down at the age of 1 year, Hilbrink et al. 2014). Little is known about turn-timing after the age of 5 years. Therefore, it is important to explore age-related timing patterns in children; however, significant individual variability can also be observed in this regard, similar to other areas of language development, which is also relevant for research. These may partly stem from differences in language development and partly from individual speech habits or strategies. Our research focuses on exploring these individual characteristics.

2. The aim and questions of the research

The present study examines the timing of turn-taking in the conversations of 5- and 9-year-old children (n=24, 12 male and 12 female participants), as well as two other relevant phenomena: the average turn length and the use of turn-taking facilitating techniques, which are closely related to turn-taking as well. The main question is what general characteristics can be observed, and what individual variability appears. The specific questions are as follows: 1) Is there a typical age pattern in the timing of turns in 5- and 9-year-olds compared to the turn-taking pattern of adults? 2) Are there differences between the 5- and 9-year-old groups concerning timing, average turn length, and turn-taking facilitating strategies? 3) Are there large individual differences in timing patterns, turn length, and the turn-taking facilitating strategies in children?

3. Data and method

In this study, we examine the conversations of 24 children with adults from the GABI database (Bóna 2017). The sample consisted of 12 five-year-olds and 12 nine-year-olds (6-6 boys and girls in each age group). All they were Hungarian monolinguals, with intact hearing and typical

speech development. Appr. 5 minutes of recording were analyzed per conversation, the total duration of the analyzed material was appr. 120 minutes, encompassing a total of 1238 turns, of which 628 turns were produced by the children. The audio recordings were annotated using Praat software (Boersma-Weenink 2008). We manually determined turn boundaries, filled pauses and turn-initial discourse markers. The timing of turn taking was measured based on the FTO value (Floor Transfer Offset, de Ruiter et al. 2006), examining the duration of pause between the end of the preceding turn and the start of the new turn (negative FTO indicates overlapping speech). Turn length was calculated for children's speech (per 100 words). The turn-taking support techniques include filled pause and discourse markers at the beginning of the turn; they were manually annotated. The quantified data were subjected to statistical testing using SPSS 20 software at a 95% confidence level (Kruskal-Wallis test).

4. Results

Regarding the timing (FTO values), we found a characteristic pattern in both age groups (Fig. 1), in which most of the turn-takings occurred with durations under 800 ms, with a mode of around 500 ms, and sometimes with 0 FTO values (in girls with a bit faster timing: mode around 250ms). Negative FTO (i.e. overlapping speech) rarely occurred in the children. However, there were large individual differences: one 5-year-old child almost always changed turns with a pause, and by two others we found many overlapping turns. (Fig. 2).

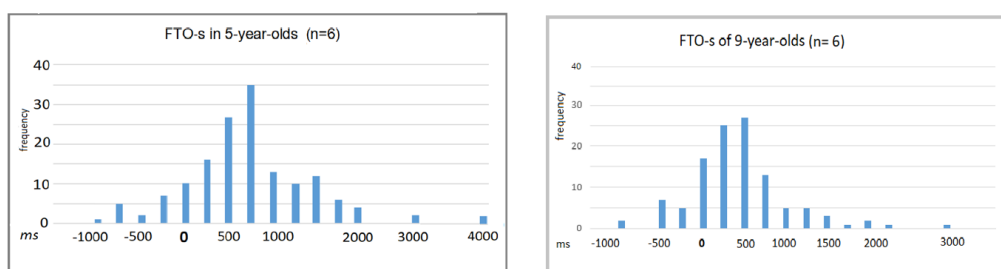


Figure 1. Duration and frequency of FTO-s at age of 5 and 9

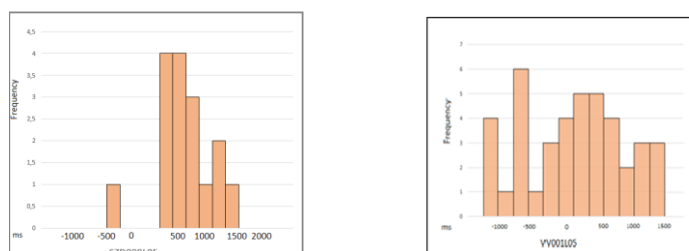


Figure 2. Individual differences in overlaps of turn-takings

The average turn length showed similar values for most children in both age groups. However, significant individual differences were found in one person each from both age groups (Fig. 3). In both age groups, turn-taking facilitating strategies were observed in every child, primarily through the use of filled pauses or a discourse marker („hát”), but the proportions and types varied depending on age and the individual (Fig. 4). At the age of 5, filled pauses dominated in all children, while at the age of 9, the DJ „hát” was the most common tool for turn-taking. The use of this turn-taking techniques differed significantly among individuals: some children started 10% of their turns this way, while another child initiated 70% of their turns using it.

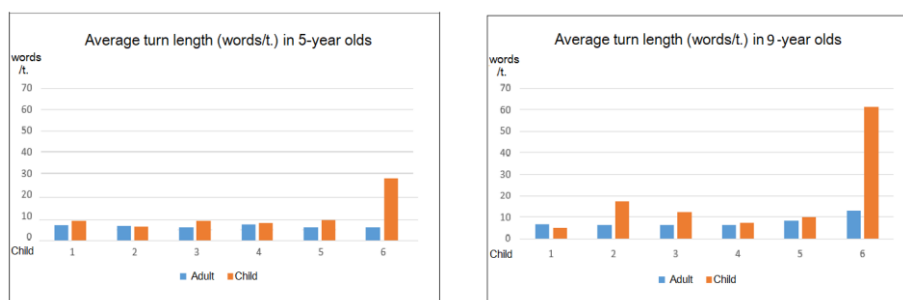


Figure 3. Average turn length at age 5 and 9

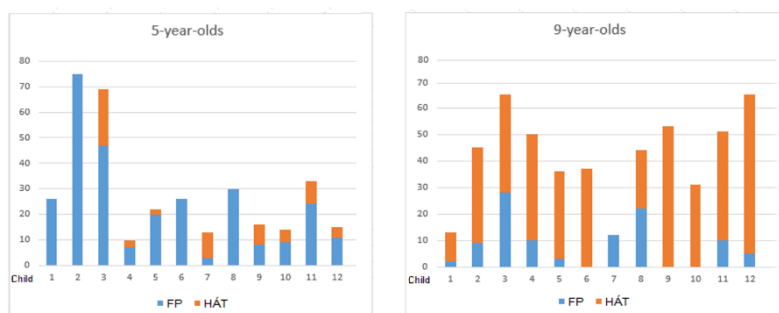


Figure 4. The proportion and type of turn-taking facilitating techniques (FP and HÁT) in turns of each child

5. Conclusion

The results indicate that there are typical age patterns in timing of turn-taking at the ages of 5 and 9, but within these, large individual differences may occur; both in terms of timing, especially the median FTO and overlapping speech, and the turn length and the strategies and means supporting rapid turn-taking. There are various possible reasons for these differences: the diverse rate of language development (most probably linked to the median FTO and average turn length), special strategies or individual speech habits, or a combination of these; but the affect of socialization and heard adult language patterns is also plausible.

Supported by the projects NKFI K-128810 and NKFH ELTE TKP2020-IKA-06, Hungary.

References

- Boersma, P. – Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43, <http://www.praat.org/>
- Bóna J. 2017. GABI – Gyermeknyelvi beszédatadátbázis a kutatásban. In Bóna J. (szerk.): *Új utak a gyermeknyelvi kutatásokban*. Budapest: ELTE Eötvös Kiadó. 35–50.
- Casillas, M. 2014. Taking the floor on time: Delay and deferral in children's turn taking. In Arnon, I. – Casillas, M. – Kurumada, C. – Estigarribia, B. (eds.): *Language in Interaction: Studies in honor of Eve V. Clark*. Amsterdam: John Benjamins. 101–114.
- De Ruiter, J. P. – Mitterer, H. – Enfield, N. J. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82/3: 515–535.
- Dér Cs. I. 2012. Beszélőváltások során használt diskurzusjelölők a magyar spontán beszédben. *Beszédkutatás* 20: 130–141.
- Hilbrink, E. E. – Gattis, M. – Levinson, S. C. 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in Psychology* 6. doi.org/10.3389/fpsyg.2015.01492
- Holler, J. – Kendrick, K. H. – Casillas, M. – Levinson, S. C. 2015. Editorial: Turn-Taking in Human Communicative Interaction. *Frontiers in Psychology* 2015/ 6: 1919.
- Horváth V. – Bóna J. – Dér Cs. I. – Gyarmathy D. – Hámori Á. – Huszár A. – Krepsz V. – Weidl Zs. 2021. Fonetika és társalgáskutatás: a beszélőváltások kérdésköre. In Markó A. (szerk.): *Tanulmányok a beszédtudományok alkalmazásainak köréből*. Budapest: ELTE Eötvös Kiadó. 251–278.

- Indefrey, P. – Levelt, W. J. M. 2004. The spatial and temporal signatures of word production components. *Cognition* 92: 101–144.
- Magyari, L. – Bastiaansen, M. C. M. – De Ruiter, J. P. – Levinson, S. C. 2014. Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience* 26/11: 2530–2539.
- Markó A. – Gósy M. 2015. A megszólalás stratégiái társalgásban. In Bárdosi V. (szerk.): A nyelvi pragmatika kérdései szinkrón és diakrón megközelítésben. Budapest: Tinta Könyvkiadó. 159–168.
- Sacks, H. – Schegloff, E. – Jefferson, G. 1974. A simplest systematic for the organization of turn-taking for conversation. *Language* 50: 696–735.
- Stivers, T. – Enfield, N. J. – Brown, P. – Englert, C. – Hayashi, M. – Heinemann, T. – Hoymann, G. – Rossano, F. – de Ruiter, J. P. – Yoon, K. – Levinson, S. C. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America* 106/26: 10587–10592.

**DOES A MORE COMPLEX PHONEME SYSTEM BIAS
PHONOLOGICAL DECISIONS IN VOWEL HARMONY? – PILOT
RESULTS FROM THE WESTERN HUNGARIAN VARIETY**

Katalin Mády

HUN-REN Hungarian Research Centre for Linguistics, Budapest

mady.katalin@nytud.hun-ren.hu

Introduction. Hungarian is an agglutinating language with front/back vowel harmony [4]. Most monomorphemic stems contain either front or back vowels, and most suffixes have a front and a back alternate (*kapu+ban* ‘in gate’, *fehér+ben* ‘in white’). Additionally, there is a substantial amount of loan words with mixed harmony including both back and front vowels. Similarly to many other languages with vowel harmony like Finnish, front unrounded vowels tend to be transparent, meaning that the harmonic suffixation rule looks through them, and the stem receives a back suffix, e.g. *papír+ban* ‘in paper’. However, the so-called height effects shows that the degree of transparency depends on vowel height [1]: the lower the vowel, the less transparent it is.

In Standard Hungarian, front unrounded vowels are high /i i:/, mid-high /e:/ and mid-low /ɛ/. Mixed stems containing back vowels and /i/ always take back suffixes, while /e:/ stems show some vacillation (*norvég+ban/ben*, ‘in Norwegian’). Mixed stems with /ɛ/ show a more complex behaviour. (1) Colloquial words such as *matek* ‘maths’ tend to take back suffixes, (2) educational terms and recent borrowings, e.g. *koncert* ‘concert’ occur only or mostly with front suffixes, while (3) the majority of loan words containing back vowels and /ɛ/, e.g. *hotel* ‘hotel’ are vacillating, i.e. the choice of front and back suffixes varies within and across speakers.

While Standard Hungarian (SH) has only three height categories for transparent (i.e. front unrounded) vowels, many dialects maintain two *e* phonemes. In these regional varieties, mid-open /ɛ/ and /æ/ are phonologically distinctive, although marginally. Standard Hungarian orthography does not distinguish between the two height degrees, they are both given with the letter *e*. Dialect speakers realise mixed stems with mid-open /ɛ/, being the third height degree on a four-item scale. According to the height effect, it might be the case that the division into two /ɛ/ height categories triggers a less transparent representation of these stems in dialect speakers, which would result in a higher preference for back suffixes in mixed stems.

Methods. The present study compares the suffix preference in two varieties: Western Hungarian (WH) with a phoneme system containing 4 heights for front unrounded vowels, and Standard Hungarian (SH) with only 3 heights. Mid-open /ɛ/ in mixed stems is the lowest vowel in the SH vowel system, while there is an additional open /æ/ in the WH vowel system that does not appear in mixed stems. We wanted to explore whether the height effect described by Hayes & Londe (2006) results in a different suffix preference in the two groups. We hypothesised that WH speakers would choose back suffixes more often than SH speakers because of the presence of an additional low vowel in their vowel inventory.

For the present study, 16 disyllabic stems containing a back vowel and /ɛ/ were chosen. The choice of the stimuli was based on an earlier production experiment investigating suffix preferences of SH speakers (for details, see [3]). 8 stimuli in the present experiment were vacillating stems (front/back preference close to 50% in the former production experiment), 4 went together with back suffixes (front selection < 20%), and 4 with front suffixes (front responses > 80%).

