

BESZÉDTUDOMÁNY – SPEECH SCIENCE

2022 (3)

Szerkesztők/Editors:

Grácsi, Tekla Etelka

Gyarmathy, Dorottya

Horváth, Viktória

Krepsz, Valéria

Mády, Katalin

Nyelvtudományi Kutatóközpont, Eötvös Loránd Kutatási Hálózat (ELKH)

Hungarian Research Centre for Linguistics,

Eötvös Loránd Research Network (ELKH)

Budapest

Szerkesztők/Editors:

Grácz, Tekla Etelka

Gyarmathy, Dorottya

Horváth, Viktória

Krepsz, Valéria

Mády, Katalin

Szerkesztőbizottság/Editorial board:

Bunta, Ferenc (University of Houston)

Hámori, Ágnes (Hungarian Research Centre for Linguistics)

Hoffmann, Ildikó (Hungarian Research Centre for Linguistics & University of Szeged)

Huntley-Bahr, Ruth (University of South Florida)

Markó, Alexandra (affiliacio)

Mildner, Vesna (University of Zagreb)

Olaszy, Gábor (Budapest University of Technology and Economics)

Siptár, Péter (Hungarian Research Centre for Linguistics, Eötvös Loránd University)

Sztahó, Dávid (Budapest University of Technology and Economics)

Trouvain, Jürgen (Saarland University)

White, Laurence (Newcastle University)

Technikai szerkesztés/Typesetting: Garai, Luca

Borítótér/cover design: Gyarmathy, Dorottya ©

Korrektúra/Proofreading: Jankovics, Julianna

A folyóiratszám kiadását az Eötvös Loránd Kutatási Hálózat (ELKH) támogatta.

This volume was supported by the Eötvös Loránd Research Network (ELKH).

©Nyelvtudományi Kutatóközpont, ELKH/Hungarian Research Centre for
Linguistics, ELKH

1068 Budapest, Benczúr u. 33.

Szerkesztői előszó

Kedves Kollégák!

A korábbi *Beszéd kutatás* folyóirat 2020 óta új néven jelenik meg, *Beszédtudomány – Speech Science* címmel. Célunk a beszédtudomány különböző területeiről érkező kutatások ismertetése. A jelen kötet tanulmányai foglalkoznak többek között beszédtechnológiával, a beszéd szegmentális és szupraszegmentális szerkezetének vizsgálatával, a diskurzusjelölők leírásával, a beszédhangok különböző realizációival, az olvasás közbeni hibázással, a dajkanyelvvel, illetve az idegennyelvi jellemzőkkel.

A jövőben is várjuk a tanulmányokat, például a következő területekről: artikuláció, akusztikum és percepció; beszédtechnológia, beszédfelismerés, beszéd-szintézis, kriminalisztikai kutatások és alkalmazások; fonológiai folyamatok érvényesülése a beszédben; az anyanyelv és idegen nyelvek elsajátítása; két- és többnyelvűség; prozódia, szintaxis; pragmatikai vonatkozások; klinikai kutatások, beszéd- és nyelvi zavarok; korpuszok, adatbázisok fejlesztése, diszharmóniás jelenségek a beszédben, valamint további, a beszéd jellemzőivel, feldolgozásával, létrehozásával kapcsolatos kérdésekről. A magyar vagy angol nyelvű tanulmányok terjedelme legalább 30.000 karakter (szóközökkel). A beküldés és a lektorálási folyamat az <http://ojs3.mtak.hu/> oldalon keresztül történik. Minden tanulmányt két független szakmai lektor véleményez a szerzők névtelensége mellett. Részletes információk a folyóirat honlapján található: <https://fonetika.nytud.hu/journal>. Amennyiben emailben is szeretne értesítést kapni a felhívásról, kérjük, jelezze a fonetika@nytud.hu címen!

Üdvözlettel: a kötet szerkesztői

Grácsi Tekla Etelka, Gyarmathy Dorottya, Horváth Viktória,

Krepsz Valéria, Mády Katalin

Editorial foreword

Dear Colleagues,

The former journal *Beszéd-kutatás* (*Speech Research*) was renamed in *Beszéd-tudomány – Speech Science* since 2020. One of our goals is to increase the number of international publications, as is signalled by the change of the journal's bilingual title. The present issue deals with speech technology, the study of the segmental and suprasegmental structure of speech, the description of discourse markers, the different realisations of speech sounds, error repairs in reading, motherese and foreign language characteristics, among others. Papers in all areas of speech science are welcome, such as: articulation, acoustics and perception; speech technology, speech recognition, speech synthesis, forensic research and applications; realisation/manifestation of phonological processes in speech; first and second language acquisition; bilingualism and multilingualism; prosody, syntax; pragmatic aspects; clinical research, speech and language disorders; development of corpora and databases; disharmonic phenomena in speech, and other research questions connected to speech characteristics, processing and production. The language of the submissions is English or Hungarian. The expected length of the studies is at least 30,000 characters (including spaces). Given that the journal is published only online, there is no upper limit for paper length. Papers are published based on a double blind peer-reviewing process. The submission and reviewing process takes place via <http://ojs3.mtak.hu/>. For details, see the journal's website <https://fonetika.nytud.hu/journal>. If you wish to be informed via email about the calls, please send us a note to fonetika@nytud.hu.

Sincerely, the Editors

Tekla Etelka Gráci, Dorottya Gyarmathy, Viktória Horváth,
Valéria Krepsz, Katalin Mády

Tartalomjegyzék/Table of contents

Dávid Sztahó – Attila Fejes – György Szaszák: <i>Comparing formant extraction methods according to speaking style and added noise in Forensic Voice Comparison</i>	7
Markó Alexandra – Deme Andrea – Juhász Kornélia – Bartók Márton – Csapó Tamás Gábor – Grácsi Tekla Etelka: <i>A mássalhangzókra ható prozódiai erősítés vizsgálata a magyarban</i>	36
Tar Cintia: <i>Az önjavítási jelenségek magyar nyelvű spontán, baráti társalgásokban</i>	75
Kohári Anna – Harmati-Pap Veronika – Mády Katalin: <i>A dajkanyelv tagolódása 6 hónapos csecsemőkhöz szóló történetmesélésben</i>	105
Zemán Andrea Anna: <i>A logopédiai terápia hatása az óvodások fonológiai tudatosságára</i>	138
Kis Orsolya – Steklács János – Jakab Katalin – Klivényi Péter: <i>Afáziás személyek lexikai hozzáféréseinek vizsgálata olvasás közben</i>	162
Felletár Fanni – Gosztolya Gábor – Hoffmann Ildikó – Babarczy Anna: <i>A megakadásjelenségek és a temporális paraméterek szerepe a borderline személyiségzavar felismerésében</i>	188
Blaskovics Ákos – Ittész Ambrus: <i>Ybl a Google-ön. A szillabikus l és a magyar magánhangzó-harmónia esete</i>	225
Csapó Tamás: <i>Audiovizuális beszédszintézis nyelvultrahang alapon</i>	273
Gocsál Ákos: <i>A FilmHír adatbázis fejlesztése régi filmhíradókban elhangzó beszéd vizsgálatához</i>	292

Comparing formant extraction methods according to speaking style and added noise in Forensic Voice Comparison

Dávid Sztahó¹, Attila Fejes², György Szaszák¹

¹*Budapest University of Technology and Economics, Faculty of Electrical Engineering and Informatics, Department of Telecommunication and Media Informatics*

²*Hungarian National University of Public Service*

Abstract

In forensic voice comparison, formant measurements are a “traditional” way of comparing speaker identities. Deep learning may offer a new way of estimating formant values; therefore, it is essential to compare its performance in a forensic way of use. In this study, four formant estimation methods are compared: three based on LPC and one on deep learning. Several aspects of formant modeling in forensic voice comparison were investigated: comparisons according to utterance lengths, speaking styles, and samples corrupted with various types of noise: reverberation and white noise. Results are reported according to C_{ur} , AUC, and EER metrics. It was found that the length of recordings used as suspect samples influences performance to a large extent. Additionally, formant tracking based on deep learning lags behind the other methods in all metrics. Same and different speaking styles also have a measurable effect on performance. Samples corrupted with reverberation do not deteriorate results but white noise does. There are no exact results on which method is better and which is to be used in studies and works. C_{ur} values show that the three LPC-based methods perform similarly. They all make large mistakes when samples are corrupted with white noise. Although the deep learning-based formant extractor performs slightly worse than the other approaches used in this study, it seems to have more resilience to white noise.

1. Introduction

The paradigm shift in forensic sciences and practice (; Saks & Koehler, 2005; Morrison, 2009a, 2011b) enabled the automatic and semi-automatic evaluation of evidence using various modalities and kinds of measurements (such as DNA, fingerprint) (Bazen & Veldhuis, 2004; Matz & Nielsen, 2005). This new paradigm, called the likelihood-ratio (LR) framework, is feasible for processing

Email addresses: `sztaho.david@vik.bme.hu` (Dávid Sztahó),
`fejes.attila@nbsz.gov.hu` (Attila Fejes), `szaszak.gyorgy@vik.bme.hu` (György Szaszák)

by computers. Such is the case in voice processing, where speaker recognition techniques are adapted to the requirements of the framework, namely to produce a probability ratio of same and different speaker for evidence (Mandasari et al., 2011; Morrison, 2011a; Kelly et al., 2019).

There are numerous ‘traditional’ acoustic features used in forensic voice comparison systems, such as fundamental frequency, mel-frequency cepstral coefficients, and intonation-based features (Chaudhary et al., 2017).

Formants, local maxima of the speech spectrum caused by resonance frequencies of the vocal tract, are also traditional features of forensic voice comparison (Titze & Martin, 1998). Although speech sounds, especially vowels, are characterized by their first two formants, formants are also speaker-dependent. Due to the possible overlap of the first formant and higher fundamental frequency, second and third formants are more frequently used in research works (Gargouri et al., 2006; Guillemin & Watson, 2008) although higher formants may also have problematic calculations (attenuation, superimposed noise, exposure to GSM coding).

There are numerous formant tracking applications and methods that are feasible for calculating formant trajectories of voice samples for forensic purposes. Commonly used methods (Kameny et al., 1974; Snell & Milinazzo, 1993) are based on LPC coefficients and direct spectrum envelope estimation, but novel techniques are also emerging that use deep learning for the same purpose (Zhang et al., 2013; Dissen et al., 2019). Although, it is an open question whether deep learning techniques have superficial performance in forensic speaker verification and also in general formant extraction compared to traditional LPC-based methods. Due to the data-driven basis of deep learning, it may be more robust against different noises speech signals are corrupted with (such as reverberation or white noise) if the training data is prepared to contain these corruptions for each target class. Otherwise, it may be fooled by background noise (Ribeiro et al., 2016). *Deepformants* applies a standard feedforward network architecture to learn LPC parameters, thus, in theory, removing potentially highly inaccurate and redundant estimates (Dissen & Keshet, 2021). The goal of the present

paper is to examine the features derived from these formant extraction methods in realistic forensic scenarios by applying a dataset that is created by forensic protocol requirements. Therefore, the used dataset must meet certain criteria.

From an engineering point of view, various modeling techniques can be adapted to the LR framework. State-of-the-art speaker verification techniques (i-vector, x-vector) (Dehak et al., 2010; Mandasari et al., 2011; Snyder et al., 2018; Kelly et al., 2019) produce speaker embeddings based on the spectrum of the voice and require large amounts of training samples. On the other hand, datasets developed directly for forensic purposes commonly contain limited samples. These corpora must meet certain special requirements (Morrison et al., 2012), such as multiple recordings of individuals separated in time, and multiple speaking styles from each speaker. Due to these highly controlled scenarios, a limited number of speakers can be recorded. Therefore, more simple modeling techniques are commonly used, among these are the multivariate kernel density estimation (MVKD) and Gaussian mixture models (GMM) that are favorably utilized in forensic voice comparison (Becker et al., 2008; Rose & Winter, 2010; Morrison, 2011a; Wang & Zhang, 2015; Hughes, 2017; Tsuge & Ishihara, 2018). These probability-based methods are easily fitted into the LR framework. Following the common practice of forensic voice comparison, we use the MVKD and GMM modeling techniques to compare formants estimated by the investigated formant trackers.

In this paper, we aim to investigate multiple matters on a Hungarian corpus specially created for forensic voice comparison. First, we introduce the Hungarian Database for Forensic Voice Comparison and evaluate the performance of MVKD and GMM on the corpus using formants as a first result obtained on the dataset. We would like to extend the existing works measuring the performance of these methods by applying them to a corpus created for forensic purposes. Coy and his colleagues (2021) compared *deepformants* to formants calculated by *Snack* and found controversial results. Formants extracted by *deepformants* were not found to be superior in all investigated scenarios. However, the ap-

plied corpus did not fulfill the requirements of forensic practices: recordings from multiple sessions from each speaker, and varying speaking styles.

Speaking styles in speaker verification were examined in some studies before, but they used artificial speaking style modifications for their experiments (Afshan et al., 2020) or the speakers were forced to change vocal efforts and speaking styles (Shriberg et al., 2008). Also, deep learning-based formant tracking was tested against noises added to the speech signal (Gowda et al., 2021). Our goal in this paper is not to examine these effects separately, but to measure the formant trackers’ performance in a realistic forensic setup that our applied corpus enables. Utterance length is also an essential factor in forensic speaker comparison. There are studies that deal with these phenomena (such as Honglin & Jiangping, 2012), but deep learning-based formant trackers were not yet investigated.

Also, there are gaps in such works in Hungarian corpora, the filling of which would prove very useful. We investigate the effect of recording length on performance. Also, we would like to measure the performance of such methods when different speaking styles are available for offender (sample of evidence from crime scene) and suspect (test sample with speaker identity in question) voice samples. Finally, we examine the effect of noise on the applied formant calculation methods by adding reverberation effects and white noises of two signal-to-noise levels to the clean samples.

In the following chapters, we introduce the corpus, the formant tracker methods, modeling techniques, and evaluation metrics applied. Next, the experimental setup is described, followed by the achieved results. In the end, a detailed discussion is given with the conclusion of the work.

2. Methods

2.1. Hungarian Database for Forensic Voice Comparison

We introduce for the first time the FORVOICE audio database for Forensic Voice Comparison in the Hungarian language. To this day, this database con-

tains samples of 80 speakers, 39 females and 41 males (between the ages of 18 and 35). Net total speaking time (without silences and pauses) is 24.74 hours. Four speaking styles are recorded: free dialogue, controlled information exchange, monologue, and prescribed answers to questions (simulation of refused answers at interrogation). Two recordings were made per speaker, separated by a period of at least two weeks. These are marked as sessions 1 and 2 in the rest of the paper. The average duration of speaking durations per speaker and session is 266.66, 153.71, 124.39, and 12 seconds for each speaking style in respective order as mentioned before. Prior to recording, subjects expressed written consent to record their voice for the given research purposes. Head-mounted microphones were used to ensure the best possible recording quality (format: PCM, 44kHz, 16-bit). Manual transcriptions and phoneme and word level segmentations are available for all recordings.