In Hungary, dialects do not have a high prestige, which leads to an overall loss of dialect usage among younger speakers. Thus, target speakers of the Western Hungarian dialect were mostly elderly people with limited knowledge of digital devices. For this reason, a paper-based questionnaire seemed to be the appropriate tool for data collection. The experimenter (the author) was present during data collection and noted whether participants were dialect speakers. They were asked to specify their gender, year of birth and the location of their first school year.

20 participants speaking the Western Hungarian dialect were recruited in Sopron, Western Hungary. Their mean age was 64.2 years (6 females, 14 males). A matching group with speakers in and around Budapest was included (mean age 56.9 years, 5 females, 11 males). All speakers in the WH group distinguished between mid-open and low *e*, and none of them did in the SH group.

Participants were asked to fill in a printed questionnaire presenting the target words embedded in a carrier sentence, in which the vowel of the target suffix was missing. The carrier sentences were identical to the ones used in the previous experiment and included four pairs of suffixes: dative *nak/nek*, instrumental *val/vel*, elative *ból/ból*, allative *hoz/hez*. Descriptive statistics along with linear regressions for the pooled data are reported below.

Results. The preference of front vs. back suffixes in the Western and Standard Hungarian group is shown in Fig. 1 for stems with an overall back, front and vacillating preference in Patay et al. (2018). Against our expectations, vacillating stems (e.g. *panel* ‘prefabricated house’) did not reveal a difference between the two dialect groups. Moreover, Western Hungarian speakers chose the back suffix less frequently with back-preferring stems like *maszek* ‘private’ than Standard speakers. At the same time, front-selecting stems like *roller* ‘scooter’ received a back suffix more frequently in the Western group.

Given the relatively small size of speaker groups, two additional comparisons were carried out. The left panel in Figure 2 shows a regression line based on the amount of front suffix choices in the experiment by [3] and the the two speaker groups in the present experiment. SH speakers’ suffix choices were close to that of the participants in Patay et al.’s experiment, whereas WH speakers preferred back suffixes slightly more, indicated by a less steeper regression line. The right panel in Figure 2 takes the written Hungarian National Text Corpus (MNSZ2) [2] as a reference. Again, speakers of WH had a slightly flatter regression line showing that they preferred back suffixes more than SH speakers. Interestingly, regression lines for both speaker groups are flatter for the text corpus than for the production experiment. It shows that a forced choice task might lead to more back suffix choices in general.

Conclusions. Preliminary results show that the height effect seems to differentiate between speakers with a 4- and a 3-level vowel system only to a small extent. A possible reason for the missing height effect is the complex relationship between factors affecting harmonic suffix choice such as the surrounding sentence context or the phonotactics of the suffix. Another explanation is that WH speakers are usually not aware of the fact that they distinguish between two *e* phonemes, since they are not distinct in orthography and have a low functional load in phonemic distinction in general.

Acknowledgements. This work was funded by grant NKFIH K 119863, *Experimental and theoretical investigations of vowel harmony patterns*. I am indebted to Péter Rebrus for his feedback on the choice of the stimulus material for this experiment.

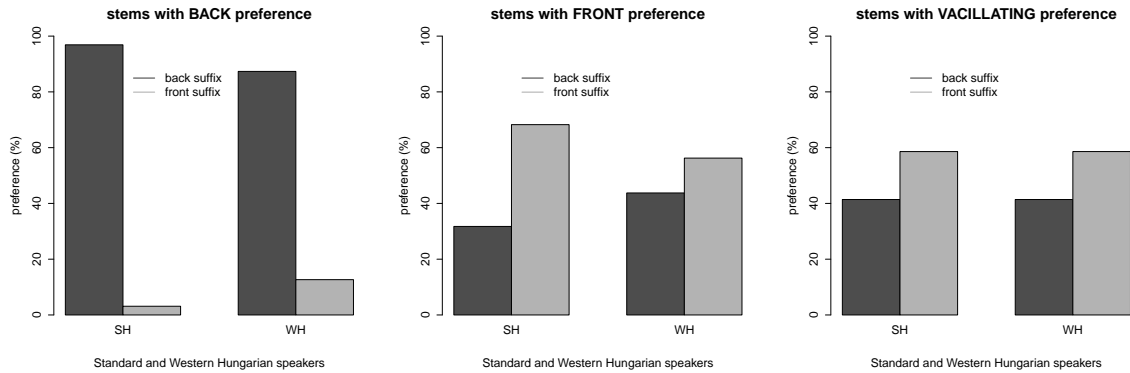


Figure 1: Barplots for stems with a general preference for (a) back, (b) front suffixes and (c) vacillating stems in Standard (1E) and Western (2E) Hungarian speakers.

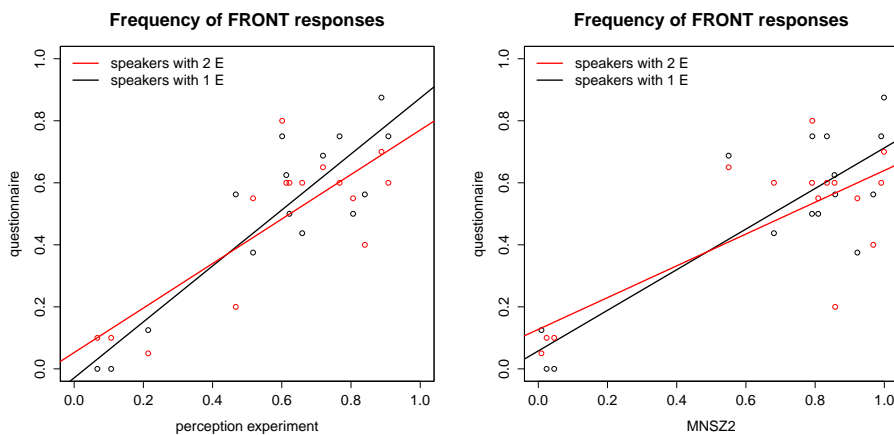


Figure 2: Linear regressions for the former production experiment (left), and for the text-based database MNSZ2 (right) and WH (red) and SH (black) speakers in the present study. Data points show the ratio of front suffixes for the 16 stems used in the questionnaire.

References

- [1] Bruce Hayes and Zsuzsa Cziráky Londe. Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology*, 23, 2006.
- [2] Csaba Oravecz, Tamás Váradi, and Bálint Sass. The Hungarian Gigaword Corpus. In *Proc. LREC 2014*, Reykjavik, 2014.
- [3] Fanni Patay, Ágnes Benkő, Ágnes Lukács, Péter Rebrus, and Miklós Törkenczy. Neutrality patterns in Hungarian backness harmony: an experimental study. In *15th Old World Conference on Phonology*, London, 2018.
- [4] Péter Siptár and Miklós Törkenczy. *The Phonology of Hungarian*. University Press, Oxford, 2000.

FILLED PAUSES AS INDIVIDUAL SPEECH CHARACTERISTICS?

Gráci, Tekla Etelka^{1,2} – Huszár, Anna¹ – Markó, Alexandra^{1,3}

¹HUN-REN Hungarian Research Centre for Linguistics

²Eötvös Loránd University

³SSNS Institute for Expert Services

graczi.tekla.etelka@nytud.hun-ren.hu, huszar.anna@nytud.hun-ren.hu,

marko.alexandra.phd@gmail.com

Introduction

Realization of filled pauses are considered speaker-specific characteristics in forensic voice comparison, including their frequency and their acoustic parameters (Jessen 2021, de Jong-Lendle 2022), what is more, it is claimed that this phenomenon is one of the most promising clues (Rose 2002; Hughes–Wood–Foulkes 2016). In forensic context, filled pauses have several potential advantages over other (vocalic) segments: they occur frequently; their frequency is speaker-specific; their occurrence is relatively independent of the linguistic structure; they are typically realized with longer duration than vowels; due to their duration and phonetic position they have a less variable acoustic structure; they are produced unconsciously, therefore their acoustic form is not under the speaker’s control (Karpiński 2013, Zdravković–Jovičić 2020).

Language-specificity of filled pauses is suggested by the literature, however, it is not clear what this means. Filled pauses, therefore, have been analysed in several languages and various aspects, including frequency, phonetic position, duration, formant frequencies, f_0 , and voice quality, among others. For example, in Polish the first two formants were found to be significantly different among speakers, but jitter-related voice parameters showed a limited level of distinctiveness (Karpiński 2013).

The comparison of temporally distant speech samples, which is the topic of interest of the present study, is relatively understudied, although this is the typical setting in the context of forensic speaker comparisons. The extent of the temporal gap between recording sessions is different case by case, which might vary from a couple of months to several years (the longest documented interval is 27 years (French–Harrison–Windsor-Lewis 2006)). A study of young Dutch speakers with a 2.5-year interval confirmed that the spectral characteristics of filled pauses remained stable over time (de Boer–Quené–Heeren 2022), however, we are not aware of studies with longer time lag with respect to filled pauses. The research question of the present study, therefore, is the following: are speakers’ filled pauses also acoustically consistent across recordings made 10 years apart?

Methods

The present study investigates the filled pauses in Hungarian spontaneous speech comparing non-contemporary recordings. 24 speakers’ interviews were chosen from the Hungarian Longitudinal Database that includes various speech types recorded with the same speakers at two time points 10 years apart (first recordings: Neuberger et al 2014, Longitudinal Database: Gráci et al 2020). 10 female and 14 male speakers were selected aged between 19 and 34 years at the first recording (mean: 26.5 ys, sd: 4.7 ys). The interviews include quasi-monologous speech about the participants’ education/job, hobby, and family. The recordings are of various lengths, therefore the first three minutes were analysed from the start of the first utterance of the speaker. The silent and filled pauses were labelled in Praat (Boersma–Weenink 2023). The following pieces of information were included in the labels: (i) “segmental realisation” based on the spectrogram and perception (e.g. [ə], [əm], [əhm], [əh]), (ii) “attachment”: possible attachment to the preceding or following words or appearing

within a chain of silent-filled-silent (combined) pause, (iii) “phonation type”: the phonation type observed by perception and the oscillogram (e.g., modal, creaky, breathy). The first three formants of the [ə] hesitations were measured at 11 equidistant time points along the duration automatically with the basic settings of Praat. The pause types were considered two ways: 1) as silent, filled, and combined pauses, 2) combined pauses segmented into silent and filled parts. The following measures were calculated for the pause types: (i) average occurrence (occurrence number per minute), (ii) occurrence frequency ratio of each pause type, and (iii) the ratio of occurrences of the filled pause types (only schwa, only nasal, multisegmental) were compared by Spearman’s correlation between the two recording sessions. (iv) Euclidean distances between the two recordings of each speaker and across the speakers within the first and second recording phases were computed based on the first three formants at the midpoint of the vowel.

Preliminary results

When dividing pauses in two groups: *silent* and *filled* types, the number of silent pauses per minute did not, but the number of filled pauses per minute did show a medium correlation between the two recording sessions. However, when including a *combined pauses* category as well, all three types of pauses did show medium correlations between the two recording sessions: both the results for occurrence frequency (number/minute; fig. 1.) and their ratio across categories (number of a type/number of all pauses together).

The filled pauses (considering all separate filled pauses and the parts of the combined pauses) were divided into 6 main segmental types: [ə], [m], [əm], [əh], [ɒm], miscellaneous (= types that appeared less than 6 times among all the data). With calculating the occurrence frequency (number/minute), only the [ə] types showed medium correlation between the two recording sessions, while calculating their ratio to all the filled pauses, the [ə] and [m] and the miscellaneous types showed medium correlation between the two recording sessions. However, in the case of the miscellaneous type, the confidence interval is too large as the data have large variability between the two sessions.

The formant values of the [ə] filled pauses showed large within-speaker variation for some speakers within the same recording sessions compared to the between-speaker variation.

Discussion

Künzel found that this ratio remains “constant” (1997, cited by Rose 2002). The ratio of filled pauses to all pauses in our study also seems to strengthen Künzel’s results: we found medium/strong correlation between the two recording sessions in both the two- and three-group calculations. Our results indicate that including three pause types (silent, filled, combined), the frequency of occurrence of the types is relevant for speaker comparison analysis. The frequency of occurrence of the [ə] and [m] as filled pauses may also be subject to further research.

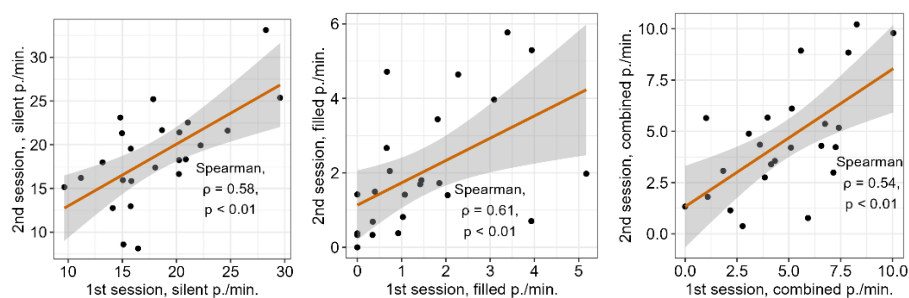


Figure 1. Correlation of the ratio of the pause types between the two recording sessions (silent, filled and combined pauses)

Acknowledgement

The research was funded by the National Research, Development and Innovation Office (project number: FK128814). The first author was also supported by the Bolyai János Research Scholarship of the Hungarian Academy of Sciences and by the ÚNKP 23-5.