2.2. Formant trackers and features

Four formant tracking methods were compared, three LPC based and one deep learning based:

- *Praat* (Boersma & Weenink, 2021) with ‘Burg’ method,
- *Snack Sound Toolkit* (Kåre, 2021),
- *Voicebox* (Brookes, 2021) (a *Matlab* based toolbox), and
- *Deepformants* (Dissen & Keshet, 2021 trained on the training set of VTR-TIMIT (Dissen et al., 2019). Based on the formant space examination in Van Heuven, 2016, it may not be a problem that the model is trained on a different language than it is used on.

25 ms as the window size and 10 ms as the time step were applied. Phoneme level segmentation (generated by Hidden Markov Model Toolkit (HTK) [Young & Young, 1993] using forced alignment based on the manual transcriptions) was available for the whole corpus. Formants (F) were measured for vowels /i/, /u/,

and /a:/, the ones with the farthest in the F1 and F2 space (nodes of a triangle) in Hungarian. For each vowel, the mean, the standard deviation, and the first three discrete cosine transform (DCT) coefficients (the 0th coefficient was skipped) of the first three formants were calculated, resulting in a 15-dimension feature vector for each vowel.

Figure 1 shows the formant space (in the space of F1 and F2 derived from /i/, /u/, and /a:/ vowels) measured in samples without noise abruption (*clean* samples). The formant space was modeled by GMM probability density functions. The figure is not derived from the extracted features but it is only a sample of the deviation between the formant tracking methods. As the figure clearly shows, neither calculation framework gives the same (or completely accurate) formant space even in the case of clean samples. With added noise, even larger differences are found. Figures 2 and 3 show the formant space of *clean* and *white noise with 10 dB SNR* in the case of *Praat* and *deepformants*. The method *deepformants* seems to have calculated formants more robustly, but even in this case, a large deviation can be observed between the two acoustic scenarios. This may affect speaker verification results.

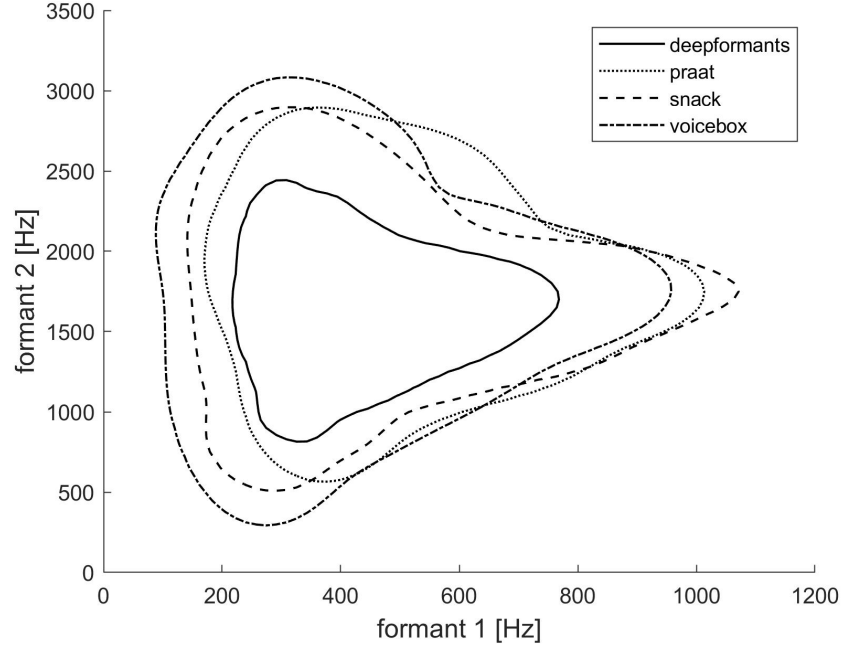


Figure 1: Formant space (F1-F2 triangle derived from /i/, /u/, and /a:/ vowels) measured at *clean* samples.

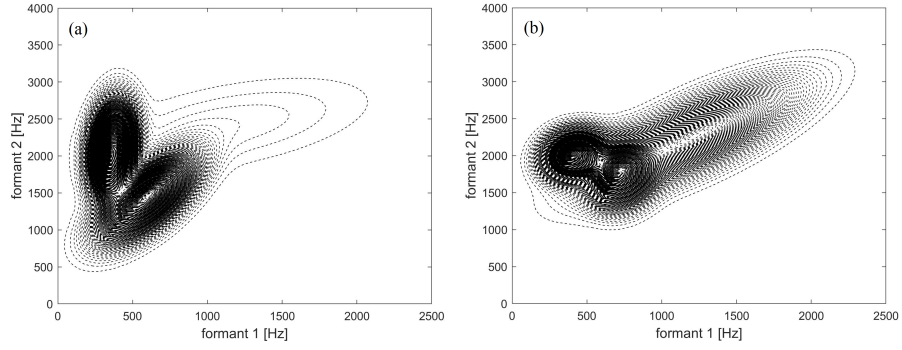


Figure 2: Formant space measured at *clean* (top) and *white noise with 10 dB SNR* (bottom) samples with *Praat*. The GMM-modeled PDFs are shown.

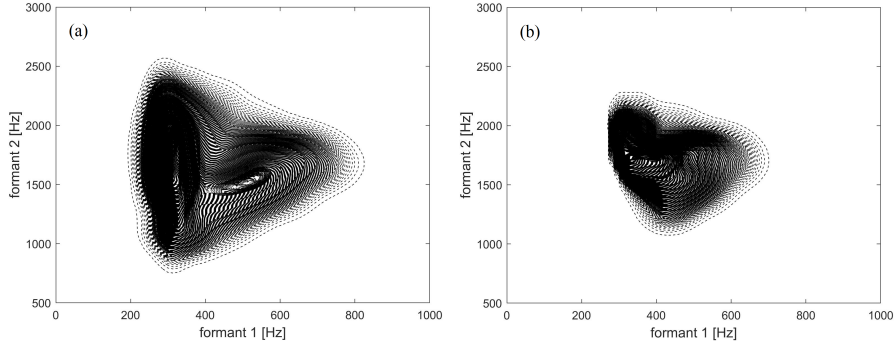


Figure 3: Formant space measured at *clean* (top) and *white noise with 10 dB SNR* (bottom) samples with *deepformants*. The GMM-modeled PDFs are shown.

2.3. LR framework

In a speaker verification method, we have to investigate two hypotheses: (1) "What is the possibility that the sample in question originates from the suspect?" and (2) "What is the possibility that the sample in question originates from a randomly selected speaker of a background population?" The ratio of these expressions expresses the strength of the evidence (Eq. 1). LR is the likelihood ratio, E is the evidence, H_{so} is the hypothesis of same-origin speakers, and H_{do} is the hypothesis of different-origin speakers.

$$LR = \frac{p(E|H_{so})}{p(E|H_{do})} \quad (1)$$

There exist several procedures to calculate LR value from univariate and multivariate continuous data. In this study, we use multivariate kernel density (MVKD) (Aitken & Lucy, 2004) and Gaussian mixture models (GMM) to estimate the density functions of features described in Section 2.2.

2.3.1. Multivariate Kernel Density

Kernel density estimation is a nonparametric technique for density estimation. It can be viewed as a generalization of histogram density estimation with improved statistical properties. In MVKD, the density function of the samples is estimated by using kernels centered at each data point. The estimated density

function is obtained by summing these kernels. This eliminates the problem of choosing the correct bins for histograms.

MVKD uses the summation of a set of equally-weighted kernels with one kernel per group centered on the mean vector of the measurements from that group for modeling between-group distribution. In forensic voice comparison, the speaker is such a group. The Gaussian kernel is used with a scaled covariance matrix of the pooled within-group covariance matrix. The degree of kernel smoothing (scaling) is determined by a function of the number of background database groups (Morrison, 2011a). The procedure is described in detail in Aitken & Lucy, 2004 and Morrison, 2011a. For detailed information and equations about calculating the LR score of MVKD, see Morrison, 2011a.

2.3.2. GMM-UBM

Gaussian mixture models (Reynolds & Rose, 1995; Hansen & Hasan, 2015) is a combination of Gaussian probability density functions (PDFs) that are commonly used to model multivariate data. It does not only cluster data in an unsupervised way, but also gives its PDF. Applying GMM to speaker modeling provides the speaker-specific PDF, from which a probability score can be obtained. Thus, by testing a sample with an unknown label, based on the probability scores of the speaker GMMs, a decision can be made.

A GMM is a mixture of Gaussian PDFs parameterized by a number of mean vectors, covariance matrices, and weights (Eq. 2). π_g , μ_g , and Σ_g indicate the weight, mean vector, and covariance matrix of the g^{th} mixture component. For a sequence of acoustic features ($X = x_n | n \in 1 \cdots T$), the probability of observing these features is computed as Eq. 3.

$$f(x_n|\lambda) = \sum_{g=1}^M \pi_g N(x_n|\mu_g, \Sigma_g) \quad (2)$$

$$p(X|\lambda) = \prod_{n=1}^T p(x_n|\lambda) \quad (3)$$

For the speaker verification scheme, a slightly different approach was developed in (Hansen & Hasan, 2015). Besides the claimed speaker’s model, an alternate

model is necessary, which represents an 'opposing' model. This alternate model is called the universal background model (GMM-UBM). The GMM-UBM represents a background speaker population and it is trained on a large number of speaker samples. This is used to answer the question if the given test sample is more likely to be sampled from a target speaker or not and helps to calculate the LR score. It was first used in Reynolds et al. (2000). Later, UBM was used as an initial model for the speaker models: rather than training GMMs on speaker data directly, the specific speaker models were created by adapting a prior UBM (Gauvain & Lee, 1994). In the GMM-UBM scheme, H_{so} and H_{do} are represented by speaker-dependent GMM and the GMM-UBM, respectively.

2.3.3. Fusing calibration

Because formants were measured at multiple vowels, it is necessary to use the information obtained from them jointly. The evaluated final results are generated by fusing scores of individual vowel LR scores. Logistic regression (Hastie et al., 2009) is a probabilistic classification method, offering a common score-to-likelihood-ratio transformation, and it is feasible to calibrate a single set of scores and fuse multiple sets of scores (Brummer et al., 2007). It takes the same and different speaker labels as target labels and the extracted features as input, then it fits a logistic curve to the data, which can be interpreted as the probability of each class (same and different speaker identities). The logistic regression-based calibration and fusion need multiple scores of comparisons from same and different speaker sample pairs combined into training data. Tokens of the three applied vowels are used as acoustic-phonetic comparisons as parallel comparisons on the same speech sample required by the algorithm. Calibration is an affine transformation to a set of scores optimizing a cost function. In forensic voice comparison, this cost function to be minimized is the log-likelihood-ratio cost (C_{llr}) (Van Leeuwen & Brummer, 2007) (Eq. 4),

$$C_{llr} = \frac{1}{2} \left(\frac{1}{N_{so}} \sum_{i=1}^{N_{so}} \log_2 \left(1 + \frac{1}{LR_{so_i}} \right) + \frac{1}{N_{do}} \sum_{j=1}^{N_{do}} \log_2 (1 + LR_{do_j}) \right), \quad (4)$$

where N_{so} and N_{do} are the number of same-origin and different-origin comparisons and LR_{so} and LR_{do} are the likelihood ratios derived from same-origin and different-origin comparisons. C_{lr} is a function measuring the balance of LR scores of same and different speaker comparisons. Ideal same-origin and different-origin comparison have $\log(LR_{so}) > 0$ and $\log(LR_{do}) < 0$, respectively. Incorrect (not as ideal as the mentioned inequalities) produce a higher C_{lr} . The better the performance of a forensic comparison system, the more correct LR values are produced, the lower C_{lr} is achieved, supplying the evidence magnitude. Thus, calibrated C_{lr} on known same-origin and different-origin sample pairs provides a metric of system validity. Calibration in this study followed Morrison (2011a) and was based on the UBM set. Same and different speaker pairs were built from all speaker pairs, skipping mirrored pairs due to symmetrical LR values (this is not exactly precise, but they are similar enough to be skipped). Throughout the paper, fused C_{lr} is the base of all evaluation metrics. C_{lr} is commonly drawn as tippet plots, showing LR components in a clean way (Morrison, 2009b).

2.4. The experimental setup

Two speaking styles were used in the study:

- **free dialogue (fd):** completely free dialogue with a talking partner without restrictions (~ 10 minutes). Due to individual head-mounted microphones, crosstalk volume was minimal between speakers;
- **monologue (m):** speakers should tell the events of their previous day objectively (~ 3 minutes per speaker).

The original clean recordings were also augmented with two types of distortions: reverberation and white noise. Reverb effects were applied by *Matlab* (Dattorro, 1997) with the following *Matlab* function arguments: density of reverb tail – 0.5 (scale: 0–1), decay factor of reverb tail – 0.5 (scale: 0–1), ratio of reverberated to original signal – 0.3 (scale: 0–1). The values were chosen by *Matlab* recommendations. White noise was added to clean recordings to achieve

two mean signal-to-noise levels: 10 dB and 15 dB. SNR was calculated as the sound intensity level ratio measured between the original clean recording and the added noise. This augmentation resulted in four final sets: clean, reverb, white noise with 10 dB SNR (*'wnoise10'*) and white noise with 15 dB SNR (*'wnoise15'*).

Multiple scenarios were considered to be evaluated: same speaking style and cross-speaking style experiments were carried out. In the same speaking style setup, suspect and offender (evidence from crime scene) samples are both taken from the *free dialogue* and *monologue* tasks and are evaluated. In the cross-speaking style setup, suspect and offender samples were taken from different tasks. In the last case, all speaking styles were used. From the 80 speakers, 60 were used to train the universal background model, the remaining 20 speakers were selected for suspect and offender sets. Sex distribution of the UBM matches the suspect-offender speakers. From the 20 speakers' samples, session 1 was used as the offender, while session 2 was applied as the suspect. Background models were trained on both speaking styles and sessions of the selected 60 speakers. The resulting scenarios for evaluation are summarized in Table 1. Formant measurements for UBM and offender models were always taken from the total sample duration. As suspect data, various durations were considered and evaluated. Sample chunks with lengths of 20, 40, 60, $\dots$, 300 seconds were applied (15 possible chunk durations). Figure 4 shows the number of vowel tokens in the function of sample chunk durations. It shows how the median and standard deviation of token numbers of speakers change as the function of sample duration increases. We would expect that the more tokens we have to extract features from, the more robust the speaker verification is. For the sake of data uniformity, if the total duration of a given sample was exceeded by the chunk length, the vowel tokens were selected from the total sample leaving the original chunk length as notation. For example, for a 120 seconds long speech sample, the 160 seconds long chunk length used the total speech sample but was still marked as '160'. Through initial experiments, the mixture number of the

Table 1: Summary of evaluation scenarios according to speaking styles

test case	UBM	#speakers	session	speaking style
1.	offender	60	1 and 2	free dialogue and monologue
	suspect	20	1	free dialogue
	suspect	20	2	free dialogue
2.	UBM	60	1 and 2	free dialogue and monologue
	offender	20	1	monologue
	suspect	20	2	monologue
3.	UBM	60	1 and 2	free dialogue and monologue
	offender	20	1	free dialogue
	suspect	20	2	monologue
4.	UBM	60	1 and 2	free dialogue and monologue
	offender	20	1	monologue
	suspect	20	2	free dialogue

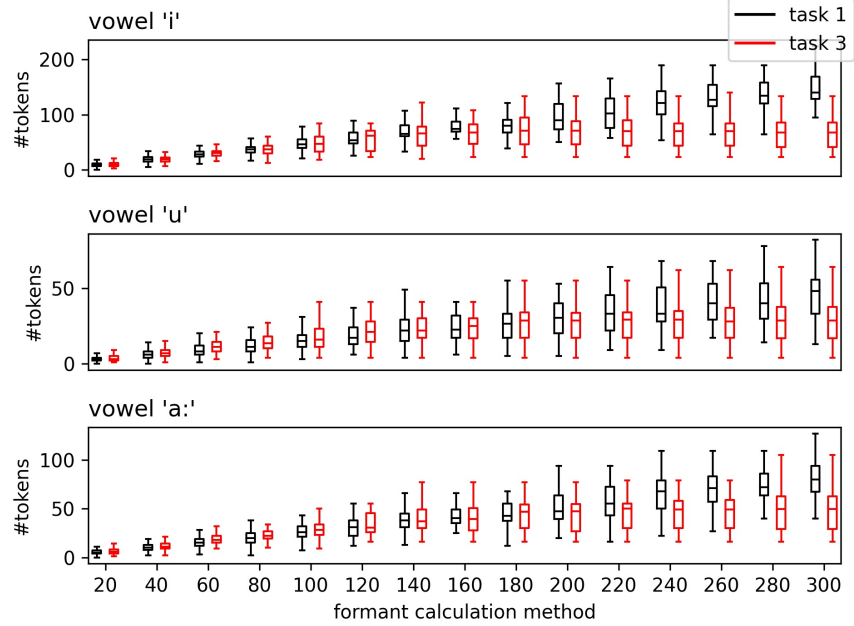


Figure 4: Number of vowel tokens according to chunk durations.