References

- Boersma, P. – Weenink, D. 2022. *Praat: doing phonetics by computer*. <http://www.praat.org/>
- de Boer, M. – Quené, H. H. – Heeren, W. F. L. 2022. Long-term within-speaker consistency of filled pauses in native and non-native speech. *JASA Express Letters* 2, 035201. <https://doi.org/10.1121/10.0009598>
- de Jong-Lendle, G. 2022. Speaker Identification. In Guillén-Nieto, V. V. – Stein, D. (eds.): *Language as Evidence. Doing Forensic Linguistics*. Cham: Palgrave Macmillan. 257–319.
- French, J. P. F. – Harrison, P. – Windsor-Lewis, J. 2006. R v John Samuel Humble: The Yorkshire Ripper Hoaxer trial. *IJSL* 13. 256–273.
- Grácsi, T. E. – Huszár, A. – Krepsz, V. – Száraz, B. – Damásdi, N. – Markó, A. 2020. Longitudinális korpusz magyar felnőtt adatközlőkről. In Berend, G. – Gosztolya, G. – Vincze, V. (eds): XVI. *Magyar Számítógépes nyelvészeti konferencia. Proceedings*. University of Szeged. 103–114.
- Hughes, V. – Wood, S. – Foulkes, P. 2016. Strength of forensic voice comparison evidence from the acoustics of filled pauses. *IJSL* 23/1. 99–132.
- Jessen, M. 2021. Speaker profiling and forensic voice comparison. The auditory-acoustic approach. In Coulthard, M. – May, A. – Sousa-Silva, R. (eds.): *The Routledge Handbook of Forensic Linguistics*. Second edition. London – New York: Taylor & Francis. 382–399.
- Karpiński, M. 2013. Acoustic Features of Filled Pauses in Polish Task-Oriented Dialogues. *Archives of Acoustics* 38/1. 63–73.
- Künzel, H. J. 1997. Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics* 4 (1). 48–83.
- Neuberger, T. – Gyarmathy, D. – Grácsi, T. E. – Horváth, V. – Gósy, M. – Beke, A. 2014. Development of a large spontaneous speech database of agglutinative Hungarian language. In: Sojka, P.– Horák, A. – Kopeček, I. – Pala, K. (eds): *Text, Speech, and Dialogue: 17th International Conference, TSD 2014*, Brno, Czech Republic, September 8-12, 2014. Proceedings. Cham, Svájc : Springer-Verlag. 424–431.
- R Core Team 2022. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Rose, P. 2002. *Forensic Speaker Identification*. London – New York: Taylor & Francis.
- Zdravković, S. – Jovičić, S. 2020. The importance of speech pauses for psychotherapeutic and forensic observations. In Jaćimovski, S. et al. (eds). “*Archibald Reiss Days*”. Belgrade, 18-19 November 2020. Thematic Conference Proceedings of International Significance. Belgrade: University of Criminal Investigation and Police Studies. 513–524.

BESZÉLŐN BELÜLI ÉS BESZÉLŐK KÖZÖTTI VARIÁCIÓ A MÚLTBAN? – JELENSÉGCSOPORTOK ÉS MÓDSZEREK LEHETSÉGES SZEREPÉRŐL A 18. SZÁZADI MEGNYILATKOZÓK AZONOSÍTÁSÁBAN

Varga Mónika¹

¹ HUN-REN Nyelvtudományi Kutatóközpont Történeti Nyelvészeti és Uralisztikai Intézet

varga.monika@nytud.hu

Hogyan vizsgálható, és mi lehet a jelentősége a beszélők azonosításának a történetiségben? Mely tényezők alapján tehetünk állításokat az egyéni nyelvhasználatról, valamint a beszélők közötti egyezkedésről írott szövegek alapján? E kérdéskör vizsgálata a grammatika és a pragmatika legkülönbözőbb területeit fedheti le, ugyanakkor az egyes időszakok és források alapján más-más jelenségcsoportok szolgálhattak viszonylag nagyobb magyarázó értékkel. Jelen előadás problémafelvető jelleggel tekint át néhányat a kutatásba vonható jelenségek és a vizsgálati módok, lehetőségek közül.

A kódexirodalom és a korszak más forrásai kapcsán már kialakultak a kézázonosítás olyan módszerei, amelyekkel a kor valamilyen tekintetben markánsnak mutakozó megnyilatkozási azonosíthatóvá váltak (így többek között a Nyulak szigeti kódexmásolók, Ráskay Lea keze Elena priorissza levelében, a pszicholingvisztikai jellemzők alapján is közös kéz az Apor- és a Lányi-kódexben, részletesen lásd Haader 2009, 2017, 2019). A paleográfiai jellemzés mellett lényegessé vált a nyelvi elemzés olyan fogódzók mentén, mint a fonéma–graféma megfelelések, a rövidítés, a szövegtagolás, beleértve a központozást, a szövegbeli javítások, valamint a nyelvjárási sajátosságok (ezek egyike sem kiindulásnak, hanem eredménynek tekinthető, lásd még Haader 2014, 2015). Egy ideje a hasonló vizsgálatok kiterjedtek a későbbi időszakokra, így a 16–18. századra is, olyan forrásokra, mint a misszilizsek (például Kocsis 2020, 2022) vagy éppen a boszorkányperek.

Utóbbi forráscsoportban a megnyilatkozók azonosítása több aspektusból is felmerül: a lejegyzői tevékenység körvonalazása és a lehetséges motivációk feltárása (javítások, módosítások alapján, Varga 2020, vagy éppen szövegváltozatok összevetésén keresztül, Varga 2022a) mellett jogi értelemben is jelentősége lehetett annak, hogy a tanúk a megnyilatkozásai alapján azonosították-e a vádlottat valamilyen károkozással összefüggésben. Magához a boszorkány szerepéhez is kötődnek olyan hiedelmek, mely szerint a viselkedéséből, egyes nyelvhasználati mintázataiból lehet a rontással gyanúsíthatók személyére következtetni. Az előadás egyik esettanulmánya néhány jegyzőkönyv metakommunikatív jelzéseire, reflexióira (jelentőségükről a nyelvi formák és a társadalmi jellemzők kapcsolatát tekintve a történeti magyarázatokban lásd Beal 2019) irányul a boszorkányperes eljárás kontextusában, beleértve a vádlottak kihallgatását is, ahol egyfajta egyezkedés tárgyává válhatott az, hogy egyes tartalmak kitől és hogyan hangzottak el.

A megnyilatkozók azonosíthatóságának problémája a nyelvtörténeti kutatások szempontjából a tekintetben is fontossá válhat, hogy inkább az egyénhez vagy társadalmi rétegekhez, illetőleg regiszterhez (bizonyos használati szituációkhoz asszociált nyelvhasználati variánsok csoportjaként értelmezve azt, vö. Biber–Conrad 2009) kapcsolható-e egyes variációk egy adott időszakban. Történeti források esetében a regiszterek nem kötött kategóriának számítanak, hanem a különféle adatok és preferenciák mentén finomhangolást igényelhetnek, így meghatározásuk is eredménynek tekinthető. Ugyancsak felmerül, hogy az egyén vagy a regiszter befolyásolja-e inkább a variánsok disztribúcióját, sőt akár az újabb variánsok diffúzióját (Gugán 2023).

E kérdéskörhöz kapcsolódva az előadás evidencialitás- és attitűdjelölő elemek megoszlásával foglalkozik (lásd még Varga 2022b) elsősorban 18. századi emlékiratok összehasonlításán keresztül felvetve azt, hogy lehetett-e bizonyos pragmatikai markerek vagy marker-csoportok használatának viszonylagos egyénítő szerepe a korszakban. A megállapítások értékelése során ugyanakkor tekintetbe kell venni az olyan módszertani problémákat is, mint a források nem kiegyensúlyozott fennmaradása és a reprezentativitás (lásd még Dömötör–Gugán–Varga 2021). Eseti jelleggel két olyan szöveg kerül előtérbe, amely különböző szerzőktől származik, de ugyanarról az utazásról számol be. A későbbi szakirodalom szerint a tartalmi különbségeket tekintve asszonyi szemmel készült leírás állítható szembe a férj diplomáciai tárgyú naplójával (S. Sárdi 2000, 2014). A kvalitatív elemzés részeredményeinek tágabb kontextusba helyezéséhez azonban további forrásokat is szükséges bevonni a vizsgálatba (egyrészt néhányat a korszakban alulreprezentált női naplóiírók műveiből, másrészt egy szerzőtől származó többféle írást, így például emlékiratokat és magánleveleket).

Irodalom

- Beal, J. C. 2019. Enregisterment and historical sociolinguistics. In: Jansen, S. – Siebers, L. (eds.) *Processes of Change. Studies in Late Modern and Present-Day English*. Amsterdam/PA: John Benjamins. 7–23.
- Biber, D. – Conrad, S. 2009. *Register, Genre, and Style*. Cambridge: University Press.

- Dömötör A. – Gugán K. – Varga M. 2021. Váltakozás, változás és nyelvtörténet: a variacionista megközelítésmódról – és a jelen kötetéről. In: Dömötör A. (szerk.) Gugán K. – Varga M. (társszerk.): *Versengő szerkezetek a középmagyar kor nyelvében*. (Nyelvtudományi Értekezések 169.) Budapest: Akadémiai Kiadó. 5–28.
- Gugán K. 2023. A kivétel erősíti a szabályt? A regiszter és az egyén szerepe a stabil variációban. (*Kézirat.*)
- Haader L. 2009. Írásbeli megakadási jelenségek történeti pszicholingvisztikai szemszögből. *Magyar Nyelvőr* 133: 48–65.
- Haader L. 2014. A kritikai forráskiadások egy hozadékáról: Elvi megfontolások egy ómagyar hibatipológiához. In: Laczkó K. – Tátrai Sz. (szerk.): *Elmélet és módszer: Nyelvészeti tanulmányok*. Budapest: ELTE Eötvös József Collegium 87–103.
- Haader L. 2015. Hibázások és hátterük: Apor-kódex. Forgács T. – Németh M. – Sinkovics B. (szerk.) *A nyelvtörténeti kutatások újabb eredményei VIII*. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék. 65–79.
- Haader L. 2017. Egy neurolingvisztikai eset a 16. század elejéről: az Apor- és a Lányi-kódex közös keze. In: Forgács Tamás – Németh Miklós – Sinkovics Balázs (szerk.) *A nyelvtörténeti kutatások újabb eredményei IX*. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék 95–108.
- Haader L. 2019. A meglepetések misszilise – avagy újra Elena prioritással leveléről. In: Haader L. – Juhász D. – Korompay K. – Szentgyörgyi R. – Terbe E. – C. Vladár Zs. (szerk.): *Forráskutatás, forráskiadás, tudománytörténet III*. Budapest: Magyar Nyelvtudományi Társaság, ELTE BTK Magyar Nyelvtudományi és Finnugor Intézet. 419–428.
- Kocsis Zs. 2020. *Kora újkori levélnyelvek. Kézazonosítási lehetőségek elmélete és módszerei*. Budapest: Szépirodalmi Figyelő Alapítvány. (Gondolat-Jel 3.)
- Kocsis Zs. 2022. Hogyan nevezzek? Egy misszilisírnok azonosításának lehetőségei és korlátai. In: Forgács T. – Németh M. – Sinkovics B. (szerk.), *A nyelvtörténeti kutatások újabb eredményei XI*. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék. 229–240.
- S. Sárdi M. 2000. *Bécsi utazások. 17–18. századi útinaplók*. Budapest. Tertia Kiadó.
- S. Sárdi M. 2014. *Napló-könyv. Magyar nyelvű naplók 1800 előtt*. Máriabesnyő: Attraktor Kiadó.
- Varga M. 2020. Rontás van a szövegen? A lejegyzési gyakorlatról boszorkányperekben. In: Boros I. (szerk.): *Fideliter servanda. II. Scriptorium konferencia, Pannonhalma, 2018. május 7–8*. Budapest: Szent István Társulat. 283–301.
- Varga M. 2022a. *Kommunikáció és szövegformálás boszorkányperekben*. Budapest: Szépirodalmi Figyelő Alapítvány. (Gondolat-Jel 6.)
- Varga M. 2022b. Pragmatikai markerekről 18. századi regiszterekben: összehasonlító elemzés. In: Forgács T. – Németh M. – Sinkovics B. (szerk.) *A nyelvtörténeti kutatások újabb eredményei XI*. Szeged: Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék. 431–445.