GMM model was selected as 11 and all features were selected to be included in the final feature vector.

2.5. Evaluation metrics

Multiple evaluation metrics are used to assess the performance of the test scenarios:

- Post-calibration fused C_{ur} value as described earlier (lower score means better performance).
- Equal error rate (EER). EER is the level where the false acceptance rate and the false rejection rate are equal, commonly used in biometric security systems. Lower EER means better performance.
- Receiver operating characteristics (ROC) curve. ROC can measure the performance of a binary classifier system as its discrimination threshold is varied. In our case, by discrimination threshold, we mean the level of LR score of accepting a speaker pair as same. The area under the ROC curve (AUC) value, when using normalized units, is equal to the probability that a classifier will rank a randomly chosen same speaker pair higher than a randomly chosen different origin speaker pair. A higher AUC value means better performance.

One-way ANOVA tests were used for checking mean value equivalence for investigated groups. To measure speaking style difference significance, generalized linear mixed models were applied (Hedeker, 2005).

3. Results

3.1. Chunk lengths

The effect of chunk lengths on performance was evaluated. Line plots are used for visualization to emphasize the possible trends of performance values. Figure 5 shows the C_{ur} values measured at *clean* samples as a function of chunk lengths. The mean performance of each formant calculation method is shown

separately. Due to the four experimental setups, a standard deviation range can also be depicted. Starting from 20 s, a continuously decreasing trend can be observed when applying both GMM and MVKD modeling techniques. Although MVKD resulted in a lower C_{lr} value at the beginning, GMM seems to outperform when enough tokens are available from all vowels. The EER and AUC values develop in the same way (Figures 6 and 7) as one would expect looking at the C_{lr} values. EERs fall (and AUC values rise) sharply until the 100 s chunk length and continue to decrease (AUC: increase) in a slighter way afterwards. Considering the various formant calculation methods, *praat* and *snack* had the lowest C_{lr} . The single method based on deep learning seems to lag behind the others, having an increased C_{lr} throughout the chunk lengths. ANOVA tests show no significant differences (p -values are 0.622 and 0.113 for GMM and MVKD, respectively) at 20 s chunk lengths between formant measurement methods. In the case of 300 s, there is a significant difference in mean values of C_{lr} in the case of GMM (p -values of ANOVA tests are 0.026 and 0.066 for GMM and MVKD, respectively).

3.2. Same and different speaking styles

Speaking style can have a significant effect on voice comparison by an LR framework (Drygajlo et al., 2015). Among the four experimental setups, two used the same speaking style for suspect and offender data (#1 and #2), and likewise, two used different ones (#3 and #4). For the experiments, the total lengths of the samples are used. All results obtained in these experimental setups are included in Table 2. By depicting the single C_{lr} values of the same and different speaking styles as a scatter plot (Figure 8), it is clear that by applying the same speaking style as enrollment and target samples, lower C_{lr} can be achieved. Same and different speaking style C_{lr} measurements are noted by ‘+’ and ‘x’, respectively. The C_{lr} values depicted are single values for a single experiment scenario. The same effect of noise corruption can also be observed: better and indistinguishable results of clean and reverberated samples, and almost identically worse results for white noise added. Generalized linear mixed

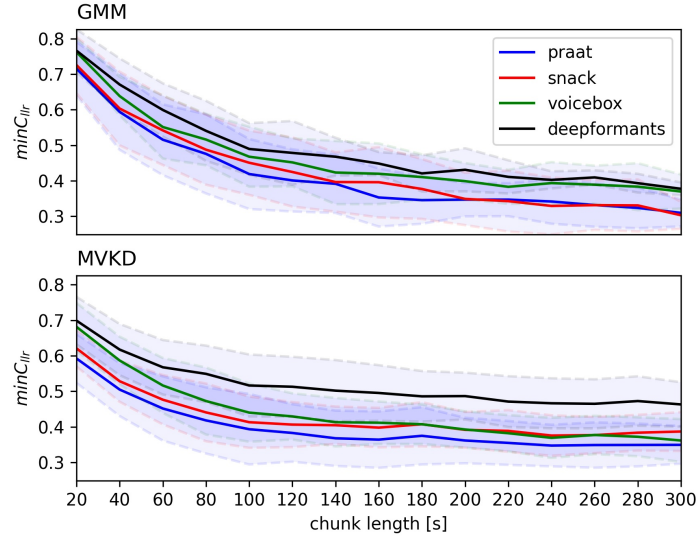


Figure 5: C_{lr} values according to split lengths using *clean* samples (top: GMM, bottom: MVKD).

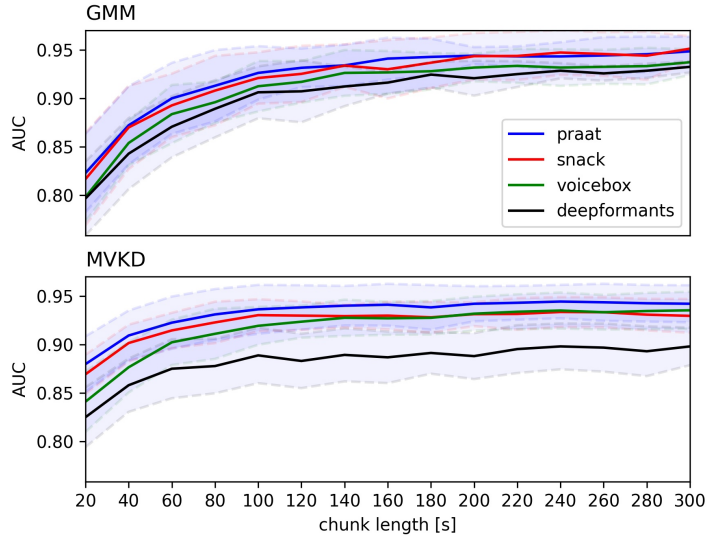


Figure 6: AUC values according to split lengths using *clean* samples (top: GMM, bottom: MVKD).

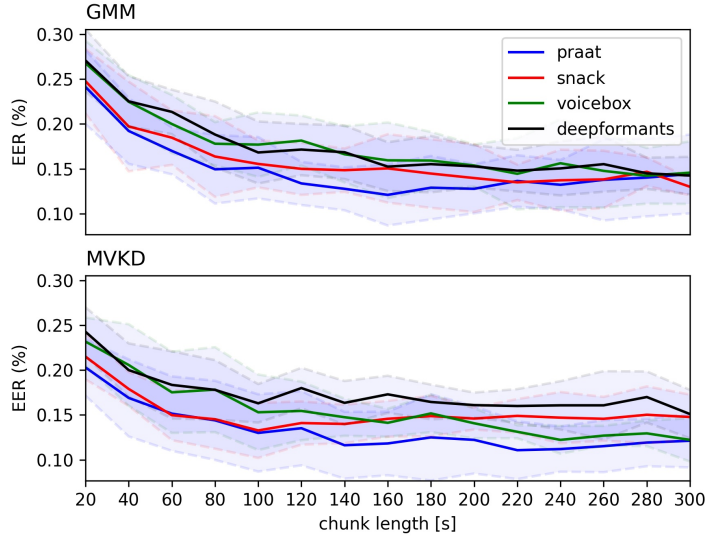


Figure 7: Equal error rates according to split lengths using *clean* samples (top: GMM, bottom: MVKD).

models were fitted to the data with C_{llr} as the target variable and speaking style (same or different), noise, and method as fixed variables. In the case of both GMM and MVKD, speaking style and noise were significant factors ($p < 0.05$) and the method was not ($p > 0.05$).

By calculating the difference of the mean of the C_{llr} values for same and different speaker styles (Figure 9), only positive values are present, showing that different speaking styles always deteriorate performance.

3.3. Noise types

Noises applied to the clean samples have different effects. Reverberation did not decrease the performance of the LR framework. On the contrary, there are multiple cases when the reverb effect did lower the C_{llr} value. Figure 10 shows mean and standard deviations (as error bars) of C_{llr} for every formant calculation method and sample quality. Only the results of 300 s long chunks are depicted because all chunk durations showed the same behavior in evaluation.

Table 2: Results aggregated according to different or same speaking styles in suspect and offender samples.

method	noise	speaking style	AUC		$\min C_{ulr}$		eer	
			GMM	MVKD	GMM	MVKD	GMM	MVKD
deepformants	clean	different	0.930	0.885	0.385	0.505	0.153	0.164
		same	0.940	0.911	0.371	0.422	0.132	0.139
	reverb	different	0.888	0.899	0.475	0.492	0.188	0.179
		same	0.949	0.926	0.310	0.378	0.125	0.140
	wnoise10	different	0.871	0.871	0.550	0.544	0.236	0.224
		same	0.908	0.950	0.449	0.316	0.157	0.108
	wnoise15	different	0.911	0.885	0.459	0.523	0.205	0.214
		same	0.909	0.951	0.458	0.324	0.165	0.141
praat	clean	different	0.938	0.929	0.332	0.384	0.153	0.132
		same	0.959	0.956	0.288	0.316	0.136	0.111
	reverb	different	0.929	0.948	0.361	0.341	0.159	0.132
		same	0.959	0.964	0.273	0.282	0.139	0.135
	wnoise10	different	0.835	0.820	0.622	0.647	0.238	0.237
		same	0.859	0.911	0.550	0.447	0.191	0.164
	wnoise15	different	0.844	0.849	0.604	0.590	0.252	0.209
		same	0.909	0.916	0.451	0.441	0.154	0.179
snack	clean	different	0.944	0.915	0.325	0.434	0.135	0.164
		same	0.959	0.945	0.282	0.340	0.125	0.132
	reverb	different	0.924	0.930	0.400	0.397	0.135	0.180
		same	0.955	0.950	0.284	0.328	0.108	0.109
	wnoise10	different	0.846	0.873	0.583	0.540	0.208	0.203
		same	0.883	0.937	0.527	0.378	0.211	0.160
	wnoise15	different	0.850	0.888	0.595	0.500	0.213	0.193
		same	0.914	0.940	0.418	0.368	0.188	0.138
voicebox	clean	different	0.933	0.925	0.394	0.394	0.153	0.135
		same	0.942	0.946	0.346	0.330	0.139	0.110
	reverb	different	0.938	0.934	0.357	0.377	0.147	0.142
		same	0.946	0.957	0.342	0.296	0.112	0.110
	wnoise10	different	0.833	0.826	0.610	0.657	0.231	0.249
		same	0.850	0.899	0.602	0.464	0.224	0.179
	wnoise15	different	0.830	0.811	0.604	0.679	0.256	0.274
		same	0.900	0.886	0.492	0.515	0.181	0.210

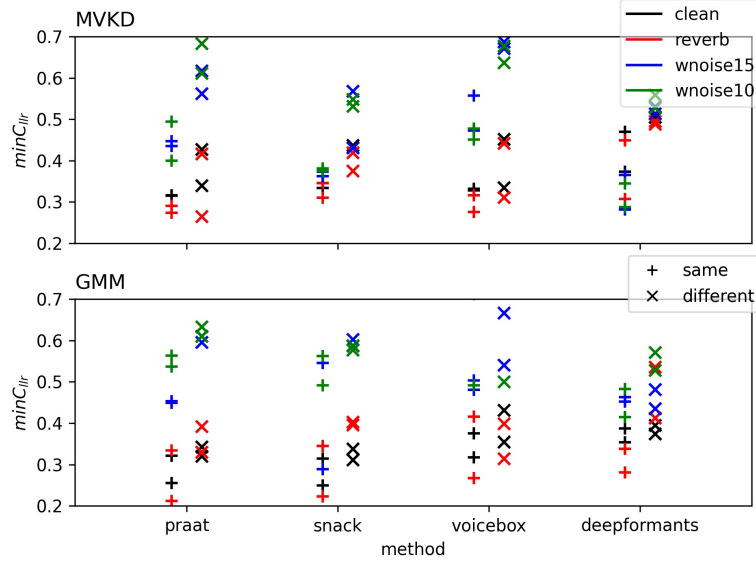


Figure 8: Scatter plot of C_{lr} values according to speaking styles (top: MVKD, bottom: GMM).

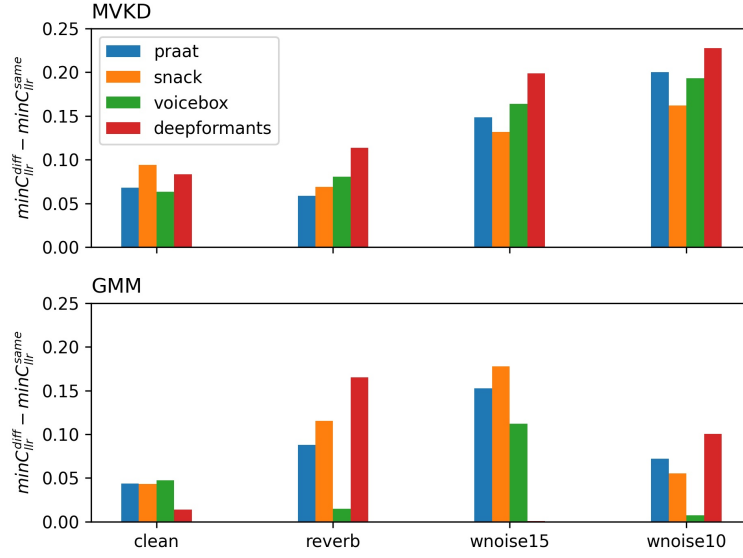


Figure 9: Difference of C_{lr} values according to speaking styles ($C_{lr}^{diff} - C_{lr}^{same}$). Positive values mean lower C_{lr} scores in same speaking style cases (top: MVKD, bottom: GMM).