A SZÓVARIABILITÁS SAJÁTOSSÁGAI 4;6-4;11 ÉVES, TIPIKUS BESZÉDFEJLŐDÉSŰ ÉS BESZÉDHANGHIBÁK TÜNETEIT MUTATÓ GYERMEKEK BESZÉDPRODUKCIÓJÁBAN

Hunyadi-Kóbor Zita – Tar Éva

ELTE Bárczi Gusztáv Gyógypedagógiai Kar

hkzita@student.elte.hu, tar.eva@barczy.elte.hu

A beszédhanghibák terminus a gyermekkori kommunikációs zavarok egyik típusát jelöli, melynek sajátossága, hogy a beszédhangok produkciója elmarad az életkor alapján várttól. A problémakörrel való gondolkodás a logopédiában az elmúlt évszázadban folyamatosan változott, fejlődött. Jelenleg a pszicholingvisztikai szemléletmód, és a zavar ernyőterminussal jelölése a szakmában általánosan elfogadott (Tar, 2017). Az elmaradás okát, felszíni tüneteit tekintve heterogén populáció felosztására több próbálkozás is történt, melyek közül az egyik legtöbbet hivatkozott Dodd (2014) nyelvi-kognitív szemléleten alapuló klasszifikációs rendszere. A szerző által létrehozott felosztásban az ismeretlen eredetű beszédhanghibák öt alcsoportja különíthető el, melyek közül kettő, a konzisztens- és az inkonzisztens fonológiai zavar elválasztását egyetlen differenciáldiagnosztikai kritérium, az ismételt beszédvariabilitásának mértéke teszi lehetővé: inkonzisztens fonológiai zavar diagnosztizálható, ha a gyermek egy adott szó háromszori (szünetekkel megszakított) ismétlése során az ismétlődő itemek minimum 40%-ában legalább kétféle ejtést produkál és ez a beszédbeli sajátosság nem az artikulációs szervek mozgásának elmaradásából adódik. Dodd angol anyanyelvű, tipikus beszédfejlődésű gyermekek ismételt beszéd-variabilitás adatait figyelembe véve fogalmazta meg a fenti differenciáldiagnosztikai javaslatát. A mutató alkalmazásával tehát véleménye szerint elkülöníthető a fonológiai természetű elmaradás konzisztens és inkonzisztens típusa, amelyek háttérben a feltételezések szerint eltérő nyelvi-kognitív műveleti zavar áll, ebből adódóan pedig az érintett gyermekek terápiája eltérő intervenciók stratégiát követ. A klasszifikációs rendszer publikálását követően a 4-5 éves életkori sáv különböző anyanyelvű gyermekek esetében is szerepelt már szóvariabilitás-vizsgálatokban, ezek az eredmények a beszéd konzisztenciájának magas arányát mutatták (Holm et al., 2022). Magyar anyanyelvű gyermekek vizsgálatáról tudósító publikáció azonban a témában még nem született.

Jelen kutatás célkeresztjében a magyar fonológiai fejlődés tipikus, illetőleg atipikus menetét követő négyéves gyermekek ismételt beszédének variabilitása áll. Vizsgálatunkban három kérdésre kerestük a választ. Az első kérdés arra irányult, hogy a tipikus fonológiai fejlődésű mintában milyen mennyiségi (százalékos) átlagértékeket mutatnak a szóvariabilitás egyes altípusai; illetve összesítve a konzisztens; és variábilis válaszok. A második kérdés azt vizsgálta, hogy a fonológiai típusú beszédhanghibák egymáshoz viszonyított aránya miként alakul akkor, ha az inkonzisztens beszéd regisztrálására a dodd-i $\geq 40\%$ -os kritériumot használjuk, illetve akkor, ha a tipikus fonológiai fejlődésű csoportban kapott átlagos variabilitásértéktől való ≥ 1 , vagy $\geq 1,5$ szórásnyi eltérést vesszük alapul. A harmadik kérdéssel azt kívántuk feltárni, hogy a két vizsgálati minta eredményei mutatnak-e eltérést a szóvariabilitás különböző típusainak előfordulási arányában, és amennyiben igen, az miben nyilvánul meg.

A kutatásban magyar egynyelvű 4;6-4;11 éves, tipikusan fejlődő (12 fő) és beszédhanghiba tüneteit mutató (6 fő) gyermekek vettek részt. A csoportba sorolás a beszédprodukciónak pontossága és a fonémaállomány fejlettsége alapján történt, a Tar (2017; 2018) által közölt normatív adatok figyelembevételével. A beszédhanghibákon belüli differenciáldiagnózis fonológiai folyamat-elemzésen alapult S. Tar (2006) szerint. Mindezen elemzések a magyar szegmentális fonológiát reprezentáló 125 szóból álló szólista képmegnevezéssel kiváltott

szóprodukción valósultak meg. A szóvariabilitás előfordulási gyakoriságának és mintázatának vizsgálata szintén képmegnevezéssel kiváltott beszédprodukciónban történt, melynek során a gyermekek a fenti szólistából kiválasztott 25 szót pár perc megszakítással még kétszer ismételték meg. A szóanyag 1-4 szótagú, egyszerű és komplex szótagszerkezetű szavakat tartalmazott. Az itemenként összesen háromszori realizációkat a variabilitás és a pontosság dimenzióiban elemeztük, a megvalósulásokat a Holm és munkatársai (2007) által is alkalmazott négy alkategóriába soroltuk: (1) konzisztensen pontos, (2) konzisztensen hibás, (3) variábilis pontos válasszal, (4) variábilis pontos válasz nélkül. Annak eldöntéséhez, hogy a beszéd variabilitása összefügg-e az artikulációs szervek működésével, a nyelv és ajkak mozgását egyszerű feladatvégzés közben megfigyeltük. Az artikulációs szervek vizsgálatára a szóanyag második ismétlését követően került sor, a Logopédiai vizsgálatok kézikönyve (Juhász, 2007) alapján. Eredményeink szerint a mintában szereplő tipikus fejlődésű gyermekek beszéde nagymértékben konzisztens volt (átlag: 92,67%, szórás: 5,07), variábilis válaszaik pedig többségében az érést tükrözték. Ezen eredmények egybevágnak a Holm és munkatársai (2007) által vezetett, brit gyermekmintán készült vizsgálat megállapításaival, valamint – a konzisztens válaszok túlsúlya tekintetében – más, egyéb nyelveken készült mérésekkel is. A beszédhanghiba tüneteit mutató csoportban a konzisztencia átlagos mértéke 67,33% (szórás: 10,25); a variábilis produkció átlagos értéke (32,67%) megközelíti, bár nem éri el a Dodd által javasolt kritériumértéket. Az eltérő besorolási kritériumok eltérő beszédhanghiba-alcsoport arányokhoz vezettek. A tipikus fonológiai fejlődéshez történő hasonlítás mindkét szórás-kritériummal 100%-ban inkonzisztens fonológiai zavaros gyermekekből álló, homogén csoportot eredményezett. Ezzel szemben akkor, ha a Dodd-i 40%-ot vesszük az inkonzisztens zavar differenciáldiagnosztikai kritériumaként, akkor az inkonzisztens fonológiai zavar alacsonyabb előfordulási aránya igazolódik a konzisztens fonológiai zavarhoz képest, a Dodd (2014) által leírtakhoz hasonlóan. A variabilitás mintázata a két minta esetén eltérő volt, az egyes típusok gyakoriságának sorrendje a következőképp alakult: a tipikus fonológiai fejlődésű csoportban a leggyakoribb választípus a konzisztensen pontos, a legritkább a variábilis pontos válasz nélkül volt, az atipikus fonológiai fejlődésű csoportban pedig a két másik típus fordult elő legnagyobb (konzisztensen hibás) és legkisebb (variábilis pontos válasszal) arányban.

Bár kutatásunk nem reprezentatív, kis elemszámú mintákon került felvételre, fontossága nem elhanyagolható a hazai logopédiai gyakorlat szempontjából: legfontosabb eredménye, hogy atipikus fonológiai fejlődés esetén az ismételt beszéd variabilitásának elemzéséhez használt különböző kritériumok eltérő beszédhanghiba-alcsoportba történő besorolást eredményeznek egyes gyermekeknél. Ez felveti a kérdést, hogy vajon milyen mutató alkalmazása vezet olyan alcsoportba soroláshoz, melyből adott gyermek a nála ez alapján alkalmazásra kijelölt intervencióból a leginkább profitál, hiszen az egyes alcsoportokra a szakirodalom eltérő terápiatípusokat javasol. Mindez dilemma elé állítja a szakmát ezen mutatók klinikai alkalmazhatóságának tekintetében, és előrevetíti a további vizsgálatok nélkülözhetetlenségét a szóvariabilitás területén mind tipikus, mind atipikus beszédfejlődésű gyermekek esetében.

Irodalom

- Dodd, B. 2014. Differential Diagnosis of Pediatric Speech Sound Disorder. *Current Developmental Disorders Reports* 1. 189–196. <https://doi.org/10.1007/s40474-014-0017-3>
- Holm, A. – Crosbie, S. – Dodd, B. 2007. Differentiating normal variability from inconsistency in children's speech: Normative data. *International Journal of Language and Communication Disorders*, 42/4. 467–486. <https://doi.org/10.1080/13682820600988967>

- Holm, A. – van Reyk, O. – Crosbie, S. – De Bono, S. – Morgan, A. – Dodd, B. 2022. Preschool children’s consistency of word production. *Clinical Linguistics & Phonetics* 37/3. 223-241. <https://doi.org/10.1080/02699206.2022.2041099>
- Juhász, Á. (szerk.) 2007. *Logopédiai vizsgálatok kézikönyve*. Budapest: Logopédia Kiadó.
- Sebestyén Tar, É. 2006. *A 3-6 éves kori fonológiai fejlődés kronológiai mintázata a magyarban*. Budapest: Open Art.
- Tar, É. 2017. *Fonológiai fejlődés, variabilitás, beszédhanghibák*. Budapest: ELTE Eötvös Kiadó
- Tar, É. 2018. Word-Initial tap-trill clusters: Hungarian. *Clinical Linguistics & Phonetics* 32/5-6. 544-562. <https://doi.org/10.1080/02699206.2017.1363292>

IDŐZÍTÉS A SPONTÁN BESZÉDBEN – TANULÁSBAN AKADÁLYOZOTT FIATALOK KÖZÖTTI ELTÉRÉSEK

Jankovics Julianna^{1,2}

¹Eötvös Loránd Tudományegyetem

²BGÉSZC Öveges József Technikum és Szakképző Iskola

jankovics.julianna@btk.elte.hu

1. Bevezetés

A tanulásban akadályozottak körébe olyan nehezen tanuló, illetve enyhén értelmi fogyatékosnak minősített emberek tartoznak, akik a gyógypedagógiai pszichológia komplex vizsgáló eljárásaival megállapított intellektuális képességzavar következtében kialakuló átfogó és tartós tanulási nehézséggel élnek együtt (Mesterházi–Szekeres szerk. 2021: 114). A jelen vizsgálat fókuszában az enyhe értelmi fogyatékosokkal élők állnak, akik egy csoportját képezik a tanulásban akadályozott személyeknek. Az értelmi fogyatékosokkal élők esetében általánosságban megállapítható az általános információfeldolgozó funkciók, továbbá a gondolkodás és a téri tájékozódás sérülése, amelyek befolyásolják a beszédprodukción és a beszédpercepciót (Lukács–Kas 2014). Szinte minden esetben jellemző rájuk a megkésett beszédfejlődés, emiatt nem ritka, hogy bizonyos beszédfejlődési szakaszon hosszabb ideig megrekednek (Hatos 2008, Lányiné Engelmayer 2017). A nyelv- és a beszédfejlődés különböző területein kisebb-nagyobb mértékben elmaradnak a tipikus fejlődésű személyektől, jellemző erre a csoportra többek között a szűkös szókincs, a beszédhibák megjelenése, a metakommunikáció nehezítettsége (Radványi 2005, Mesterházi–Szekeres szerk. 2021). A legtöbb kutatás az értelmi fogyatékos személyek beszédének tartalmi részét vizsgálta, a szupraszegmentális szerkezet jellemzőit is főként középsúlyos értelmi fogyatékos gyermekek és felnőttek beszédében elemezték (vö. az alaphangjellemzőkre vonatkozóan Albertini et al. 2010, Seifpanahi et al. 2011, Zampini et al. 2016). Magyar nyelven a temporális jellemzőket megemlítő két kutatás fókuszában is a középsúlyos Down-szindrómával élő (DSZ) személyek állnak. Homor (2008) 10 Down-szindrómás (az átlagos életkor 13,8 év) és 10 más értelmi fogyatékosokkal élő (a tanulmányból nem derül ki, hogy pontosan milyen súlyossági kategóriát ért ez alatt a szerző; az átlagos életkor 14,1 év) fiatal társalgásában és narratív beszédében lévő beszédhibák mellett a beszédtempót is elemezte. Az eredmények szerint mindkét beszéd típusban a DSZ személyeknél volt gyorsabb az átlagos beszédtempó. A narratívában a DSZ gyermekek beszédtempója 36,42 szó/perc volt, míg a más értelmi fogyatékosokkal élők értéke 30,19 szó/perc volt. A társalgásban a DSZ személyeknél 28,92 szó/perces értéket, a másik csoportnál 25,50 szó/perces értéket mért. Rohár (2016) 15 és 18 éves kor közötti DSZ fiatal és más értelmi fogyatékosokkal élő személy (ebben a tanulmányban sincs részletezve az értelmi fogyatékos típusa), valamint 4 és 7 éves kor közötti tipikus fejlődésű óvodás szókincsét, morfológiai és szintaktikai készségét, valamint beszédtempóját elemezte. A feladatban egy képtörténetet kellett elmesélniük, az eredmények pedig azt mutatták, hogy a DSZ személyek átlagosan 26,07 szót, a más értelmi fogyatékosokkal élők 47,45 szót, míg az óvodások 65,27 szót mondtak percenként. A csoportok közötti eltérések szignifikánsak voltak. Jankovics és Garai (2019) enyhe értelmi fogyatékos nők megakadásjelenségeit (többek között a szünetezést) elemezték a spontán beszéd különböző típusaiban, és ezeket az értékeket vetették össze a némben és korban illesztett kontrollszemélyekével. Eredményeik szerint mind a néma, mind a kitöltött szünet hosszabb időtartamban valósult meg az enyhe értelmi fogyatékosoknál, továbbá mindkét csoportban a néma szünet fordult elő leggyakrabban (minden beszéd típusban).