However, various observations can be made by observing the measured values. As expected, white noise added to the samples has a higher impact on performance. LPC-based calculation methods exhibit a high degree of deterioration in all evaluation metrics. The method *deepformants*, however, seems to have a more robust resistance to white noise. In the case of MVKD, the C_{lr} did not decrease at all, and in the case of GMM, *deepformants* had the lowest mean values across white noise corrupted samples. There are no differences between the 10 dB and 15 dB SNR samples. The p -values of one-way ANOVA tests are shown in Table 3 marking if there are any significant differences between sample qualities and formant calculation methods. Besides *deepformants*, all other methods show significant differences across noise types. On the other hand, if we consider noise types, neither case reaches a significance level to show that the methods would differ for the given noise type.

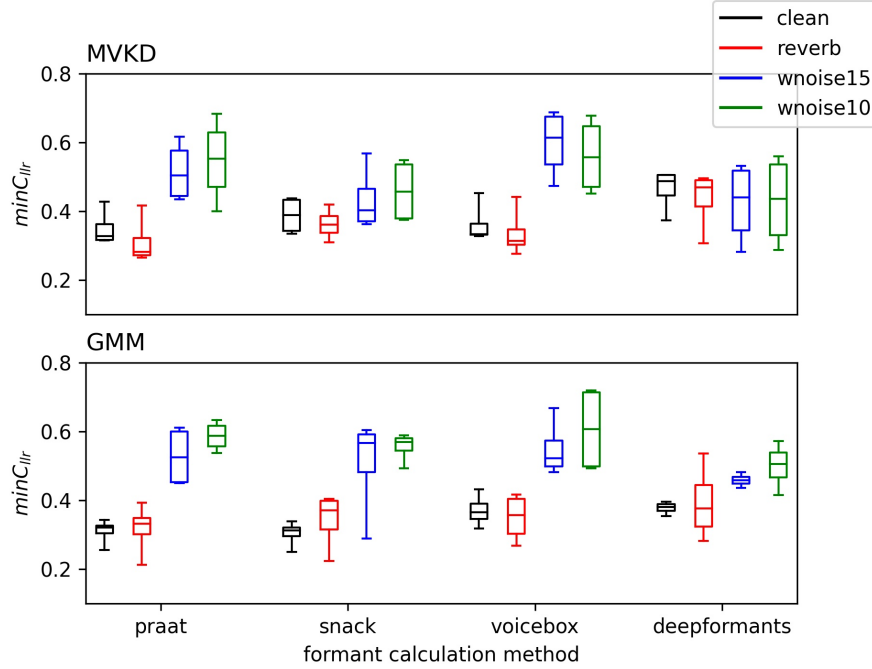


Figure 10: C_{lr} values according to formant calculation methods and noises (chunk size: 300s). Top: MVKD, bottom: GMM.

Table 3: The p -values of one-way ANOVA tests in case of 300 ms chunk lengths for noises across methods and methods across noises. Values < 0.05 are marked with ‘*’.

		GMM	MVKD
method	praat	0.0001*	0.0054*
	snack	0.0042*	0.3079
	voicebox	0.0025*	0.0021*
	deepformants	0.0720	0.9501
noise	clean	0.0257	0.0656
	reverb	0.6741	0.1329
	wnoise15	0.5985	0.1053
	wnoise10	0.2848	0.3498

4. Discussion

In this study, four formant trackers were systematically evaluated for forensic voice comparison. We analyzed the performance of various formant trackers in a forensic voice comparison setup depending on speaking style, length of suspended recording, and noises. Based on our results, the deep learning approach (*deepformants*) showed significant differences. However, the other methods, although all are based on LPC, also did not produce the same results (see Figure 1 for initial formant measurement example). On the contrary, significant differences could be observed. The method *deepformants* shows a narrower F1–F2 space than the other methods, explained by the real ‘tracking’ nature of its algorithm. While the LPC-based ones are estimating formants in each frame independently, *deepformants* takes neighboring frames into consideration when estimating formants in a given frame. This results in a narrower F1–F2 space but its advantage is its more robust nature. As Figure 8 shows, there are fewer differences between samples with different noise corruption in its case (although it performs basically worse also in the case of *clean* samples).

Our results confirm that the longer speech segments are used (more formant tokens are involved), the better the performance can be measured. Below 120 – 140 seconds, there is a rapid decline in C_{lir} . If we look at EERs, a value of

around 0.15 can be achieved from 60 – 80 seconds long chunks. This implies that formant estimates are rather unreliable if they are based on short segments.

Speaking style had a significant effect on performance measurements. When the offender and suspect speech samples originated from the same speaking style (free dialog and monologue in this study), lower C_{ur} , EER and higher AUC could be calculated compared to cases when the offender and suspect samples contained different speaking styles. Figure 10 shows that when the mean of the C_{ur} of different speaking style test trials was subtracted from same speaking style mean values, no negative values could be calculated. This implies that when comparing formants in a forensic voice comparison situation, speaking style matters a lot. It is always advised to compare the speech material of suspects with evidence of the same style.

Reverberation didn't seem to influence formant estimation in a bad way. In some cases, even lower C_{ur} and EER could be calculated compared to *clean* samples. White noise, however, mostly deteriorated the performance to a large extent. Besides *deepformants*, which seemed to be robust against white noise, but performed worse even with *clean* samples, only *snack* was robust against *white noise* with 15 dB SNR. *White noise* with 10 dB SNR always deteriorated performance, only *deepformants* was resistant to it using MVKD modeling.

There are no exact results on which method is better and which is to be used in studies and works. C_{ur} values of Figure 8 shows that the three LPC based methods perform similarly. They all make large mistakes when samples are corrupted with white noise. Although the method *deepformants* performs slightly worse than the others used in this study, it seems to have more resilience to white noise. This may be due to its real formant tracking approach, as mentioned before. This implies that it cannot be stated that the DNN approach outperforms the LPC-based ones.

5. Conclusion

In this study, several aspects of formant modeling in forensic voice comparison were investigated. A corpus containing 80 speakers with multiple speaking styles is used to estimate first and second formant values by four different formant estimation methods (three based on LPC, one on deep learning). Formants were modeled by multivariate kernel density estimation and Gaussian mixture models. It was found that the length of recording used as suspect samples influences performance to a large extent. Additionally, formant tracking based on deep learning lags behind the other methods in all metrics. Same and different speaking styles also have a measurable effect on performance. Samples corrupted with reverberation do not deteriorate results but white noise does. The continuation of the work will be to investigate these effects also by fully automatic voice comparison systems, such as x-vector and i-vectors.

Acknowledgements

The work was funded by project no. FK128615, which has been implemented with the support provided from the National Research, Development, and Innovation Fund of Hungary, financed under the FK_18 funding scheme.

References

- Afshan, A., Guo, J., Park, S. J., Ravi, V., McCree, A., & Alwan, A. (2020). *Variable frame rate-based data augmentation to handle speaking-style variability for automatic speaker verification*. arXiv (Cs, Eess). doi:<https://doi.org/10.48550/arXiv.2008.03616>. arXiv:2008.03616.
- Aitken, C. G., & Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53, 109–122. doi:<https://doi.org/10.1046/j.0035-9254.2003.05271.x>.

- Bazen, A. M., & Veldhuis, R. N. (2004). Likelihood-ratio-based biometric verification. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*, 86–94. doi:<https://doi.org/10.1109/TCSVT.2003.818356>.
- Becker, T., Jessen, M., & Grigoras, C. (2008). Forensic speaker verification using formant features and Gaussian mixture models. In *Ninth Annual Conference of the International Speech Communication Association*.
- Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer*. [Computer program] (Version 6.1.42). Retrieved 15 April 2021. URL: <http://www.praat.org/>.
- Brookes, M. (2021). *Voicebox*. URL: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.
- Brummer, N., Burget, L., Cernocky, J., Glembek, O., Grezl, F., Karafiat, M., Van Leeuwen, D. A., Matejka, P., Schwarz, P., & Strasheim, A. (2007). Fusion of heterogeneous speaker recognition systems in the stbu submission for the nist speaker recognition evaluation 2006. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*, 2072–2084. doi:<https://doi.org/10.1109/TASL.2007.902870>.
- Chaudhary, G., Srivastava, S., & Bhardwaj, S. (2017). Feature extraction methods for speaker recognition: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, *31*, 1750041. doi:<https://doi.org/10.1142/S0218001417500410>.
- Coy, T., Hughes, V., Harrison, P., & Gully, A. J. (2021). A comparison of the accuracy of dissen and keshet’s (2016) deepformants and traditional lpc methods for semi-automatic speaker recognition. *Interspeech 2021*, (pp. 406–410). doi:<https://doi.org/10.21437/Interspeech.2021-1487>.
- Dattorro, J. (1997). Effect design, part 1: Reverberator and other filters. *Journal of the Audio Engineering Society*, *45*, 660–684.

- Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., & Ouellet, P. (2010). Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19, 788–798.
- Dissen, S., & Keshet, J. (2021). *DeepFormants*. URL: <https://github.com/MLSpeech/DeepFormants>.
- Dissen, Y., Goldberger, J., & Keshet, J. (2019). Formant estimation and tracking: A deep learning approach. *The Journal of the Acoustical Society of America*, 145, 642–653. doi:<https://doi.org/10.1121/1.5088048>.
- Drygajlo, A., Jessen, M., Gfroerer, S., Wagner, I., Vermeulen, J., & Niemi, T. (2015). *Methodological guidelines for best practice in forensic semiautomatic and automatic speaker recognition*. Verlag f"ur Polizeiwissenschaft.
- Gargouri, D., Kammoun, M. A., & Hamida, A. B. (2006). A comparative study of formant frequencies estimation techniques. In *Proceedings of the 5th WSEAS International Conference on Signal Processing, Istanbul, Turkey* (pp. 15–19).
- Gauvain, J.-L., & Lee, C.-H. (1994). Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Transactions on Speech and Audio Processing*, 2, 291–298. doi:<https://doi.org/10.1109/89.279278>.
- Gowda, D. N., Bollepalli, B., Kadiri, S. R., & Alku, P. (2021). Formant tracking using quasi-closed phase forward-backward linear prediction analysis and deep neural networks. *IEEE Access*, 9, 151631–151640. doi:<https://doi.org/10.1109/ACCESS.2021.3126280>.
- Guillemin, B. J., & Watson, C. (2008). Impact of the gsm mobile phone network on the speech signal: Some preliminary findings. *International Journal of Speech, Language & the Law*, 15. doi:<https://doi.org/10.1558/IJSLL.V15I2.193>.

- Hansen, J. H., & Hasan, T. (2015). Speaker recognition by machines and humans: A tutorial review. *IEEE Signal Processing Magazine*, 32, 74–99. doi:<https://doi.org/10.1109/MSP.2015.2462851>.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.
- Hedeker, D. (2005). *Generalized linear mixed models*. Encyclopedia of Statistics in Behavioral Science.
- Honglin, C., & Jiangping, K. (2012). Speech length threshold in forensic speaker comparison by using long-term cumulative formant (ltcf) analysis. In *2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control* (pp. 418–421). doi:<https://doi.org/10.1109/IMCCC.2012.103>.
- Hughes, V. (2017). Sample size and the multivariate kernel density likelihood ratio: How many speakers are enough? *Speech Communication*, 94, 15–29. doi:<https://doi.org/10.1016/j.specom.2017.08.005>.
- Kameny, I., Brackenridge, W. A., & Gillmann, R. (1974). Automatic formant tracking. *The Journal of the Acoustical Society of America*, 56, S28–S28. doi:<https://doi.org/10.1121/1.1914097>.
- Kåre, S. (2021). *Snack Sound Toolkit*. URL: <http://www.speech.kth.se/snack>.
- 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 *Audio Engineering Society Conference: 2019 AES International Conference on Audio Forensics*.
- Mandasari, M. I., McLaren, M. L., & van Leeuwen, D. A. (2011). *Evaluation of i-vector speaker recognition systems for forensic application*. Florence, Italy.

- Matz, M. V., & Nielsen, R. (2005). A likelihood ratio test for species membership based on dna sequence data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1969–1974. doi:<https://doi.org/10.1098/rstb.2005.1728>.
- Morrison, G. S. (2009a). Forensic voice comparison and the paradigm shift. *Science & Justice*, 49, 298–308. doi:<https://doi.org/10.1016/j.scijus.2009.09.002>.
- Morrison, G. S. (2009b). Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs. *The Journal of the Acoustical Society of America*, 125, 2387–2397. doi:<https://doi.org/doi.org/10.1121/1.3081384>.
- Morrison, G. S. (2011a). A comparison of procedures for the calculation of forensic likelihood ratios from acoustic–phonetic data: Multivariate kernel density (mvkd) versus gaussian mixture model–universal background model (gmm–ubm). *Speech Communication*, 53, 242–256. doi:<https://doi.org/10.1016/j.specom.2010.09.005>.
- Morrison, G. S. (2011b). Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*, 51, 91–98. doi:<https://doi.org/10.1016/j.scijus.2011.03.002>.
- Morrison, G. S., Rose, P., & Zhang, C. (2012). Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice. *Australian Journal of Forensic Sciences*, 44, 155–167. doi:<https://doi.org/10.1080/00450618.2011.630412>.
- Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10, 19–41. doi:<https://doi.org/10.1006/dspr.1999.0361>.

- Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, 3, 72–83. doi:<https://doi.org/10.1109/89.365379>.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. arXiv. URL: <http://arxiv.org/abs/1602.04938>. arXiv:1602.04938.
- Rose, P., & Winter, E. (2010). Traditional forensic voice comparison with female formants: Gaussian mixture model and multivariate likelihood ratio analyses. *SST*, (pp. 42–45).
- Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science*, 309, 892–895. doi:<https://doi.org/10.1126/science.1111565>.
- Shriberg, E., Graciarena, M., Bratt, H., Kathol, A., Kajarekar, S. S., Jameel, H., Richey, C., & Goodman, F. (2008). Effects of vocal effort and speaking style on text-independent speaker verification. In *Ninth Annual Conference of the International Speech Communication Association*. 22–26 September 2008, Brisbane, Australia.
- Snell, R. C., & Milinazzo, F. (1993). Formant location from lpc analysis data. *IEEE Transactions on Speech and Audio Processing*, 1, 129–134. doi:<https://doi.org/10.1109/89.222882>.
- Snyder, D., Garcia-Romero, D., Sell, G., Povey, D., & Khudanpur, S. (2018). X-vectors: Robust dnn embeddings for speaker recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5329–5333). 15–20 April 2018, Calgary, Alberta, Canada.
- Titze, I. R., & Martin, D. W. (1998). *Principles of voice production*. Acoustical Society of America.
- Tsuge, S., & Ishihara, S. (2018). Text-dependent forensic voice comparison: Likelihood ratio estimation with the hidden markov model (hmm) and gaus-

- sian mixture model. In *Proceedings of the Australasian Language Technology Association Workshop 2018* (pp. 17–25).
- Van Heuven, V. J. (2016). An acoustic characterisation of English vowels produced by american, dutch, chinese and hungarian speakers. *Hungarian Journal of Applied Linguistics*, *16*, 1–20.
- Van Leeuwen, D. A., & Br"ummer, N. (2007). An introduction to application-independent evaluation of speaker recognition systems. In *Speaker classification I* (pp. 330–353). Springer. doi:https://doi.org/10.1007/978-3-540-74200-5_19.
- Wang, H., & Zhang, C. (2015). Forensic automatic speaker recognition based on likelihood ratio using acoustic-phonetic features measured automatically. *Journal of Forensic Science and Medicine*, *1*, 119. doi:<https://doi.org/10.4103/2349-5014.169617>.
- Young, S. J., & Young, S. (1993). *The HTK hidden Markov model toolkit: Design and philosophy*. Cambridge University Engineering Department.
- Zhang, C., Morrison, G. S., Enzinger, E., & Ochoa, F. (2013). Effects of telephone transmission on the performance of formant-trajectory-based forensic voice comparison—female voices. *Speech Communication*, *55*, 796–813. doi:<https://doi.org/10.1016/j.specom.2013.01.011>.

A mássalhangzókra ható prozódiai erősítés vizsgálata a magyarban

Markó Alexandra¹, Deme Andrea^{2,1}, Juhász Kornélia^{2,1,3}, Bartók Márton¹,
Csapó Tamás Gábor^{4,1}, Grácz Tekla Etelka^{1,3}

¹MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport

²Eötvös Loránd Tudományegyetem

³Nyelvtudományi Kutatóközpont

⁴Budapesti Műszaki és Gazdaságtudományi Egyetem
Távközlési és Médiainformatikai Tanszék

Abstract

Prosodic strengthening is the spatial and/or temporal extension of articulatory gestures that occurs at prosodically salient locations, such as the edges of larger prosodic units or stressed syllables. Hungarian consonants have not been analysed with reference to traits of prosodic strengthening so far. In a material of pseudowords with the structure $pV_1pV_1pV_1pV_1$ forming independent utterances (recorded previously for the examination of prosodic strengthening in vowels), bilabial voiceless plosive consonants were analyzed in terms of the number of bursts, duration, VOT, spectral moments and intensity, as a function of syllable position (first and stressed, internal and final) and adjacent vowels' quality (/i/, /u/, /v/ and /a:/). The presence and the number of bursts indicated prosodic strengthening (especially) at the domain-initial edge of the phrase. A lack of burst was not detected in the first, stressed syllable. Realizations containing one burst were also most typical for the first syllable, and the occurrence of more than one burst was the least typical here. As for VOT, the interaction of syllable position and vowel quality had a significant effect, and the model explained the data to a large extent. In the first syllable, shorter VOT was measured; however, the difference between the syllable positions (first vs. final) reached the significance level only in close vowels. CoG and SD were lower, while skewness and kurtosis were higher in the first syllable than in the final syllable, reaching the significance level only in open vowels. The intensity of bursts was also higher in the first syllable with a gradual and significant decrease towards the end of the phrase, independently of vowel quality. In general, prosodic strengthening was observed in the first and also stressed syllables' consonant; however, in most parameters, this effect was dependent on the height of the adjacent vowel. Furthermore, in the analysed parameters the final syllable differed significantly from the first one, which may serve as a clue for boundary perception. Results support the theory that different languages use different combinations of acoustic keys to indicate stress and prosodic boundaries.

Keywords: prosodic strengthening, consonants, Hungarian

Email addresses: marko.alexandra.phd@gmail.com (Markó Alexandra),
deme.andrea@btk.elte.hu (Deme Andrea), juhasz.kornelia@nytud.hu (Juhász Kornélia),
bartokmarton@gmail.com (Bartók Márton), csapot@tmit.bme.hu (Csapó Tamás Gábor),
graczi.tekla.etelka@nytud.hu (Grácz Tekla Etelka)

1. Bevezetés

A prozódiai szerkezet két, a beszédprodukciónak szempontjából lényeges funkciót lát el: az egyik a határjelölés, azaz a prozódiai alkotóelemek hierarchikus csoportosítása, a másik pedig a prominencia, azaz a prozódiai alkotóelemek közötti viszony jelölése (Pierrehumbert, 1999: 679). Egy adott megnyilatkozás prozódiai szerkezete természetesen a szegmentális szerkezet sajátosságaira is hatással van. *Prozódiai erősítésnek* (*prosodic strengthening*, vö. Cho, 2005, 2015) nevezik azt a jelenséget, hogy a prozódiai jellemzők kifejeződnek a szegmentális fonetikai sajátosságokban. A terminus az artikulációs gesztusok olyan térbeli és/vagy időbeli expanziójára utal, amely feltűnő helyeken jelentkezik, mint amilyenek a prozódiai egységek szélei vagy a hangsúlyos szótagok. (Más megközelítésben a szerzők – vö. pl. White et al., 2020 – szétválasztják a gesztusok térbeli expanzióját, ezt nevezve *erősítésnek* (*strengthening*), valamint az időbeli expanziót (*nyújtás/lengthening*). A jelen tanulmányban azonban a fenti definíciót alkalmazzuk, amely mind a térbeli, mind az időbeli expanziót, attól függetlenül, hogy ezek együtt járnak-e, erősítésként értelmezi a prozódiailag kitüntetett helyeken.)

A jelenség kapcsán meg kell említenünk Lindblom (1990) H&H elméletét, amelynek az elnevezése a *hyperspeech* és a *hypospeech* kifejezések rövidítéséből állt elő (418). Az előbbi a fonetikai információban gazdag (*overarticulate/d*], *túl/hiperartikulált*), az utóbbi pedig az információhiányos (*underarticulate/d*], *alulartikulált*) jelre utal, amelyeket egy kontinuum két végpontjaként képzelhetünk el. Lindblom szerint a beszélő az információgazdagság és az információ alulszabottság végpontjai közötti megvalósítási formák közül választ az alapján, hogy mit feltételez a hallgató számára (a jeltől függetlenül) elérhető információról.

Mind a hangsúlyos szótagokban tapasztalható, mind a nagyobb prozódiai egységek határán megvalósuló prozódiai erősítés vezethet lokális hiperartikulációhoz, amelyet az motivál, hogy a beszélő a hallgatói feldolgozáshoz kíván

több kulccsal szolgálni (ami persze nem feltétlenül tudatos). Lindblom (1990) terminusát továbbvive de Jong (1995) a hangsúlyos szótagok elemzése kapcsán az általa *lokális hiperartikulációnak* (*localized hyper-articulation*) nevezett jelenségre alkalmazta, azaz a gesztusoknak a prominencia hatására megvalósuló expansziójára. A *lokális hiperartikuláció/túlargikulálás* terminust kiterjeszthetjük a prominens szótagok mellett más kitüntetett fonetikai helyzetekre is, és Cho és munkatársai (2011: 345) alapján mindezekre alkalmazhatjuk a *prozodikusan motivált (lokális) túlargikulálás* (*prosodically driven (local) hyper-articulation*) terminust. A jelenséget már 1997-ben dokumentálta Fougeron & Keating (ők így nevezték: *articulatory strengthening at prosodic-domain edges*), és mind a térbeli (extrém artikulációs gesztusok), mind az időbeli (tartamtöbblet) dimenzióban dokumentált eltéréseket ide értették.

A frázisvégi nyújtás/nyúlás (az angol nyelvű szakirodalomban *preboundary* vagy *domain/phrase-final lengthening*) az egyik legszembevetőbb és leggyakoribb példája annak, hogy a prozódia hatással van a szegmentális szerkezet megvalósítására Cho (2015). Ez a hatás abban jelentkezik, hogy a prozódiai határhoz közeledve a szegmentumok időzítése megváltozik, azaz a szegmentumok időtartama hosszabb az intonációs frázis végén, mint a frázis belsejében (pl. Edwards et al., 1991; Byrd, 2000). A frázis eleji nyújtás/nyúlást (az angol nyelvű szakirodalomban *postboundary* vagy *domain-initial lengthening*, vö. Cho, 2015: 519) elsősorban mássalhangzók esetében dokumentálták, és bár jellemzően *nyúlásként* említik, itt is a gesztusok nagyobb (és nem feltétlenül időbeli) expansziójáról van szó. Artikulációs vizsgálatokban, elektropalatográfias módszerrel elemelve a koronális mássalhangzók esetében nagyobb kiterjedésű lingvoplatális érintkezést találtak magasabb szintű prozódiai egységek (intonációs frázis) elején, mint alacsonyabb szintű határ (pl. szóhatár) esetében (összefoglalóan lásd Cho et al., 2007; Ip & Cutler, 2022), tehát az erősítés valóban bármely dimenzióban (térben, azaz a gesztusok térbeli megvalósításában, illetve időben, azaz a gesztusok időtartamában) megvalósulhat. Akusztikai vizsgálatok alapján pedig intenzitás-erősödésre és a spektrális momentumok (azaz a spektrum

frekvenciaátlag, csúcossága és ferdesége, vö. Forrest et al., 1988) eltérésére is vannak adatok (lásd a vonatkozó szakirodalom összefoglalását alább).

Az a tény, hogy mind a frázis végén, mind a fráziskezdő helyzetben mutatnak időbeli és/vagy térbeli erősítést a különféle nyelvek, akár a mássalhangzókra, akár a magánhangzókra vagy mindkét szegmentumtípusra vonatkozóan, arra utalhat, hogy mindkét esetben univerzális tendenciáról van szó. A nyelvek közötti eltérések alapján ugyanakkor arra lehet következtetni, hogy a jelenség fonetikai implementációja a beszélő által kontrolláltan, nyelvenként sajátos módon megy végbe, amely további tanulmányozást igényel. A szakirodalom áttekintése nyomán azt is láthatjuk, hogy a különféle prozódiai szerkezettel rendelkező nyelvek eltérő módon és mértékben jelenítik meg a prozódiai erősítés mintázatait ezeken a kitüntetett helyeken. Mivel a frázisvégi nyúlásnak és a hangsúly indukálta prozódiai erősítésnek magyar nyelven is viszonylag nagy irodalma van (lásd az alább hivatkozott forrásokat), az alábbiakban a frázis kezdetén mérhető prozódiai erősítésre koncentrálunk, és azon belül is a mássalhangzókra, illetve az explozívákra, hiszen a jelen vizsgálatunk a /p/ különféle fonetikai helyzetekben mutatkozó realizációira irányul.

Akusztikai vizsgálatok alapján az angol, a német és a koreai nyelvben a zöngétlen hehezetes (aspirált) felpattanó zárhangok zöngékezdési ideje (azaz VOT-ja) hosszabbnak bizonyult fráziskezdő helyzetben, mint frázis belsejében (lásd Ip & Cutler, 2022 összefoglalását); másrészt a holland zöngétlen (aspirálatlan) explozíva /t/ esetében VOT-rövidülést figyeltek meg a hangsúlyos szótagokban és a fráziskezdetben a frázis belsejéhez képest (Cho & McQueen, 2005). A VOT-t az artikulációs erőfeszítéssel szokás összefüggésbe hozni, amit a fortisz-lénisz skálával képeznek le (vö. pl. Lisker & Abramson, 1964). Itt a fortisz az erőteljesebb ejtésre, vagyis nagyobb izomfeszítettségre utal, és a zöngétlenekre jellemző, tehát nagyobb pozitív VOT társul hozzá, szemben a lénisz (zöngétlen, negatív VOT) ejtéssel. Ezen a skálán a hehezetes ejtésű explozíva a fortisz végponthoz közelebb helyezkedik el (a nem hehezetes zöngétlenhez képest is), ami lényegében azt jelenti, hogy a hangsúly hatására az angolban a nagyobb erőfeszítéssel

képzett felé tolódik el az ejtés (nagyobb pozitív VOT), míg a hollandban ellenkezőleg.

Vicenik (2010) a grúz nyelv három explozívátípusát (zöngétlen hehezetes, ejektív és zöngés) vizsgálta a bilabiális, alveoláris és veláris képzéshelyeken, tekintettel többek között a prozódiai erősítésre is. Akusztikai méréseket végzett egyebek mellett a zárfeloldás intenzitására és a spektrális momentumokra (azaz a spektrum frekvenciaátlagára, csúcsosságára és ferdeségére) vonatkozóan. A magasabb rendű prozódiai helyzetnek (intonációs frázis eleji, szó eleji) kismértékű hatása volt a zárfeloldás intenzitására a hehezetes és a zöngés felpattanó zárhangokban, ugyanakkor az ejektívákban éppen az alacsonyabb szintű (szó belseji) pozíció esetén volt intenzívebb a zárfeloldás. A grúz alveoláris explozívák nem mutattak eltérést a zárfeloldást kísérő turbulens zörej átlagos frekvenciaértékében, de kisebb csúcsosság- és ferdeségértékeket kapott a szerző a magasabb rendű (intonációs frázis eleji, szó eleji) prozódiai helyzetekben. A veláris zárhangok magasabb átlagos frekvenciaértékeket mutattak, ami alacsonyabb csúcsosság- és ferdeségértékekkel párosult a magasabb rendű prozódiai helyzetekben. A bilabiális zöngés zárhangoknál hasonlóképpen magasabb átlagos frekvenciaértékek, valamint alacsonyabb csúcsosság- és ferdeségértékek adódtak a magasabb rendű prozódiai helyzetekben, míg a bilabiális hehezetes zárhang és a bilabiális ejektíva alacsonyabb frekvenciaátlagot mutatott ugyanabban a helyzetben, de nem volt eltérés a csúcsosságban és a ferdeségben az alacsonyabb rendű prozódiai helyzethez (szó belseje) képest.

Mindezek alapján tehát azt mondhatjuk, hogy a prozódiai erősítés megjelenése az adott nyelv mássalhangzókészletétől és az abban található fonemikus kontrasztoktól függően alakul, ugyanakkor hatással vannak rá az adott nyelv prozódiai sajátosságai is. Cho és munkatársai (2011) az angol és a koreai nyelv között talált eltéréseket a nyelvspecifikus prozódiai mintázatokra vezették vissza, mégpedig arra, hogy a lexikai hangsúly (lexical stress) és dallamhangsúly (pitch accent) nélküli nyelvekben (pl. koreai) nagyobb az artikulációs gesztusok térbeli és időbeli expanziójának a mozgástere, mint az olyan nyelvekben (pl. az angolban), ahol a lexikai prominencia korlátozza ennek a lehetőségeit. Korábban

a frázis eleji erősödést négy nyelvben, köztük szintén az angolban és a koreai-ban is vizsgálva Keating és munkatársai (2004) arra a következtetésre jutottak, hogy a koreai-ban robusztusabb a fráziskezdet erősítő hatása, mint az angolban, mivel a koreai-ban a prozódiai szerkezet elsősorban a tagolást szolgálja, míg az angolban a tagolás mellett a prominenciát is jelöli. A magyar nyelv sajátosnak mondható ebből a szempontból, mert ugyan lexikai szinten nincs a hangsúlynak megkülönböztető szerepe, és a fonémarealizációk jellegzetességeit tekintve sem térnek el a hangsúlyos és a hangsúlytalan szótagbeli magánhangzók (nincs szisztematikus magánhangzó-redukció a hangsúlytalan szótagban, vö. Gósy, 1997), egyes tanulmányok eredményei mégis a szóhangsúlyos magánhangzó akusztikai szálenciájára utalnak (pl. Szalontai et al., 2016; Mády et al., 2017). Ugyanakkor az első szótagi kötött hangsúly miatt a magyarban igen gyakori, hogy egybeesik a fráziskezdet és a (szó)hangsúly, így nem is mindig választható szét a két prozódiai erősítő hatás.