A kontrollpopulációra vonatkozóan számos kutatás alapján elmondható, hogy a beszéd típus nagyban befolyásolja a különböző időzítési jellemzőket, mint például a tempót, a beszédszakaszok időtartamát és a szünetezést (vö. Markó 2014). Továbbá a beszélő nemére vonatkozóan ellentmondásosak az eredmények: Gocsál (2001) eredményei szerint a nők beszédtempója gyorsabb a férfiakénál, több kutatás viszont azt bizonyította, hogy a fiúk, férfiak beszélnek gyorsabban (például Markó 2015, Tóth 2017).

A jelen kutatás célja, hogy megvizsgálja a tanulásban akadályozott fiatalok csoportjába tartozó enyhe értelmi fogyatékos fiatalok spontán beszédében az időzítési jellemzőket, majd összehasonlítsa azt a nemben és korban illesztett kontrollszemélyek értékeivel. A csekély számú nemzetközi és magyar kutatás alapján a vizsgálat hipotézisei szerint (1) a spontán beszéd különböző típusainak a temporális mintázatában különbség mutatkozik a tanulásban akadályozott és a kontrollszemélyek között; valamint (2) a nemek között eltérés tapasztalható a beszéd típusokat és a mentális állapotot tekintve.

2. Módszertan

A jelen kutatásban 16 tanulásban akadályozott fiatal (8 nő, átlagos életkor: 19,4 év; 8 férfi, átlagos életkor: 19,6 év), valamint hozzájuk nemben és korban illesztett 16 kontrollszemély (a nők átlagos életkora: 20,0 év; a férfiak átlagos életkora: 19,5 év) hanganyagait elemeztem. Mindegyik tanulásban akadályozott személy egy-egy budapesti speciális szakiskola tanulója, akiket a különböző szakértői bizottságok enyhe értelmi fogyatékosnak diagnosztizáltak (BNO-kód: F70; BNO-10, 1995). A saját gyűjtésű hanganyagok között kétrészes interjú, képleírás és történetmondás szerepelt, ezek képezik a jelen kutatás spontán beszéd anyagát. Mindegyik beszéd típus a GABI adatbázis (Bóna et al. 2014), valamint a Tóth (2017) által használt feladatok alapján lett kiválasztva. A kutatás etikai szempontokat figyelembe véve mindegyik adatközlő szóban és írásban is tájékoztatást kapott a vizsgálat menetéről, a feladatok típusáról, az enyhe értelmi fogyatékos adatközlők esetében a szülő, a gyám vagy a gondnok írásban nyilatkozott arról, hogy a kísérletvezető megismerheti a szakértői bizottságok által kiállított dokumentációt. A hanganyagot a Praat programmal (Boersma–Weenink 2016–2020) annotáltam, a temporalitásra vonatkozó adatokat egy erre a célra készített script segítségével határoztam meg. Az időzítésre vonatkozó elemzések – beszédszakasz-gyakoriság, artikulációs és beszédtempó, átlagos szótagidőtartam, szünetarány – során a tanulásban akadályozott személyek és a kontrollszemélyek értékeit vetettem össze a beszéd típusok és a nemek szerint. A statisztikai elemzéshez az SPSS 20.0 szoftvert használtam.

3. Eredmények

Az eredmények alátámasztották az első hipotézist, amely szerint a temporalitás egyes tényezőiben különbség mutatkozik a tanulásban akadályozott és a kontrollszemélyek értékei között. Ennek hátterében az áll, hogy az enyhe értelmi fogyatékosoknál bizonyos képességek másképpen működnek a kontrollszemélyekhez képest, és a különböző feladatokban aktiválódó készségek (kognitív, narratív készségek, memória) nagyban befolyásolták adatközlői és csoportszinten is a temporális tényezők alakulását. A nemek közötti különbségek is igazolódtak a legtöbb paramétert tekintve, de ezek elsősorban a tanulásban akadályozott személyeknél mutatkoztak meg. A beszédszakaszok gyakorisága mind a négy csoport esetében az interjú második részében volt a legnagyobb. A tempóértékek minden esetben az enyhe értelmi fogyatékos csoportokban voltak kisebbek. Mindegyik csoport esetében a tartalomösszegzésben volt a legnagyobb az átlagos szótagidőtartam. Továbbá a nőknél mértem magasabb értékeket a férfiakhoz képest. Mindegyik csoportnál az interjú első részében volt a legtöbb, míg az interjú második részében volt a legkevesebb szünet, továbbá a néma szünetek voltak a leggyakoribbak a jellel kitöltött és a kombinált szünetekhez képest. A csoporton belüli egyéni különbségek jelentősek voltak a szünetek gyakoriságát kivéve mindegyik változó esetében. Továbbá a vizsgálat érdekessége volt, hogy a kontrollférfiak által produkált mintázat a legtöbb esetben teljesen ellentétes volt a másik három csoporthoz viszonyítva.

4. Összegzés, kitekintés

Mind a nemzetközi, mind a magyar irodalmat nézve látható, hogy a tanulásban akadályozott, enyhe értelmi fogyatékosággal élő személyek spontán beszédét szupraszegmentális szempontból elemző művek száma elenyésző. A jelen kutatás eredményei több területen is felhasználhatóvá válhatnak (fonetika, pszicholingvisztika, gyógypedagógia). Továbbá a kutatás korpusza több vizsgálat alapját is képezheti, példaként említve a tanulásban akadályozott fiatalok beszédének zöngjellemzőit, megakadásjelenségeit és dallamszerkezetét.

Irodalom

- Albertini, G. – Bonassi, S. – Dall’Armi, V. – Giachetti, I. – Giaquinto, S. – Mignano, M. 2010. Spectral analysis of the voice in Down syndrome. *Research in Developmental Disabilities* 31/5. 995–1001.
- BNO-10 1995. *A betegségek és az egészséggel kapcsolatos problémák nemzetközi statisztikai osztályozása. 10. revízió. 1. kötet.* Népjóléti Minisztérium, Budapest.
- Boersma, P. – Weenink, D. 2016–2020. *Praat: doing pfonetics by computer (Version 6.0.43).* http://www.fon.hum.uva.nl/praat/download_win.html (A letöltés ideje: 2016. szeptember)
- Bóna J. – Imre A. – Markó A. – Váradi V. – Gósy M. 2014. GABI – Gyermeknyelvi beszédAdatBázis és Információtár. *Beszéd kutatás 2014.* 246–251.
- Gocsál Á. 2001. Gyorsabban beszélnek-e a nők, mint a férfiak? *Beszéd kutatás 2001.* 61–72.
- Hatos Gy. 2008. *Az értelmi akadályozottsággal élő emberek: nevelésük, életük.* 4. átdolgozott, időszerű kiegészítésekkel bővített kiadás. Gyula: APC Stúdió.
- Homor V. 2008. *Down-szindrómások és más értelmi akadályozottak spontán beszédének összehasonlítása.* Szakdolgozat. Budapest: ELTE BGGYK.
- Jankovics J. – Garai L. 2019. Disfluencies in mildly intellectually disabled young adults’ spontaneous speech. In: Rose, R. L. – Eklund, R. (eds.): *Proceedings of DISS 2019.* Budapest: ELTE. 79–82.
- Lányiné Engelmayer Á. 2017. *Intellektuális képességszavar és pszichés fejlődés.* Második, átdolgozott és bővített kiadás. Budapest: Medicina Könyvkiadó.
- Lukács Á. – Kas B. 2014. Nyelvvelésajátítás és értelmi fogyatékoság. In: Pléh C. – Lukács Á. (szerk.): *Pszicholingvisztika 1–2: Magyar pszicholingvisztikai kézikönyv.* Budapest: Akadémiai Kiadó. 1383–1404.
- Markó A. 2014. A beszéd temporális szerkezete a beszédmód és a beszédhelyzet függvényében. In: Bátyi Sz. – Navracsics J. – Vigh-Szabó M. (szerk.): *Nyelvelésajátítási-, nyelv tanulási- és beszéd kutatások.* Budapest, Veszprém: Gondolat Kiadó, Pannon Egyetem Modern Filológiai és Társadalomtudományi Kar. 33–45.
- Markó A. 2015. *A spontán beszéd prozódiai szerkezete. Időzítés és beszéddallam.* Budapest: Akadémiai Kiadó.
- Mesterházi Zs. – Szekeres Á. (szerk.) 2021. *A nehezen tanuló gyermekek iskolai nevelése. Egyetemi tankönyv a Gyógypedagógia szak Tanulásban akadályozottak pedagógiája szakirány számára. 2., javított kiadás.* Budapest: ELTE BGGYK.
- Radányi K. 2005. A kommunikáció és a beszéd fejlesztése az értelmileg akadályozott gyermekeknél. In: Varga I. (szerk.): *Speciális didaktika I. Az értelmi akadályozottsággal élő gyermekek tanítása.* Szeged: Szegedi Tudományegyetem, Juhász Gyula Tanárképző Főiskolai Kar. 28–69.
- Rohár A. 2016. Down-szindrómás személyek nyelvi képessége. *Gyógypedagógiai Szemle* 44/3. 195–215. <https://epa.oszk.hu/03000/03047/00074/pdf/> (A letöltés ideje: 2021. január 10.)
- Seifpanahi, S. – Bakhtiar, M. – Salmalian, T. 2011. Objective vocal parameters in Farsi-speaking adults with Down syndrome. *Folia Phoniatrica et Logopaedica* 63/2. 72–76.

- Tóth A. 2017. *A spontán beszéd a nem és az életkor függvényében gyermek- és fiatal felnőttkorban*. PhD-értekezés. Budapest: ELTE BTK.
- Zampini, L. – Fasolo, M. – Spinelli, M. – Zanchi, P. – Suttora, C. – Salerni, N. 2016. Prosodic skills in children with Down syndrome and in typically developing. *International Journal of Language & Communication Disorders* 51/1. 74–83.

A HIÁTUSTÖLTŐ ÉS FONEMIKUS [j]-VARIÁNSOK AKUSZTIKAI MEGVALÓSULÁSA VALÓDI SZAVAKBAN A NYELVÁLLÁS AKUSZTIKAI VETÜLETÉNEK ÉS A VOKALIKUS SZAKASZUK IDŐTARTAMÁNAK SZEMPONTJÁBÓL

Deme Andrea¹, Juhász Kornélia^{1,2}

¹ Eötvös Loránd Tudományegyetem

² HUN-REN Nyelvtudományi Kutatóközpont

deme.andrea@btk.elte.hu, juhasz.kornelia@nytud.hun-ren.hu

A beszédben megjelenő beszédhangok számos okból mutatnak változatosságot: ugyanazon beszédhang kisebb-nagyobb mértékben eltérő módon valósul meg például különböző beszélők esetében, de eltéréseket okozhat az, ha a beszéd lassabb vagy gyorsabb, illetve az is, ha ugyanazon beszédhangok más-más beszédhangok szomszédságában állnak. Felmerül, hogy a változatosság egy különleges esete az, amikor egy adott beszédhang eltérően jelenik meg a beszéd reprezentációjában az egyik és a másik szóban vagy fonetikai helyzetben. Ilyen például a magyar [j], amelyről azt feltételezzük, hogy a) bizonyos esetekben teljesértékű fonémaként megjelenik a szavak absztrakt reprezentációjában, illetve a helyesírás is rögzíti (pl. *zoknija*), míg b) más esetekben nem része az absztrakt reprezentációnak, és a megvalósulása fiziológiai törvényszerűségek, illetve a hiátustöltés fonológiai folyamatának bejósolható következménye, így a helyesírás sem jelöli (pl. *szoknia*). Feltételezhető, hogy ezeknek a körülményeknek hatása van az egyébként minőségében azonosnak vélt beszédhang kétféle helyzetben tapasztalható megvalósulására, amely feltételezés mind az elméleti, mind a kísérletes szakirodalomban felbukkan.