Tabain és munkatársai (2016) öt, különböző mássalhangzókészlettel rendelkező és eltérő prozódiai mintázatokat mutató nyelvben vizsgálták a szóhangsúly hatását a felpattanó zárhangok megvalósulására, zöngétlen zárhangokat elemezve (az egyes nyelvekben 3, 4, 5 vagy 6 képzéshely alkotott fonológiai kontrasztot), referenciaként az angol nyelv 3 zöngétlen felpattanó zárhangját alkalmazva. A magyarhoz a vizsgált nyelvek közül az Indonéziában beszélt makassar nyelv hasonlít a leginkább az ott található négy képzéshellyel, melyek a bilabiális (/p/), az alveoláris (/t/), a palatális (/c/) és a veláris (/k/). Ugyanakkor ebben a nyelvben a szóhangsúly az utolsó előtti (vagy elég gyakran az azt megelőző) szótagra esik – amely (két, illetve három szótagos szó esetén) akár egybe is eshetne az első szótaggal, de a tanulmány a vizsgált szavak szótagszámáról nem szolgál információval. A vizsgálat tárgyát két olyan nyelv is képezte, amelyben a magyarhoz hasonlóan első szótagi lexikális hangsúly van, ezek a pitjantjatjara és a warlpiri, ezekben azonban öt-öt zöngétlen zárhang alkot kontrasztot. Habár egyik nyelv sem vethető össze egy az egyben a magyarral, az elemzések tanulságosak a jelen vizsgálatunk szempontjából is. A kutatásban az időtartamot, a spektrális lejtést és a négy spektrális momentumot (spektrális súlypont, azaz

CoG; szórás, ferdeség, csúcsosság) elemezték, az utóbbiak közül csak az első kettőt mutatták be a tanulmányban. A zárfeloldás időtartama az angolban a hangsúlyos szótagok mássalhangzóiban hosszabb volt, mint a hangsúlytalankéban; a makassarban nem találtak ilyen eltérést. Ez a magyar szempontjából azért releváns, mert a magyarhoz hasonlóan az angolban és a makassarban is van zöngésségi kontraszt az explozívákban, de míg az angol aspiráló nyelv, ahol a zöngétlen aspirált (magas pozitív VOT) és a zöngétlen nem aspirált (alacsony pozitív VOT) áll szemben egymással (azaz pozitív VOT-kontraszt van), addig a makassar és a magyar zöngésítő nyelv, ahol a zöngés (negatív VOT) és a zöngétlen (alacsony pozitív VOT) áll szemben (azaz negatív VOT-kontraszt van). A szerzők más kutatások alapján is úgy találták, hogy az angolban a zöngétlen (és aspirált) explozívák kevésbé aspiráltak (értsd: kisebb a VOT-értékük) akkor, ha hangsúlytalan szótagban jelennek meg; ugyanakkor úgy vélték, hogy ha a makassar mássalhangzókészletből figyelembe vették volna az előzöngés hangzókat (negatív VOT), abban a körben eltérést találtak volna a hangsúlyos és a hangsúlytalan szótagbeli előfordulások VOT-értékei között. A pitjantjatjara esetében kismértékű eltérést találtak a hangsúlyos és a hangsúlytalan szótagbeli /p/-k VOT-je között (kisebbit, mint a /c/ és a /k/ esetében), nem találtak ugyanakkor ilyen hatást a warlpiri nyelvben. A CoG az angolban csak a /k/ esetében mutatott eltérést a hangsúlyos és a hangsúlytalan szótagok között (alacsonyabb a CoG a hangsúlyosakban), a makassarban viszont nem volt eltérés egyik mássalhangzóban sem. A pitjantjatjara az angollal egyezően a /k/-ban alacsonyabb CoG-t mutatott a hangsúlyos szótagban, a warlpiriban viszont semmilyen hatás nem volt mérhető. A spektrális szórásban kizárólag az angol /k/ esetében volt eltérés. Összességében tehát a mássalhangzókészletét tekintve a magyarhoz leginkább hasonló makassarban semmilyen (a szignifikanciát is elérő) hatást nem találtak a szóhangsúlyosság függvényében a VOT-t és a spektrális momentumokat illetően, sem a /p/, sem a többi zöngétlen zárhang esetében, ugyanakkor a szerzők utalnak a nagymértékű egyéni variációra. Azt is említik, hogy ebben a nyelvben nagy időtartambeli különbség van a hangsúlyosság függvényében a magánhangzók között, és úgy vélik, ez a tény is támogatja azt

az elméleti elgondolást, amely szerint a különféle nyelvek az akusztikai kulcsok eltérő kombinációit alkalmazzák a hangsúly jelzésére (Cutler, 2012).

A magyarra vonatkozóan a prozódiai erősítés hatását (még ha nem is feltétlenül nevezték így a szerzők) több tanulmány vizsgálta a hangsúlyos helyzetben lévő magánhangzók körében (pl. Magdics, 1965, 1966; Kassai, 1979; Gósy, 1997; Kovács, 2001; Mády et al., 2008; Gósy & Beke, 2010; Szalontai et al., 2016; Mády et al., 2017; Markó et al., 2018, 2019; Deme et al., 2019a,b, 2021, 2022, megj., kézirat), illetve a legújabban hangsúlyos és frázisszállóan lévő magánhangzókban is (Markó et al., 2022). Bár az eredmények nem vágnak egybe, ami vélhetően a számos jellegzetességet tekintve eltérő módszertanokból adódik, többségükben mégis azt igazolják ezek a kísérletek, hogy a hangsúlyos szótagi magánhangzók akusztikai és artikulációs sajátosságai esetenként (a magánhangzó-minőségtől nem függetlenül) eltérnek a hangsúlytalan szótagbeli megvalósulásuktól. Céltan mássalhangzókra vonatkozó vizsgálatokról a jelen tanulmány szerzőinek nincs tudomása, de mivel a magyarban gyakran fráziskezdő is az, ami hangsúlyos, az alábbiakról adhatunk számot a szakirodalom alapján.

A magyar fráziskezdő mássalhangzókra vonatkozó adatok ellentmondásosak: míg Magdics (1966) rövidebbnek találta a mássalhangzókat a felolvasott mondatok elején (ugyanazon beszédhangoknak a hangsor belsejében található előfordulásaihoz képest), addig Kassai (1979) éppen ellentétes tendenciát talált. Az eltérések mögött módszertani eltérések is húzódnak (vö. Kohári, 2018), ugyanakkor más nyelvek esetén is dokumentáltak hasonlóképpen ellentmondó tendenciákat. White (2002) angol nyelvű kísérleteiben például hangsúlyos és hangsúlytalan helyzetben is vizsgált különböző szavakat mind frázis elején, mind frázis belsejében, és kevert mintázatokat talált a különféle szócsoporthoz az egyes pozíciók tekintetében.

A különféle prozódiai határokhoz közeledve megjelenő nyújtás jelenségét a fentebb idézett szerzők univerzálisnak tartják, illetve a nyúlás – bár eltérő mértékben és módon – érintheti mind a magánhangzókat, mind a mássalhangzókat. A nyúlás mértéke függ a frázis végéhez képest elfoglalt helyzettől: a frázis utolsó szótagjában a legnagyobb mértékű a nyúlás, de már a frázis végéhez közeledve is

megfigyelhető az időtartamtöbblet (ha nem is feltétlenül fokozatosan/lineárisan nő a szegmentumok tartama). A nyúlás mértéke a szótagsajátosságoktól (zárt vagy nyílt szótag), továbbá az érintett szegmentumok különféle jellemzőitől is függ, ilyen például a fonológiai kvantitás, a magánhangzó-minőség, illetve a mássalhangzók képzésmódja (összefoglalóan vö. Kohári, 2018).

A magyarban a frázisvégi nyúlást több tanulmány is adatolta (pl. Magdics, 1966; Kassai, 1979; Kovács, 2001; Kohári & Markó, 2015; Gósy & Krepsz, 2016), mind magánhangzók, mind mássalhangzók időtartamában. Habár a módszertanok eltérnek, és az eredmények ezért nehezen összevethetők, a magyar beszéd is egyértelműen mutatja ezt a globális tendenciát. A komplex időszerkezetű beszédhangok (például a felpattanó zárhangok) esetében ugyanakkor a belső időszerkezetre (például a zöngekezdési időre) vonatkozóan nincsenek még magyar nyelvű kutatások.

A jelen tanulmány célja a prozódiai erősítés vizsgálata, azaz a prozódiai szempontból száliens fonetikai helyzeteknek (a fráziskezdő és egyben hangsúlyos, illetve frázisvégi helyzetnek) a szegmentális szerkezetben kifejeződő akusztikai mintázatait elemezzük, ezúttal a szótagkezdő (CV szótagbeli) mássalhangzóban. A vizsgálat jelentőségét főként abban látjuk, hogy a magyarban – az eddig legtöbbet vizsgált nyelvektől több szempontból eltérően – nem elsősorban prozódiai, hanem szintaktikai fókuszjelölés jellemző, ritmikai szempontból szótag-időzítésű nyelv, kötött szóhangsúllyal rendelkezik, és hangsúlytalan szótagban nem mutat fonológiai magánhangzó-neutralizációt.

A vizsgálat tárgyát négy szótagos álszavak önálló megnyilatkozásként való meghangosítása képezte, amelyben a kezdő szótag viseli a frázishangsúlyt, azaz egyszerre jelentkezhet a fráziskezdethez és a prominenciához kötődően prozódiai erősítés. Vizsgáltuk a frázisvégi nyújtás jelenségét is. E két szélső pozícióban jelentkező mássalhangzók szegmentális jellemzőit viszonyítjuk a közbülső mássalhangzókéhoz a következő paraméterek tekintetében:

- a felpattanások száma (ideértve a felpattanás esetleges elmaradását is);

- a zárhang időtartama (zöngétlen felpattanóról lévén szó ezt csak a 2–4. szótagban tudtuk elemezni);
- a zöngekezdesi idő (VOT);
- a négy spektrális momentum, azaz a spektrális súlypont (CoG), a szórás, a ferdeség és a csúcsosság;
- a zárfeloldás intenzitásának mediánja és annak a követő magánhangzó legnagyobb intenzitásától való eltérése.

A vizsgálat központi kérdése tehát az, hogy a fenti paraméterekben eltérnek-e a feltételezett prozódiai erősítést mutató első és utolsó szótagokban álló mássalhangzók a frázis belseji szótagokban ejtett mássalhangzóktól. Azt, hogy a prozódiai erősítés hatása legalábbis bizonyos esetekben megmutatkozik a frázisok kezdetén és végén, egy korábbi, ugyanezen az anyagon végzett mérésorozatunk (Markó et al., 2022) a magánhangzók tekintetében igazolta: magánhangzó-minőségtől függő mértékű és hatókörű expansziót láttunk az időtartamban a frázis végén; a frázis kezdetén (és egyben a hangsúlyos szótagban) a hosszú és legalsó nyelvállású /a:/ mutatott egyértelmű artikulációs gesztusbeli expansziót (nagyobb ajaknyílást). Mivel a prozódiai erősítés a különféle nyelvekben és esetekben eltérő kombinációkban jelenhet meg, elképzelhető, hogy a jelen vizsgálatban a zármássalhangzók elemzése kapcsán a magánhangzóknak látottaktól eltérő (vagy éppen ellentétes) tendenciákat kapunk.

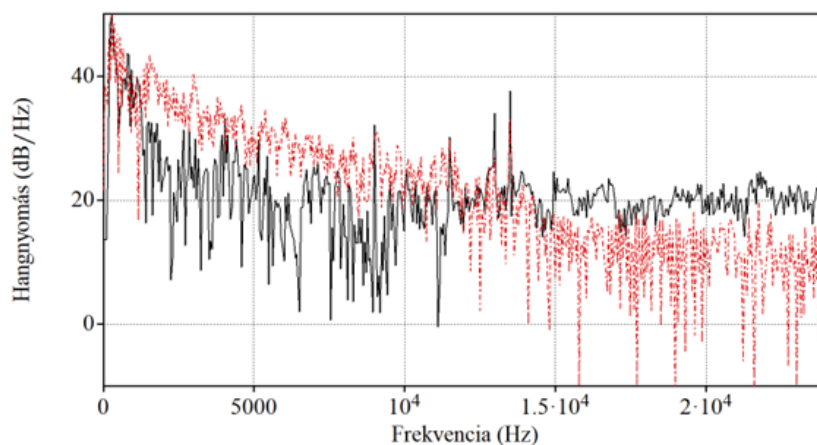
Meg kell jegyeznünk, hogy jelentősége van annak, hogy a vizsgált mássalhangzók CV szótagok onszetjai, hiszen így az utolsó mássalhangzó nem a hangsort záró szegmentum. Ennélfogva nem számítottunk erőteljes gyengülésre az utolsó szótagi mássalhangzóban (vö. Keating et al., 2004).

Mivel a követő magánhangzó minőségének szerepe van a VOT alakulásában (vö. pl. Gósy, 2000), a zár-felpattanások számának alakulásában (vö. Grácz & Kohári, 2012), valamint különféle szótagpozíciókban megjelenő magánhangzók artikulációs és akusztikai jellemzői is eltérnek a magánhangzó-minőség függvényében (vö. Markó et al., 2022), ezért a szótagpozíció mellett a magánhangzó-

minőség hatását is vizsgáltuk. A magyarban az első szótag egyben hangsúlyos is, ezért itt a prozódiai erősítés hatását vártuk. Grácz & Kohári (2012) szerint a zár-felpattanás elmaradásának hátterében az áll, hogy az izmok feszítettsége és/vagy a zár mögött kialakult nyomás mértéke nem elegendő, ilyen jelenséget tehát az első szótagban nem feltételeztünk, a további szótagokban azonban számítottunk rá. A magyar zöngétlen zárhangok közül a bilabiálisokra a legkevésbé jellemző a több felpattanás Neuberger & Beke (2017) szerint (a rövid /p/-k 27,6%-a volt ilyen a frázisbeli helyzetre nem kontrollált anyagukban). Más irodalmi előzményről nem lévén tudomásunk, nem volt hipotézisünk a zár-feloldás többszöröződésére vonatkozóan a szótaghelyzet függvényében, de azt (Grácz & Kohári, 2012 alapján) feltételeztük, hogy ennek gyakorisága függ a követő magánhangzó minőségétől. Gósy & Ringen (2009) magyar (izoláltan, szólista-ként) felolvasott szavakban hosszabb VOT-értékeket mért szó belsejében álló zöngétlen zárhangokban, mint szó elején.

A spektrális momentumoknak nevezett (Forrest et al., 1988) paraméterek az energiaeloszlást jellemzik a spektrumban. A spektrális középpontot vagy súlypontot (CoG) a különböző frekvenciájú energiakomponensek súlyozott átlagaként határozzuk meg: ha magasabb frekvenciákon koncentrálódik az energia, akkor magasabb a CoG értéke, és fordítva. Ehhez viszonyítva határozható meg a spektrális szórás, azaz a spektrum frekvenciaátlagától mért átlagos eltérés. Az eloszlás szimmetrikus voltát ragadja meg a ferdeség paramétere; illetve a csúcsosság jellemzi az eloszlás formáját. Ha normális az eloszlás, így például a spektrum energiájának, illetve a felpattanó zárhang felpattanási zöreijének spektrális eloszlása, akkor a ferdeség és a csúcsosság értéke is 0. Attól függően, hogy balra vagy jobbra tolódik el a spektrum energiaeloszlása, a ferdeségi érték negatív vagy pozitív. A pozitív csúcsosságérték azt jelzi, ha az eloszlás a CoG-érték környezetében csúcsosabb, mint a normál eloszlás, a negatív pedig azt, ha lapultabb, azaz, hogy mennyire koncentrálódik az energiaeloszlás egy területre. Az 1. ábrán mindezt a jelen kutatás anyagából származó két /p/ mássalhangzó (ugyanazon ejtésből a frázis első = fekete és összefüggő vonallal jelölt, valamint negyedik = piros és szaggatott vonallal jelölt zárhangja) ese-

tében mutatjuk be a 2. fejezetben ismertetett mérési beállítások mentén. A folytonos (fekete) görbe olyan mássalhangzót reprezentál, amelynek a spektrális középpontja 2178 Hz-nél van, a szórás pedig 4657 Hz. A ferdesége 2,97, a csúcsossága 8,076. A szaggatott (piros) görbe által reprezentált /p/ zárfelepattanási zöreijének spektrális súlypontja alacsonyabban, 1744 Hz-nél mérhető, hiszen az alsóbb frekvenciákon magasabb intenzitásértékeket látunk, a magasabb frekvenciartományban pedig alacsonyabb intenzitásértékeket. Ennek a spektrális szórása (vagyis a CoG-től mért átlagos eltérés) a folytonossal jelölt mássalhangzóénál kisebb, 3345 Hz. A ferdeség a másik mássalhangzóhoz hasonlóan pozitív (2,82), hiszen mindkét esetben az alsóbb frekvenciartomány felé tolódik el az eloszlás. A csúcsosság 8,071, hasonlóképpen a folytonos görbével reprezentált /p/-éhez, azaz mindkét esetben pozitívak, ami a CoG-érték körüli frekvenciartományban mindkét irányban alacsonyabb intenzitást jelent.



1. ábra. Két /p/ felpattanási zöreijének energiaeloszlása a frekvencia függvényében (200 Hz alatt szűrés 100 Hz-es simítással, majd 1 kHz-től oktávonként 6 dB-es előerősítéssel).

Habár általában a képzési hellyel összefüggésben vizsgálják a spektrális momentumokat, a korábbi kutatások alapján ezek a zöngésség-zöngétlenség tekintetében is mutatnak eltéréseket. Magasabb frekvenciaértékek adódtak több vizsgált nyelvben is a zöngétlen zárhangok esetében (vö. van Alphen & Smits,

2004; Sundara, 2005; Chodroff & Wilson, 2014), ami percepció kulcsként is szolgál a zöngésségi kategória felismerésében. A grúz bilabiális zöngés zárhangok alacsonyabb átlagos frekvenciaértéket, illetve magasabb ferdeség- és csúcsosságértékeket mutattak az azonos képzéshelyű más explozívákhoz/ejektívákhoz képest (Vicenik, 2010). Ha a zöngés-zöngétlen közötti eltérést a lénisz-fortisz skálára vetítjük, ez azt jelenti, hogy a magasabb frekvencia (CoG) és az alacsonyabb ferdeség és csúcsosság fortiszabb képzésre utal.

A magyar /p t k/ zárhangokat a spektrális momentumokat illetően vizsgálták már ugyan, de annak a kutatásnak a középpontjában nem a mássalhangzó szó- vagy frázisbeli helyzete, hanem a fonológiai hosszúsága állt. Neuberger & Beke (2017) 10 férfi beszélő spontán beszédéből származó 221 darab /p/ átlagos CoG-értékét 1243 Hz-nek mérte. Konkrét számadatokat nem ismertetnek a szerzők, de az kiderül, hogy a ferdeség pozitív értékeket adott, azaz az eloszlás aszimmetrikus, és jobbra nyúlik el; valamint a csúcsosság értékei is pozitívak voltak, vagyis az eloszlás csúcsosabb volt, mint a normál eloszlás. Az intenzitás átlagértékéként a /p/ zárhang esetében 62 dB-t mértek. Meg kell jegyeznünk, hogy az idézett kutatás és a jelen mérések anyaga és módszertana több tekintetben eltért, ezért az ismertetett adatok nem szolgálnak közvetlen referenciaként a jelen elemzéshez.

A nemzetközi és a magyar szakirodalom ismeretében nem tudtunk a vizsgálati kérdésekre és paraméterekre hipotéziseket felállítani, vizsgálatunk exploratív jellegű.

2. Kísérleti személyek, anyag, módszer

A kutatásban kilenc magyar anyanyelvű női beszélő vett részt, átlagéletkoruk $25,2 \pm 5,9$ év volt, ismert hallás- vagy beszédproblémájuk nem volt.

A kutatás anyagát négy szótagos álszavak képezték, amelyeket önálló megnyilatkozásként hangosítottak meg a résztvevők. Az álszavak CV szerkezetű szótagokból álltak, amelyek mindegyikében a /p/ bilabiális zöngétlen explozív volt a mássalhangzó. Azért a bilabiális /p/ kontextust használtuk a kísérletben,

mert ez a hang csak kismértékben lép lingvális koartikulációba a szomszédos magánhangzókkal. Egy álszón belül a vizsgált anyagban a magánhangzó minősége sem változott – ezzel kívántuk elérni azt, hogy ne érvényesülhessen a magánhangzók közötti koartikulációs hatás (vö. Deme et al., 2019a,b, 2022, megj., kézirat), és (a lehetőségek mértékében) állandó legyen a szomszédos mássalhangzó és magánhangzó közötti koartikulációs hatás is. Négy, a magyar magánhangzókészletben legszélsőnek számító magánhangzót vontunk be a vizsgálatba: a hátul képzett és felső nyelvvállású /u/-t, az elől képzett és felső nyelvvállású /i/-t, a hátul képzett és alsó nyelvvállású /ɒ/-t és a nyelv vízszintes helyzetét tekintve centrálisnak tartott és legalsó nyelvvállású /a:/-t. Így végül a következő álszavak (négyelemű szótagsorok) szerepeltek a vizsgálati anyagban: *pupupupu*, *pipipipi*, *papapapa*, *pápápápá*.

A célszavakból típusonként és beszélőnként hat-hat előfordulást vizsgáltunk (összesen 216 hangsort, azaz $864 /p/-\text{előfordulást} = 9 \text{ beszélő} \times 4 \text{ magánhangzó-minőség} \times 4 \text{ szótag} \times 6 \text{ ismétlés}$), melyeket random sorrendben, disztraktorok között rögzítettünk. Az álszavakat képernyőn jelenítettük meg a résztvevők számára, egyszerre egyet, és arra kértük őket, hogy „kijelentő mondat”-ként olvassák fel őket. Az itt bemutatott elemzés az akusztikus csatornán (44,1 kHz-es mintavételi frekvencián, a szájjughoz helyezett omnidirekcionális, kondenzátoros fejmikrofonnal) rögzített felvételen készült, de megjegyezzük, hogy ezzel párhuzamosan artikulációs vizsgálatra alkalmas felvétel is készült elektromágneses artikulográfiával (Markó et al., 2022). (Az eredeti felvételi módszertan indokolja a résztvevők viszonylag alacsony számát.) A jelen tanulmányban csak akusztikai méréseket mutatunk be.

A szegmentumokat az akusztikai jel alapján szegmentáltuk és címkéztük fél-automatikusan a *BAS* webszolgáltatás graféma-fonéma konverterének (Reichel, 2012) és a *MAUS* rendszernek (Schiel, 1999) a segítségével, majd a szükséges helyeken kézzel javítottuk a felismertetett hanghatárokat a *Praat* szoftverben (Boersma & Weenink, 2022). Ezt követően kézileg felcímkéztük az adott mássalhangzó zárfeloldásának kezdetét (az első felpattanás kezdetét), valamint dokumentáltuk a zárfelattanások számát is. Az így felcímkézett mássalhangzók

időtartamát (a 2. szótagtól a 4. szótagig) a megelőző magánhangzó második formánsának végétől a követő magánhangzó második formánsának kezdetéig mértük, a zöngelkedési idejét pedig az első felpattanás kezdetétől a követő magánhangzó zöngéjének megjelenéséig. A zöngelkedési idő és a spektrális mérések esetében is (értelemszerűen) kizártuk azokat a megvalósulásokat, amelyekben nem volt detektálható a zárfeloldás. Ez összesen 30 előfordulást érintett (a realizációk 3,5%-a). A spektrális jellemzőket az alábbi módon mértük a *Praat* szoftver segítségével. A hangfájl intenzitásmenetét 100 Hz-es minimális alaphfrekvencia-beállítással 0,002 s-os ablakolással, átlagolással nyertük ki. A zárfeloldás spektrális jellemzőit annak kezdetétől a zöngé kezdetéig mértük.

A mássalhangzóban a spektrális momentumokat az alábbi módon határoztuk meg. A hangot 200 Hz alatt szűrtük (100 Hz-es simítással), majd 1 kHz-től oktávonként 6 dB-es előerősítést alkalmaztunk. Ezután a gyors Fourier-transzformált spektrumon nyertük ki a spektrális súlypontot, szórást, ferdeséget és csúcosságot a *Praat*-ban (Vicenik, 2010 módszerével egyezően).

A magánhangzóban 5 ms-os ablakkal, 0,5 ms-os eltolással megkerestük a legnagyobb intenzitású időszakaszt, és ennek középpontja körül ± 10 ms-mal mértünk intenzitásmediánt. A zárfeloldásban egyrészt elemeztük a feloldás intenzitását (mediánérték), másrészt a magánhangzóhoz képest mért intenzitását, azaz a magánhangzó fentebb leírt módon megkeresett legintenzívebb pontja körül 20 ms-os intervallumban kimértük az intenzitás mediánját, és ebből kivontuk a /p/ zárfeloldásában mért intenzitásmediánt.

A statisztikai elemzéseket az *R* szoftverben (R Core Team, 2021) végeztük, az ábrák a *ggplot2* csomagban készültek (Wickham, 2016). A kiugró értékeket (a mediántól több mint az interkvartilis tartomány 1,5-szeresére lévő értékeket) kizártuk az elemzésből.

A magánhangzó-kontextus és a szótagsorszám hatását a VOT-re, az időtartamra, a spektrális jellemzőkre, valamint az intenzitásjellemzőkre lineáris kevert modellekkel elemeztük (*lme4* csomag: Bates et al., 2015; *lmerTest* csomag: Kuznetsova et al., 2017). A vizsgált paraméterből kiindulva az időtartam, a VOT, a spektrális jellemzők vagy az intenzitásjellemzők egyike volt a függő

változó, a magánhangzó-kontextus és a szótag sorszáma a fix hatások, közöttük interakciót megengedve. A random hatások a két fix hatásra illesztett beszélőnkénti random meredekségek voltak, a beszélőre és az ismételésre illesztett random konstansokkal. (Ismétlések alatt az egyes logatompéldányokat értjük.) A random meredekségek több esetben szinguláris illesztést okoztak, így azokat ezekben az esetekben elhagytuk, és csak a random konstansokat tartottuk meg. Amennyiben szingularitási probléma nem, hanem konvergációs hiba lépett fel, a *BOBYQA*-optimalizációt (*Bound Optimization BY Quadratic Approximation*, Powell, 2009) alkalmaztunk a maximális függvényértékelés értékének 20000-re emelésével (alapérték: 10000) (Singmann & Kellen, 2019 ajánlásaiból kiindulva). A random, majd a fix hatások alapján történt a modellszelekció az *anova()* paranccsal. A legszűkebb olyan modellt tartottuk meg, amely a legbővebb (fent leírt) modelltől még nem tért el szignifikánsan. Amennyiben mind a magánhangzó, mind a szótagsorszám elhagyása a kritériumoknak megfelelő modellt eredményezett, az alacsonyabb AIC-számút választottuk (Akaike, 1974). A p -értékeket Satterthwaite-approximációval nyertük ki az *anova()* parancs segítségével. A hatásnagyságokat a *MuMIn* (Bartoń, 2022) csomaggal elemeztük. A modellen Tukey-féle post hoc tesztet is futtattunk az interakcióra (*emmeans*: Lenth, 2020). Akkor is az interakcióra készítettünk post hoc próbát, ha az interakció nem bizonyult szignifikáns hatásnak.

A zárfelpattanások számát külön elemeztük szótagonként, magánhangzó-minőségenként és beszélőnként. Az adott faktor mentén összegeztük a zárfelpattanások számát, majd a gyakori cellánkénti alacsony előfordulások és a magas általános elemszám miatt Fisher-féle egzakt próbát alkalmaztunk a p -érték 2000-es szimulációjával (Harrison & Pius, 2021). A magánhangzók, illetve a szótagok páronkénti összevetését is elvégeztük ugyanilyen módon. A szignifikanciaszint korrekcióját a 4 csoportra alkalmaztuk mindkét faktor esetében, a 0,05-ös határértéket ($4 * 3/2 =$) 6-tal osztva, a 0,008-as p -értéket tekintettük szignifikanciaküszöbnek ezekben a post hoc tesztekben.