A jelen akusztikai fonetikai kísérletben a fentebb említett két [j]-variáns realizációját elemezzük valódi, jelentéssel rendelkező szavakban: a hiátustöltőként megjelenő [j]-variánst főnévi igenevekben (pl. *szoknia*), a fonemikus [j]-t pedig toldalékolt főnevekben (pl. *zoknija*), de szegmentálisan azonos környezetben vizsgáljuk. A kísérletben e két [j]-variánst dinamikus elemzéssel hasonlítjuk össze az /iɒ/ és /ijɒ/ vokalikus szakaszok időtartama, illetve a képzés során létrejövő szűkület akusztikai vetületének szempontjából. Egyes fonológiai elemzések szerint a fonemikus /j/ realizációja „mássalhangzóssabb”, mint a hiátustöltő [j]-é, ezért a fonetikai megvalósításuk is eltér (Siptár–Törkenczy 2000, Siptár 2013). Míg a /j/ képzésében magasabb nyelvállásra számítunk, azaz jelentősebb toldalékcso- szűkületre, addig a hiátustöltő [j] esetében például az /ia:/ hangsorban pusztán egy vokalikus hangátmenet megjelenését várjuk a /j/-nél nyíltabb, felső nyelvállású, elől képzett /i/ és a legelső nyelvállású centrális /a:/ között úgy, hogy abban nem találunk egy önálló, a hiátustöltő [j]-hez tartozó akusztikai/artikulációs célt, azaz jelentősebb szűkületet a toldalékcso- bben (vö. (Siptár–Törkenczy 2000, Siptár 2013, Davidson–Erker 2014). Korábbi akusztikai vizsgálataink minimális párokba rendezett álszavakon részben megerősítették ezt a feltételezést, ahol a toldalékcso- bben megjelenő szűkület nagyságát az első formáns (F₁) frekvenciájának értékével számszerűsítettük, ami a szűkület átmérőjével fordított arányban áll (Hunt 2003, Jagers 2018). Az /ia:/ és /ija:/ vokalikus hangsorok F₁-görbéit összevetve az /i/ és /a:/ akusztikai céljai közötti tranzíciós fázis realizációjában (ahol a [j]-variánsok megjelenésére számítottunk) a fonemikus [j]-variánst tartalmazó F₁-menet a hiátustöltő [j]-hez képest alacsonyabb F₁-frekvenciát, azaz kisebb szűkületet mutatott (Juhász–Deme 2023). Ugyanezen vizsgálat és korábbi eredmények továbbá az időtartam tekintetében is alátámasztották azt, hogy a hiátustöltő [j]-variáns „gyengébb és átmenetibb jellegű” (Siptár–Törkenczy 2000: 91): ez a beszédhang időtartamában is rövidebben valósult meg, mint a fonemikusan megjelenő [j] (Gósy 2014, Juhász–Deme 2022, 2023). A jelen kísérletben arra keressük a választ, hogy a hiátustöltő és fonemikus [j]-variánsok között korábban megmutatott akusztikai eltérések valódi, jelentéssel bíró szavak esetében is megjelennek-e.

A kísérletben 15 magyar anyanyelvű felnőtt női beszélő produkcióját vizsgáljuk akusztikailag. (A nem kontrollálására azért volt szükség, hogy elminimáljuk a férfiak és a nők testméretbeli különbségeiből fakadó, akusztikai eltéréseket okozó hatásokat.) A vizsgált vokalikus /iɒ/ és /ijɒ/ célszekvenciákat szótagszámukban, illetve a közvetlen hangkörnyezetet illetően is megegyező kontextusban elemezzük létező szavakban, de úgy, hogy a kérdéses szavak a teljes szóalakra nézve nem képeznek minimális párt (pl. *szoknia* vs. *zoknija*). Ennek megfelelően a beszélők a hangsorok produkciója során vélhetően nem voltak előfeszítve arra, hogy a szavak közötti, a [j] megvalósulásában tetten érhető különbségeket kihangsúlyozzák, illetve felerősítsék. A célszavakat rövid mondatokba ágyaztuk, amely megegyezett minden vizsgált szó esetében (*Legyen <célszó>*). Az elemzés során egyfelől összehasonlítjuk a fonemikus és hiátustöltő [j]-t tartalmazó szavak vokalikus szekvenciáinak (/ijɒ/ és /iɒ/) időtartamát, másfelől pedig összevetjük a nyelvállásfokkal fordított arányban álló F₁ frekvenciaértékének idői lefutását. Ez utóbbihoz az F₁ értékét 5 ms-onként nyerjük ki az /iɒ/ és /ijɒ/ vokalikus szakaszokon belül a Praat szoftver segítségével (Boersma–Weenink 2020), és az így létrejövő F₁-görbéket GAM-modellekkel

hasonlítjuk össze az R programban (R Core Team 2020). A vokalikus szakaszok időtartamát kevert modellekkel vetjük össze.

A korábbi, álszavas kísérletekből származó eredmények alapján a jelen kísérlet eredményeit illetően azt várjuk, hogy a két vizsgált [j]-variáns tartalmazó vokalikus szakasz elkülönül egymástól mind időtartamában, mind az F₁ menetét illetően. Egyfelől a fonemikus [j]-t tartalmazó /ijɒ/ hangsor hosszabb időtartamú, mint a hiátustöltéssel megvalósuló /iɒ/ hangsor. Másfelől az /ijɒ/ esetében az /i/ és az /ɒ/ közötti tranzíció alacsonyabb F₁-értékekkel (tehát magasabb nyelvállással) valósul meg, mint a hiátustöltő [j]-vel realizálódó /iɒ/ szekvencia esetében. A kísérlet eredményei hozzájárulhatnak a hiátustöltés folyamatának mélyebb megértéséhez, valamint bővítik az ismereteinket arról, hogy a beszéd vélt fonológiai (mögöttes) reprezentációja miként hat a beszédhangok megvalósítására, illetve hogyan függ össze azzal.

Irodalom

- Boersma, P., – D. Weenink 2020. Praat: Doing phonetics by computer [Computer program]. 6.1.15-ös verzió. <http://www.praat.org>
- Davidson, L. – Erker, D. 2014. Hiatus resolution in American English: The case against glide insertion. *Language*, 90/2, 482–514.
- Gósy M. 2014. A palatális közelítőhang kétféle funkcióban. *Beszédkutatás*, 22, 17–40.
- Hunt, E. 2003. *Acoustic Characterization of the Glides /j/ and /w/ in American English*. PhD Disszertáció. Princeton: Princeton University.
- Jaggers, Z. S. 2018. Evidence and characterization of a glide-vowel distinction in American English. *Laboratory Phonology* 9, 1–27.
- Juhász K. – Deme A. 2022. Palatális approximánsok a kínaiiban és a magyarban – a kínai alveolopalatális /ɕ/ szibiláns követő vokális szakasz produkciója kínaiul tanuló magyar anyanyelvűeknél. *Általános Nyelvészeti Tanulmányok XXXIV*. 287–332.
- Juhász K. – Deme A. 2023. A hiátustöltőként és fonémarealizációként megjelenő [j]-variánsok elemzése a nyelvállás akusztikai vetületének szempontjából. Konferenciaelőadás. Beszédkutatás Konferencia 2023. Budapest: Nyelvtudományi Kutatóközpont.
- Siptár P. – Törkenczy M. 2000. *The phonology of Hungarian*. Oxford: Oxford University Press.
- Siptár P. 2013. Palatálisok. In: Benő, A. – Fazakas, E., – Kádár, E. (eds.) „...hogyan legyen a víznek lefolyása...” : *Köszöntő kötet Szilágyi N. Sándor tiszteletére*. Kolozsvár: Erdélyi Múzeum Egyesület, 433–448.

A BESZÉDSAJÁTOSSÁGOK VÁLTOZÁSA 10 ÉV ELTELTÉVEL – A BESZÉDLEIRATOK LEXIKAI VIZSGÁLATA ÉS AZ EREDMÉNYEK FELHASZNÁLHATÓSÁGA A BESZÉLŐAZONOSÍTÁS MŰVELETÉBEN

Főző Eszter

NBSZ Szakértői Intézet

fozo.eszter@nbsz.gov.hu

Ha egy bűnügyben a szakkérdés annak megállapítására irányul, hogy a kérdéses hangfelvétel(ek)en és a – hangtechnikai szakértő által rögzített – hangmintán ugyanannak a személynek a hangja hallható-e, a nyomozóhatóság igazságügyi hangtechnikai szakértői vizsgálat elvégzését rendeli el. A szakértő a beszédhang elemzésével és összehasonlításával elvégzi a személyazonosság megállapítását, azaz a beszélőazonosítást.

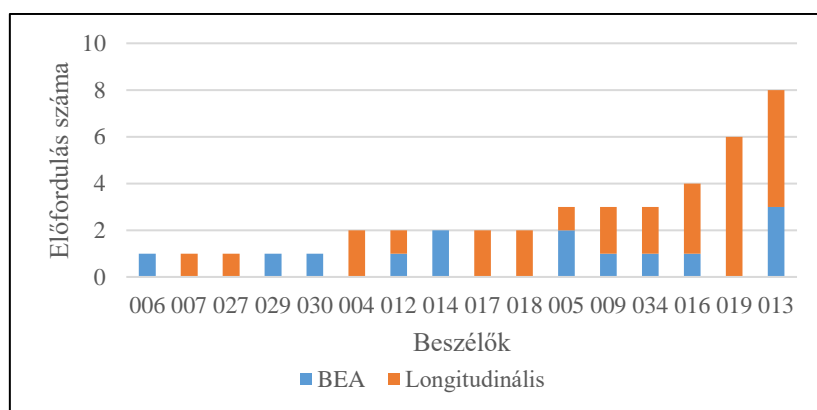
Vannak azonban olyan esetek, amikor az akusztikai-fonetikai és/vagy a biometrikus azonosítás nem végezhető el, például mert a kérdéses hangfelvétel tisztítás után is zajos; a kérdéses hangfelvételen a beszélő szándékosan torzítja a hangját; a kérdéses hangfelvétel túl rövid (Gold–French 2019); a hangrögzítő eszköz távol van a célszemélytől, így a beszéd érthető, de a felvétel az akusztikai-fonetikai mérésekhez nem megfelelő minőségű, vagy felmerül annak a lehetősége, hogy a hangfelvételen hallható beszéd DeepFake technológiával készült. Ilyenkor van érthető és értelmezhető beszédproduktum beszédleirat formájában, melyet a nyelvész szakértő szövegnyelvészeti módszerekkel – elsősorban stilisztikai, morfológiai, lexikai, szemantikai és pragmatikai elemzésekkel – képes feldolgozni.

Annak érdekében, hogy a korábban felsorolt esetekben a nyelvész hatékonyan tudja támogatni a hangtechnikai szakértők feladatvégzését, a beszélt nyelv egyedi sajátosságait kutatjuk. A vizsgálatok során a következő kérdésekre kerestük a választ: amennyiben a beszéd-rögzítés feltételei változatlanok (ugyanaz a beszédhelyzet, ugyanaz a beszédfeladat, esetenként ugyanaz a beszédtema is), változnak-e az idő múlásával ugyanazon személynél a beszélt nyelvi összetevők (szó- és szókapcsolat-használata, szófordulatai), és ha igen, milyen mértékben? Vajon, ha hosszú idő – pl. 10 év – telik el a kérdéses hangfelvétel és a hangminta rögzítési ideje között, megfigyelhetők-e a nyelvhasználatban olyan idiolektális jellemzők, melyeket a beszélőazonosítás művelete során nyelvi bizonyítékként fel lehet használni?

A jelen kutatásban 36 különböző beszélő (18 férfi, 18 nő) beszédsajátosságait elemeztük, különös tekintettel szó- és szókapcsolat-előfordulásokra (diskurzusjelölők, funkciószavak stb.), az előfordulások relatív gyakoriságára (előfordulások száma a szószámhoz viszonyítva), valamint pozíciójára. A korpuszt a BEA adatbázis (Gósy et al. 2012) és a megközelítőleg 10 évvel később rögzített Longitudinális (Grácsi et al. 2020) adatbázis spontán és félspontán beszédadataiból (interjú, társalgás, vélemény) állítottuk össze. Arra törekedtünk, hogy a korpusz ne pusztán a nemek, hanem egyéb csoportjegyek vonatkozásában is kiegyensúlyozott legyen, hiszen a profiljegyek tekintetében a lehető legkevesebb változót hordozó korpuszon vizsgálhatók legjobban a csoportjegyek, emellett a csoportjegyeiktől eltérő egyedi nyelvhasználati jegyek is jobban detektálhatók. Az első (BEA-s) felvétel idején a férfi beszélők 100%-a, a női beszélők 94%-a a kriminalisztikai szempontból releváns életkori csoportba esik (Tatár 2013, Skarnitzl–Vaňková 2017), tehát fiatalok vagy középkorúak; a férfi beszélők 44%-a alap- vagy középfokú, 56%-a felsőfokú iskolai végzettségű, a női beszélőknél ez az arány 50-50%.

Az egyes beszédfeladatok hanganyagait gépi beszédleiratozóval jegyeztük le, majd kézzel ellenőriztük. A szókészlet vizsgálatát a következőképpen végeztük: meghatároztuk a beszélőkhöz tartozó 3-3 beszédfeladat együttes szószámát, és ehhez képest néztük meg a szóelőfordulások gyakoriságát (relatív előfordulás), hiszen a beszéd (szöveg) hossza jelentősen befolyásolja az egyes lexikális elemek előfordulási példányszámát. Emellett az abszolút előfordulásokat is figyeltünk, hiszen például a diskurzusjelölőket elemző szakirodalomra inkább ez a megfigyelési módszer jellemző (Dér–Markó 2010, Schirm 2018). A teljes korpusz 215654 token (28385 szó). A vizsgálatok során a teljes korpusz 500 leggyakoribb szavának előfordulásaira koncentráltunk, azok közül is elsősorban az indulat-, módosító- és határozószókra, a kapcsolódásaikra és esetenként a pozíciójukra.

Példának okáért a 426. helyen álló *ügymond* diskurzusjelölő előfordulási adatai is szolgálhatnak információval. A BEA-s anyagban összesen 14, a Longitudinális anyagban 28 db található belőle, a 36 beszélő közül 16-nál fordul elő legalább egyszer vagy a BEA-s, vagy a Longitudinális korpuszrészben (a korpusz beszélőinek 44%-a), ám mindössze 6 beszélőnél van az *ügymond*-ra adat mindkét alkorpuszban (a korpusz beszélőinek 16,6%-a). A 16 beszélő közül 11-nél a Longitudinális szöveghalmazban fordul elő az *ügymond* többször (a használók 68,75%-a); azon 6 beszélőnél, akiknél mind a BEA-s, mind a Longitudinális alkorpuszban volt találat az *ügymond* szóra, 5-nél fordul elő többször a Longitudinális anyagban (83,3%), tehát ezeknél a beszélőknél van életkori vonatkozása az *ügymond* használatának. Az *ügymond*-ot használó 16 beszélő közül 9 férfi, vagyis valamivel jobban jellemzi a korpuszban lévő férfiak nyelvhasználatát (a 6 konzekvens használó közül 5 férfi). Mind a közép-, mind a felsőfokú iskolai végzettségű beszélőknél előfordul, a magasan iskolázottaknál valamivel gyakrabban (56,25%). A korpusz beszélőinek 80,5%-a következetes az *ügymond* használatát illetően (vagy nem használja egyáltalán, vagy közel azonos előfordulással használja): az egyes beszélőkhöz tartozó BEA-s és Longitudinális anyagokban az *ügymond* előfordulási számadatok különbsége 0 vagy 1. Mivel egy ritka előfordulású lexikai elemről van szó, a relatív gyakorisági értékek nagyon alacsonyak (0,000162–0,002345). Annyi megállapítható, hogy az *ügymond* szót 10 év távlatában is visszatérően használó 6 beszélőnél a BEA-s és Longitudinális relatív gyakorisági mutatók között alacsony a differencia (0,000258 és 0,00139 közötti), így a korábbi előfordulási számadatra vonatkozó megállapítást a relatív gyakorisági értékek is alátámasztják.



1. ábra 'Ügymond'-előfordulások számának alakulása beszélőnként és az alkorpuszokban

Irodalom

Gold, E. – French, J. P. (2019) International Practices in Forensic Speakers Comparisons: Second Survey. *Internatinal Journal of Speech, Language and the Law*. Vol. 26. No. 1. 1–20. doi: 10.1558/ijssl.38028

- Gósy, M. – Gyarmathy, D. – Horváth, V. – Grácsi, T. E. – Beke, A. – Neuberger, T. – Nikléczy, P. (2012) BEA: Beszélt nyelvi adatbázis. Gósy M. (szerk.) *Beszéd, adatbázis, kutatások*. Akadémiai Kiadó. 9–24.
- Grácsi, T. E. – Huszár, A. – Krepsz, V. – Száraz, B. – Damásdi, N. – Markó, A. (2020) Longitudinális korpusz magyar felnőtt adatközlőkről. Berend G., Gosztolya G. és Vincze V. (szerk.), *XVI. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, Szegedi Tudományegyetem Informatikai Intézet. 103–115.
- Markó A. – Dér Cs. I. (2011) Diskurzusjelölők használatának életkori sajátosságai. In: Navracsics J., Lengyel Zs. (szerk.), *Lexikai folyamatok egy- és kétnyelvű közegben. Pszicholingvisztikai tanulmányok II*. Budapest: Tinta Könyvkiadó. 49–61.
- Schirm A. (2018) Diskurzusjelölő-társulások a Szögedi Szociolingvisztikai Interjúban. *Alkalmazott Nyelvtudomány*, XVIII. évfolyam, 2018/1. szám. doi: 10.18460/ANY.2018.1.007
- Skarnitzl, R – Vaňková, J. (2017) Fundamental frequency statistics for male speakers of common Czech. *Acta Universitatis Carolinae: Philologica*, 2017/3, 7–17. doi: 10.14712/24646830.2017.29
- Tatár Z. (2013) Beszélőprofil-alkotás lehetőségei a kriminalisztikai fonetikában. *Alkalmazott Nyelvtudomány*, XIII/1-2, 121–130.

Köszönetnyilvánítás

A kutatást az NKFIH az FK128814-es és az Európai Unió az RRF-2.3.1-21-2022-00004 azonosítójú, Mesterséges Intelligencia Nemzeti Laboratórium projekt keretében támogatta.

AUTOMATIKUS BESZÉLŐPROFIL KÉSZÍTÉSE – AZ ÉLETKOR MEGHATÁROZÁSA

Tatár Zoltán

NBSZ Szakértői Intézet

tatar.zoltan@nbsz.gov.hu

Bűnügyi nyomozás során egy kérdéses hangfelvételen beszélő ismeretlen személyről készült profil segíti a nyomozóhatóságot. A kriminalisztikai beszélőprofil az igazságügyi hangtechnikai szakértők készítik percepciós és fonetikai paraméterek számítógépes elemzésével. Ezirányú kutatások kimutatták, hogy a legpontosabban a beszélő nemét és életkorát lehet meghatározni, az egyéb jellemzők – pl. testalkat, magasság – jóval nagyobb hibaarányal állapíthatók meg (Gocsál, 1998, Gósy, 2001).

A beszélőazonosítás területén az elmúlt két évtizedben kialakult a szakmai igény a biometrikus azonosítás alkalmazására. Ennek előnye a gyorsaság és az objektivitás, ugyanis két beszélő összehasonlítása gyorsabban elvégezhető, mint a fonetikai alapú összehasonlítás, illetve a hasonlóság a valószínűségi arányszámmal (Likelihood Ratio) leírható. Egyes beszélőazonosító alkalmazások már képesek automatikusan nem, életkor és egyéb jellemzők meghatározására is

Jelen kutatásomban a Vocalise nevű beszélőazonosító szoftver beszélőprofil-készítő képességeit, azon belül kifejezetten az életkor-meghatározó képességét vizsgálom. Ez a rendszer nem kínál automatikus lehetőséget erre, vagyis egy ismeretlen beszélő életkorát nem állapítja meg egy gombnyomásra. A kérdéses személyeket összehasonlítom az adatbázisunkban szereplő minden egyes beszélővel (kérdéses beszélő–adatbázis 1. beszélő, kérdéses beszélő–adatbázis 2. beszélő, ...), és a szoftver minden pár esetében ad egy hasonlósági értéket, amely minél nagyobb, annál hasonlóbb a két beszélő hangja. Hipotézisem szerint a magas hasonlósági értéket kapott beszélők életkora is hasonló lesz.

A vizsgálatokhoz a BEA adatbázist (Neuberger et al., 2014), a Longitudinális beszédkorpuszt (Grácsi et al., 2020) és az NBSZ kutatási céllal készített hangfelvételeit használom fel. Az 530 beszélőtől származó összesen 1102 spontán beszédet tartalmazó hanganyag az alábbi eloszlást mutatja: BEA: férfi: 123, nő: 128; Longi: férfi: 14, nő: 11; NBSZ: férfi: 402, nő: 424.

Az életkor meghatározása minden vizsgálati módszer esetében tűrési hibahatár mentén működik. A célom, hogy a vizsgálatokkal a gyakorlatban is alkalmazható módszert dolgozzak ki. A való életben egy nyomozónak az is segítség, ha a rendszer megmondja, hogy egy 30 év alatti vagy egy 50 év feletti személyt keressen, hiszen ennek a megállapítása sem minden esetben könnyű. A gyakorlati tapasztalatom alapján a hipotézisem az, hogy az 5 évenkénti életkor meghatározás túl szűk intervallum, és nagy hibaarányal működne egy ilyen rendszer. Legalább 10 éves korcsoportok alkalmazása reális cél lehet, vagyis az, hogy pl. négy (20–29, 30–39, 40–49, 50–65 éves) korcsoportba legyenek sorolhatók a beszélők.

A kutatást az NKFIH az FK128814-es és az Európai Unió az RRF-2.3.1-21-2022-00004 azonosítójú, Mesterséges Intelligencia Nemzeti Laboratórium projekt keretében támogatta.

Irodalom

- Gocsál, Á. 1998. Életkorbecslés a beszélő hangja alapján. *Beszédkutatás '98*. 122–134.
Gósy, M. 2001 A testalkat és az életkor becslése a beszéd alapján. *Magyar Nyelvőr* 125. 478-488.

- Neuberger, T. – Gyarmathy, D. – Grácsi, T. E. – Horváth, V. – Gósy, M. – Beke, A. 2014. Development of a large spontaneous speech database of agglutinative Hungarian language. In: Sojka, P. – Horák, A – Kopeček, I. – Pala, K. (Eds.): *Proceedings of TSD*. New-York – Berlin: Heidelberg–Springer. 424–431.
- Grácsi, T. E. – Huszár, A. – Krepsz, V. – Száraz, B. – Damásdi, N. – Markó, A. 2020. Longitudinális korpusz magyar felnőtt adatközlőkről. In: Berend, G – Gosztolya, G. – Vincze, V. (Eds.): *XVI. Magyar Számítógépes Nyelvészeti Konferencia kötete*. Szeged: JATEPress. 103–114.

A ZÖNGÉSSÉGI HASONULÁS A NYUGATI NYELVJÁRÁSTERÜLETEN – EGY ESETTANULMÁNY TANULSÁGAI

Markó Alexandra

NBSZ Szakértői Intézet, HUN-REN Nyelvtudományi Kutatóközpont

marko.alexandra.phd@gmail.com

1. Bevezetés

A kriminalisztikában a hangfelvétel alapján történő profilalkotást olyan esetekben lehet alkalmazni, amikor rendelkezésre áll hangfelvétel, de a feltételezett elkövető még nem került a nyomozóhatóság látókörébe. A beszédhangzás alapján bizonyos jellegzetességek megállapíthatók. Ilyen, a származásra vagy a lakhelyre utaló beszédbeli jellegzetesség lehet a beszélő nyelvjárása (vö. pl. Tatár 2013, Tatár–Varga–Főző 2021).

A magyar nyelvjárások elsősorban a magánhangzókészletükben térnek el egymástól (Vargha 2022), ugyanakkor a magánhangzók variabilitása artikulációs sajátosságai miatt egy adott nyelvjáráson belül is igen nagy lehet. Robusztusabb hangtani mintázatokat találhatunk a mássalhangzók fonológiai folyamataiban, és bár a magyar nyelvjárások nem sok példát szolgáltatnak erre, a köznyelvitől eltérő hasonulási folyamatok a gyakorlott elemző számára jó támpontot adhatnak.

A nyelvjárástani szakirodalom széleskörűen dokumentálja azt a jelenséget, hogy a *v* a zöngésségi hasonulásban a köznyelvitől eltérő módon viselkedik (elsősorban) a nyugati nyelvjáróterületen (Imre 1971; Fodor 2001; Juhász 2001). Nemcsak a *v* hasonul zöngésség szerint (ahogyan a köznyelvben, pl. *hívtál* [hi:fta:l]), hanem a Nyugat-Dunántúl északi részén triggerként regresszíven zöngésíti az előtte álló zöngétlen mássalhangzót (pl. *ötven* [ødvɛn], *Vasvár* [vɔʒva:r]; akár szóhatáron is, pl. *sok volt* [ʃog vuot]; vö. Fodor 2001). Emellett néhány területen (Zala megye, a Hetés déli része és Baranya) az is megfigyelhető, hogy ugyanezen fonetikai helyzetekben (zöngétlen obstruens + *v*) a *v* progresszív hasonulás targetje (pl. *hatvan* [ɦɒtfɒn], látva [la:tfɒ], ez azonban szóhatáron már nem érvényesül, vö. Imre 1971). Fodor (2001) arra is kitér, hogy Vas megye déli és Zala északi részén lexémán belül a két jelenség akár ugyanannak a beszélőnek a beszédében szabadon váltakozva is előfordul, amire a *Magyar dialektológia* böi szövegmintájában is találunk példát (Juhász 2001: 269); de akár egyazon szóalakbeli váltakozás is megfigyelhető (pl. *hatvan* [ɦɒdvɒn] vagy [ɦɒtfɒn]).

A szakirodalomban vita tárgyát képezi, hogy a zöngésítő *v*-hez hasonlóan a nazálisok is zöngésítik-e az őket megelőző zöngétlen obstruens. Kiss Jenő (2001) szerint ez a feltételezés tévedés, ugyanakkor több szerző (pl. Fehér 2009; Kis 2016; Bárth 2022) tényként kezeli a nazális előtti obstruens zöngésülését a nyugati nyelvjáróterületen – ha nem is tartják olyan gyakorinak, mint a *v* előtt (pl. *hát nem* [ɦa:d nɛm], *kismiska* [kizmiʃkɒ]).

A jelen kutatás arra irányul, hogy a zöngésségi hasonulásnak a köznyelvitől eltérő hangtani mintázatai segítik-e a beszélő nyelvjárásának a megállapítását, ezzel hozzájárulva a profilalkotáshoz. Ehhez azt vizsgáljuk meg, hogy milyen mintázatok és milyen gyakorisággal jelennek meg egy adott beszélő esetében.

2. Anyag és módszer

A kutatásban egy, a nyugati nyelvjáróterületről származó, és a köznyelvitől eltérő hasonulási mintázatot jól észlelhetően mutató középkorú férfi hangfelvételein vizsgáljuk az érintett mássalhangzó-kapcsolatokat. Az adatközlő győri születésű, de két éves korától 18 éves koráig Celldömölkön élt, ezt követően költözött Budapestre, ahol azóta is él. A szülei mindketten celldömölki születésűek, és a régióban (pl. Pápa, Szombathely) nevelkedtek, és élnek azóta is. Az eddigi elemzések valamivel több mint 11 percnyi hanganyagon készültek (a szüneteket nem számítva), melynek a nagyobb része spontán megnyilatkozás (811 szó), míg kisebb része (664

szó) olvasott szöveg. Mind a felolvasott, mind a spontán szövegben szerepelnek ismétlődő szavak, szerkezetek.

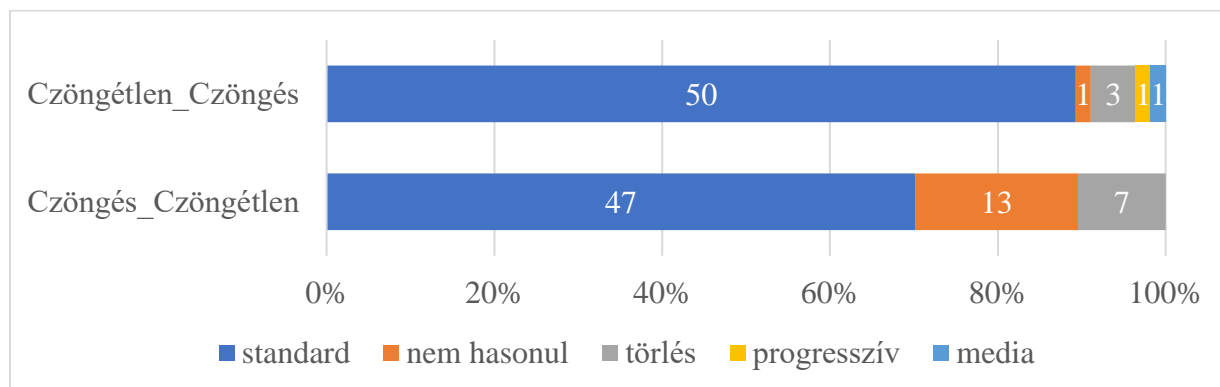
A stúdióban rögzített hangfelvételt a Praat segítségével annotáltuk szöveg- és szószinten, valamint címkéztük a fonémakapcsolatokat és a beszédhangrealizációkat is. A fent említett típusokon (zöngétlen obstruens + *v*, zöngétlen obstruens + nazális) túl vizsgáljuk a zöngétlen obstruens + *l* kapcsolatokat is (mivel a magyar köznyelvben is megfigyelhető bizonyos esetekben az *l* zöngésítő hatása, erre a leggyakrabban idézett példák a *hőmérséklet* [fiø:me:rfe:glet] és a *bicikli* [bitsigli]). Referenciaként a köznyelvi zöngességi hasonulás prototipikus példáit is elemezzük: zöngességükben eltérő obstruensek bármilyen sorrendű kapcsolatait (kivéve azokat az eseteket, amelyeknek a kimenete összeolvadás, pl. *egyszer* [ɛts:ɛr]). Nem vesszük számításba azokat a szóhatáron lévő mássalhangzó-kapcsolatokat, amelyek határán szünet észlelhető.

Első közelítésben a típusok gyakoriságát és a konkrét mássalhangzó-kapcsolatok realizációs jellemzőit elemezzük, különös tekintettel ezek variabilitására (az egyes szavak-szerkezetek ismétlődése lehetővé teszi egyazon fonémakapcsolat variabilitásának elemzését). A zöngesség/zöngétlenség megállapításának támpontja az oszcillogram és a spektrogram akusztikai elemzése volt, a kérdéses esetekben a Praat alapbeállításai mellett a Voice Report Voicing: Fraction of locally unvoiced frames funkcióját használtuk. Amennyiben a zöngétlen periódusok aránya meghaladta a 60%-ot, a beszédhangot zöngétlenként soroltuk be; amennyiben nem érte el a 40%-ot, a beszédhangot zöngésnek minősítettük; a két érték között pedig félzöngés (media) besorolást kapott a beszédhang.

3. Eredmények

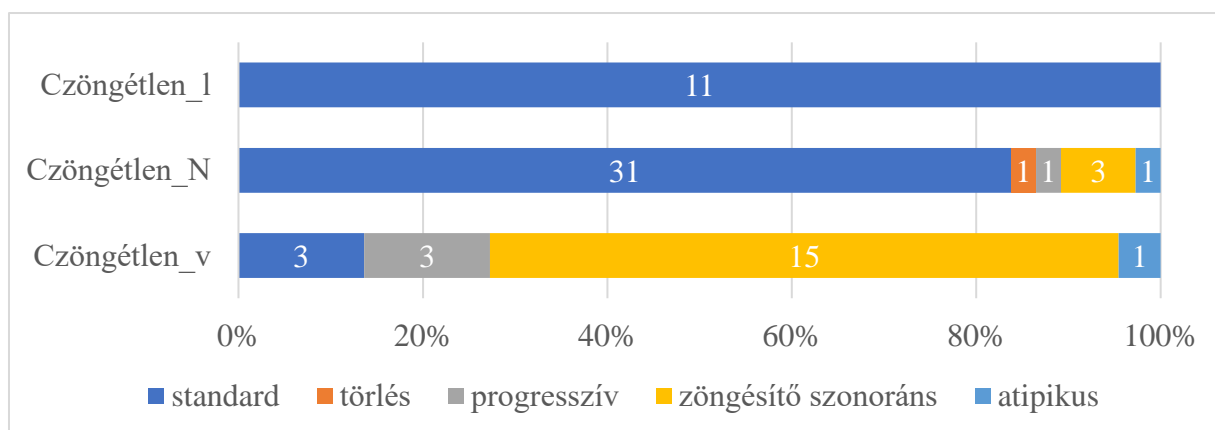
Az eddig megvizsgált 193 kételemű fonémakapcsolat közül 56 zöngétlen obstruens + zöngés obstruens; 67 zöngés obstruens + zöngétlen obstruens; 70 zöngétlen obstruens + szonoráns. A szonoránsok közül 37 nazális (22 *m* és 15 *n*); 11 *l* és 22 *v*.

Az 1. ábrán az obstruenskapcsolatok realizációtípusait látjuk, ahol megfigyelhető, hogy (a vizsgált anyagban) eltérő a standard és nemstandard realizációk aránya attól függően, hogy zöngésítésről vagy zöngétlenítésről van szó.



1. ábra A kételemű obstruenskapcsolatok realizációtípusainak megoszlása

A 2. ábrán az obstruens + szonoráns kapcsolatok realizációtípusai kapcsán látható, hogy az *l* egyetlen esetben sem zöngésítette az előtte álló obstruens, a nazális kismértékben, a *v* azonban túlnyomórészt. A *v* három esetben volt progresszív zöngétlenítés targetje. Az atipikus előfordulások között olyan példák szerepeltek, amelyekben progresszív hasonulás targetjeként a nazális vált zöngétlenné (ezt a lehetőséget nem említi szakirodalom), vagy pedig (a *v* esetében) a progresszív zöngétlenítést a trigger zöngétlen obstruens zöngéssé (sőt approximánssá) válása kísérte.



2. ábra Az obstruens + szonoráns kapcsolatok realizációtípusainak megoszlása

Az ismétlődő lexémákban és szókapcsolatokban megfigyelhető hasonulási mintázatok variabilitása – a szakirodalommal egyezően – nagymértékű. Az obstruens + szonoráns kapcsolatokban (egyelőre) nem találtunk arra példát, hogy ugyanazon szóban egyszer regresszív zöngésítés, máskor progresszív zöngétlenítés fordulna elő, de a többször többféle lexémában vagy lexémahatáron előforduló $k(\#)v$ és $t(\#)v$ kapcsolatok mindkét módon megjelentek.

4. Következtetések

Mivel a vizsgálat esettanulmány (mindössze egy beszélő hangfelvételeit volt mód elemezni), nem tudunk választ adni arra a kérdésre, hogy mennyiben általánosíthatók az eredmények, és mennyiben egyéni ejtési jellemzők. Elképzelhető, hogy a ritka (kivételes) ejtészváltozatok az adott beszélő esetében is csak egyszeri előfordulásúak voltak. Az igazságügyi hangtechnikai szakértésben ugyanakkor gyakran éppen az egyéni sajátosságok vezethetnek eredményre.

A vizsgálat folytatásának szükségességét másrésztől indokolja, hogy a magyar dialektológiai hangtani leírások alig építenek akusztikai adatokra, sőt az irodalom egy részének megállapításai nem hangfelvételeken, hanem egyszeri hallás alapján történő lejegyzéseken alapulnak.

Irodalom

- Bárth M. J. 2022. „Vazs megye, Vazsvár” – egy nyelvjárási jelenség különös története. *Édes Anyanyelvünk* XLIV/5. 9–11.
- Fehér K. 2009. Dialektális hangjelenségek és az ún. zöngességi hasonulások. In: É. Kiss K. – Hegedűs A. (szerk.): *Nyelvelmélet és dialektológia*. Piliscsaba: PPKE BTK Elméleti Nyelvészeti Tanszék – Magyar Nyelvészeti Tanszék. 85–96.
- Fodor K. 2001. A nyelvjárási hangtani jelenségek. In: Kiss J. (szerk.): *Magyar dialektológia*. Budapest: Osiris Kiadó. 325–350.
- Imre S. 1971. *A mai magyar nyelvjárások rendszere*. Budapest: Akadémiai Kiadó.
- Juhász D. 2001. A magyar nyelvjárások területi egységei. In: Kiss J. (szerk.): *Magyar dialektológia*. Budapest: Osiris Kiadó. 262–324.
- Kiss J. 2001. A nyelvjárási jelenségek. In: Kiss J. (szerk.): *Magyar dialektológia*. Budapest: Osiris Kiadó. 38–41.
- Kis T. 2016. A bilabiális β és a v -féle hangok a magyarban. *Magyar Nyelvjárások* 54. 23–50.
- Tatár Z. 2013. Beszélőprofil-alkotás lehetőségei a kriminalisztikai fonetikában. *Alkalmazott Nyelvtudomány* XIII/1–2. 121–130.
- Tatár Z. – Varga Z. – Főző E. 2021. Beszélőazonosítás a kriminalisztikában. In: Markó A. (szerk.): *Tanulmányok a beszédtudomány alkalmazásainak köréből*. Budapest: ELTE Eötvös Kiadó. 229–250.
- Vargha F. S. 2022. Vokalizmus és konzonantizmus a magyar nyelvjárások hasonlósági mintázataiban. *Általános Nyelvészeti Tanulmányok* XXXIV. 333–360.