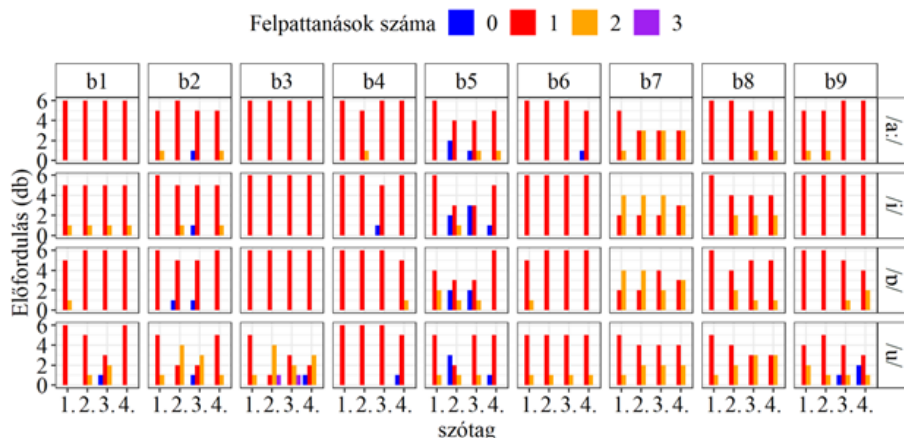
3. Eredmények

3.1. A zárfelpattanások száma

A zárfelpattanás nélkül realizálódott mássalhangzók száma (mint említettük) 30 volt, ami az összes példány 3,5%-át teszi ki. Egy felpattanással 703 előfordulás, azaz a vizsgált beszédhangok 81,4%-a realizálódott, egynél többel pedig 131, azaz 15,2%. A 2. ábra a felpattanások számának eloszlását mutatja beszélőnként, szótagonként és magánhangzó-minőségenként. Látható, hogy a zár feloldásának elmaradása és a többszöri feloldás egy, illetve két beszélőre jellemző főként, de mindkét jelenség jelentősen gyakoribb az /u/ magánhangzó környezetében. A szótagpozíció kapcsán megállapítható, hogy a felpattanás elmaradása az első szótagban nem fordult elő, a többszöri felpattanás pedig inkább beszélőnként változóan alakult a további szótagok között. A második szótagban mindösszesen 10 (4,6%), a harmadikban 14 (6,5%), a negyedikben pedig 6 (2,8%) esetben maradt el a zárfelpattanás. Egy felpattanással realizálódott az első szótagi mássalhangzók 89,4%-a (193 db), a második és a harmadik szótagbeliek 77,3%-a (167 db), valamint a negyedik szótagbeliek 81,5%-a (176 db). Egynél több felpattanással az első szótagban 23 /p/ realizálódott (az első szótagiak 10,6%-a), a másodikban 39 (18,1%), a harmadikban 35 (16,2%), a negyedikben pedig 34 (15,7%). Az eredményeket Fisher-féle egzakt próbával elemeztük a három faktorra külön-külön. Mindhárom faktor esetében szignifikáns hatást találtunk (szótagpozíció: $p = 0,002$; magánhangzó-minőség: $p = 0,001$, beszélő: $p < 0,001$). A szótagpozícióra készített post hoc tesztben azt az eredményt kaptuk, hogy az első szótag szignifikáns eltérést mutat az összes többi szótagtól, míg a többiek szótagok között nincs szignifikáns eltérés a felpattanások számának tekintetében. A magánhangzó-környezet esetében az /u/ mutatott szignifikáns eltérést az összes többi magánhangzótól, míg a többi vizsgált magánhangzó-minőség között nem kaptunk szignifikáns eltérést.

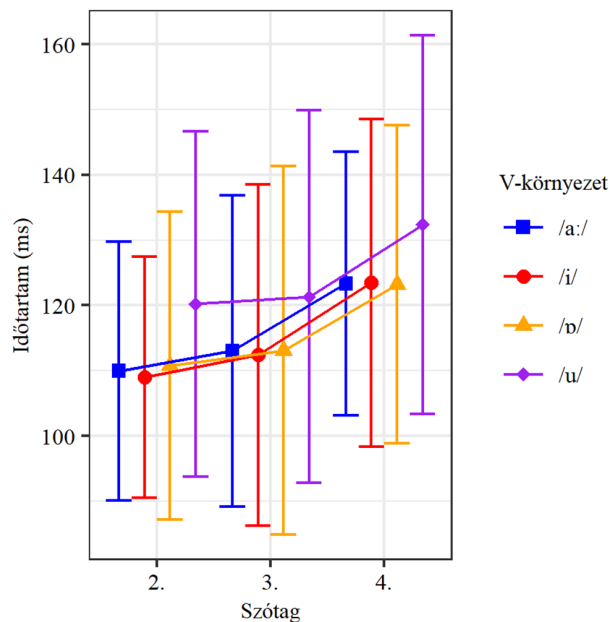
3.2. Időtartam

Az időtartamot a mássalhangzó kezdetén lévő néma fázis miatt a magánhangzókat követő kontextusban, azaz a 2., a 3. és a 4. szótagban tudtuk mérni,



2. ábra. A felpattanások számának eloszlása beszélőnként (b1–b9), szótagpozícióként (1–4.) és magánhangzó-minőségenként.

ennélfogva nincs adatunk arról, hogy a mássalhangzó teljes időtartamára hatással lehetett-e a frázis kezdeti és egyben hangsúlyos helyzet. A csak a szótagok sorszámát mint fix hatást tartalmazó modell bizonyult az adatokat legjobban leíró modellnek. A szótag sorszáma (2–4.) szignifikáns hatással volt a mássalhangzó időtartamára ($F(2, 410, 49) = 92,986, p < 0,001$). Míg a második szótagban mért átlagos időtartamoktól (/a/: 110 ms, /i/: 109 ms, /o/: 111 ms, /u/: 120 ms) a harmadik szótagban mért értékek (113 ms, 112 ms, 113 ms, 121 ms – azonos sorrendben) csak kis mértékben tértek el, a negyedik szótagban az elsőhöz képest legalább 10%-kal hosszabb átlagos időtartamértékeket kaptunk (/a:, i, o/: 123 ms, /u/: 132 ms). A Tukey-féle post hoc teszt alapján a 4. szótagban mért időtartamok bizonyultak szignifikánsan hosszabbnak a korábbi szótagokban mérttől (3. ábra), azaz a frázis végi nyújtás hatását találtuk az adatokban – hasonlóan a magánhangzókhoz (vö. Markó et al., 2022). A modellre kapott hatásnagyságok alapján ($r_m^2 = 0,057, r_c^2 = 0,813$) azonban fontos kiemelni, hogy a szótagpozíció csak kis mértékben magyarázza az adatokat, míg a beszélők közötti variabilitás jelentős.

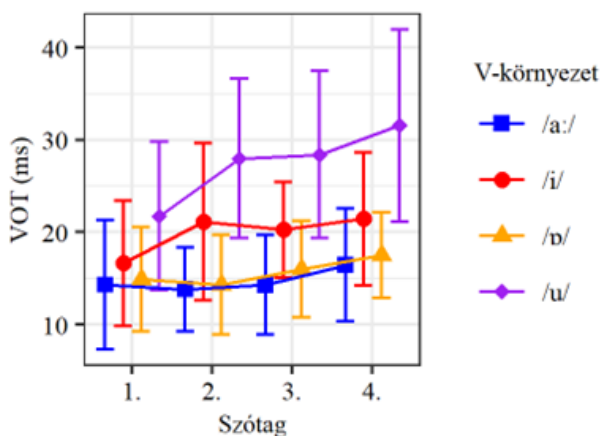


3. ábra. Az időtartam alakulása a szótagpozíció (2–4.) és a magánhangzó-környezet függvényében (átlag és szórás).

3.3. Zöngelkedési idő (VOT)

Az adatok leírására a legjobb modellnek az bizonyult, amely a két faktor (szótagpozíció és magánhangzó-minőség) interakcióját, a beszélőnkénti és ismétlésekénti random konstanst és a magánhangzó-minőségekénti random meredekséget tartalmazta. A szótagpozíció és a magánhangzó-minőség interakciója szignifikáns hatású volt ($F(9, 571, 31) = 2,681$, $p = 0,005$). A marginális hatásnagyság ($r_m^2 =$) 0,344, a kondicionális ($r_c^2 =$) 0,574, tehát a modell viszonylag nagymértékben magyarázza az adatokat. A post hoc tesztek alapján szignifikánsan eltért a /p/ zöngelkedési ideje a következő kombinációk közötti párokban (csak a szélső és a belső szótagok viszonylatában ismertetve az azonos magánhangzó-minőségek esetében): az /i/-s kontextusban az első (16,14 ms) és a második (19,47 ms) szótag között, valamint az első és a negyedik (20,11 ms) szótag között; az /u/-s kontextusban pedig az első szótagbeli /p/ VOT-értéke (20,48 ms) tért el az összes többi (második: 23,88 ms, harmadik: 24,34 ms, ne-

gyedik: 25,43 ms) szótagbeli értéktől. Habár a szignifikanciaküszöböt nem éri el a különbség minden magánhangzó-minőség (mint kontextus) esetében, a 4. ábrán az látszik, hogy a frázis végén a zöngékezdesi idő általánosságban (valamennyivel) hosszabb, mint a frázis kezdetén, ami a frázisvégi prozódiai erősítés (egyik) jele lehet. A 2–4. szótagokban a felső és az alsó nyelvállású magánhangzós kontextusok szinte minden esetben szignifikáns eltérést mutattak (/a:/ vs. /i/ és /a:/ vs. /u/ a 2., 3., 4. szótagban; /ɒ/ vs. /i/ a 2.; /ɒ/ vs. /u/ a 2. és 3. szótagban).

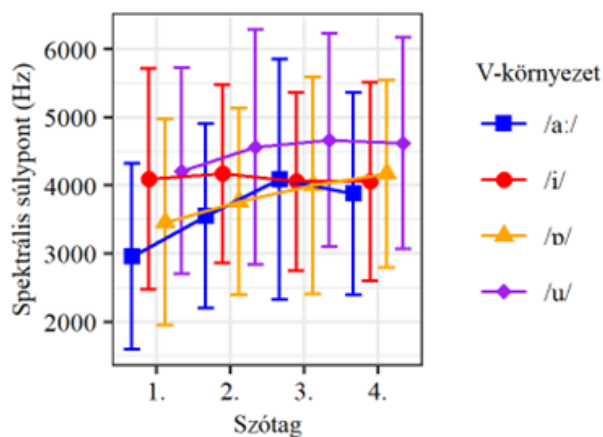


4. ábra. A zöngékezdesi idő (VOT) alakulása a szótagpozíció és a magánhangzó-környezet függvényében (átlag és szórás).

3.4. Spektrális súlypont (CoG)

A spektrális súlypont esetében a legjobb modell tartalmazta a két faktor interakcióját, beszélőnkénti és ismétlésekénti random konstanssal, random meredekség nélkül. A szótagpozíció és a magánhangzó-minőség interakciójának hatása szignifikáns ($F(9, 551, 37) = 2,569$; $p = 0,007$), a marginális hatásnagyság azonban alacsony ($r_m^2 = 0,103$), ugyanakkor a kondicionális hatásnagyság közepes ($r_c^2 = 0,585$), azaz a beszélőnkénti és ismétlésekénti variabilitás bír valamelyes magyarázó erővel. Az 5. ábrán az látszik, hogy a CoG-átlagértékek magánhangzó-minőségenként eltérő mintázatokat mutatnak a különböző szótag-

pozíciókra vonatkozóan. Az /i/ esetét tekintve lényegében nincsenek különbségek a szótagpozíciók között, az /u/ előtti mássalhangzóban vannak ugyan kis-mértékű eltérések, de nem érik el a szignifikanciaküszöböt a post hoc tesztek alapján. Az /v/-t megelőző /p/ spektrális súlypontja az első (3460 Hz) és az utolsó szótag (4172 Hz) között mutat szignifikáns eltérést; az /a:/-környezetben pedig az első (2959 Hz) és a harmadik (4093 Hz), valamint az első és a negyedik (3880 Hz) szótag között. Ezekben az esetekben a magasabb CoG a frázis vége felé fortisabb képzésre utal. Az egyes magánhangzó-környezetek között azonos szótagokban az /a:/ és az /u/ esetében minden szótagban szignifikáns eltérést találtunk a mássalhangzók spektrális súlypontjában, míg az egyéb párok között nem.

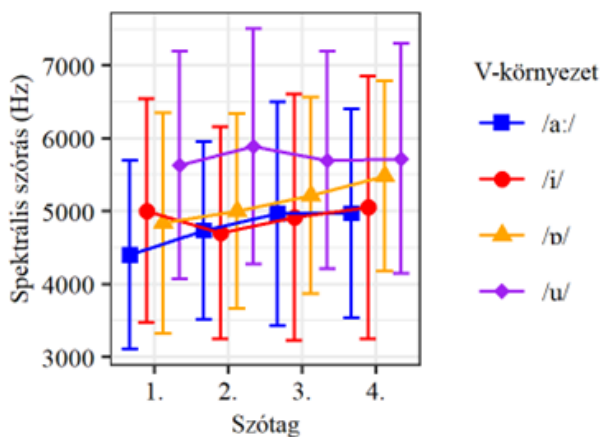


5. ábra. A spektrális súlypont (CoG) alakulása a szótagpozíció és a magánhangzó-környezet függvényében (átlag és szórás).

3.5. Spektrális szórás (SD)

A spektrális szórást leíró legjobb modellnek az bizonyult, ahol mindkét faktor szerepel, a beszélőnkénti és ismétlésenkénti random konstanssal. A két faktor interakciója szignifikáns ($F(9, 585,75) = 2,938$, $p = 0,020$), de kicsi a marginális hatásnagyság ($r_m^2 = 0,076$). A kondicionális hatásnagyság alapján ($r_c^2 = 0,752$) ugyanakkor a beszélő és az ismétlés jelentős mértékben magyarázza a variabili-

tást. A 6. ábráról leolvasható tendenciák ismét meglehetősen változatos képet mutatnak magánhangzó-minőségenként, és a CoG-hez hasonlóan szintén a nyílt magánhangzók esetében mutattak eltérést a post hoc tesztek. Az /v/-t megelőző /p/ spektrális szórása az első (4837 Hz) és az utolsó szótag (5482 Hz) között mutat szignifikáns eltérést; az /a:/-környezetben pedig az első (4399 Hz) és a harmadik (4968 Hz), valamint az első és a negyedik (4970 Hz) szótag között. Az /u/ környezetében álló /p/ CoG-értéke minden más magánhangzó-kontextushoz képest szignifikánsan magasabb volt (az /a:/-s környezettől minden szótagban, az /i/-stől a 2–4. szótagokban, az /v/-stől az első két szótagban mutatott eltérést).

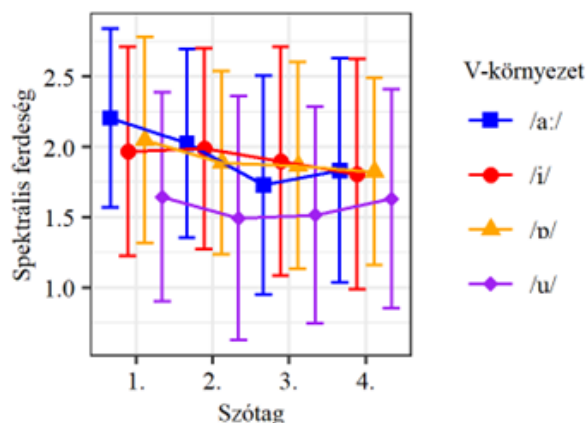


6. ábra. A spektrális szórás (SD) alakulása a szótagpozíció és a magánhangzó-környezet függvényében (átlag és szórás).

3.6. Spektrális ferdeség

Mint említettük, a spektrum eloszlásának szimmetriáját ragadja meg a ferdeség paramétere, és a /p/ esetben pozitív értéket várunk, hiszen a középponttól jobbra nyúlik el az eloszlás. A kérdés az, hogy az eltolódás mértéke összefüggést mutat-e a szótagpozícióval és a magánhangzó-minőséggel. A legjobb modellnek az bizonyult, ahol mindkét faktor szerepel, a beszélőnkénti és ismétlésenkénti random konstanssal. A két faktor interakciója nem volt szigni-

fikáns, míg egyesével mind a szótagpozíció ($F(3, 554, 40) = 6,700$, $p < 0,001$), mind a magánhangzó-minőség ($F(3, 200, 42) = 12,410$, $p < 0,001$) főhatása igen. A marginális hatásnagyság alacsony ($r_m^2 = 0,079$), a kondicionális pedig közepes ($r_c^2 = 0,542$), vagyis bizonyos mértékig magyarázzák az adatokat a beszélők és az ismétlések egyéni jellegzetességei, de valamilyen további, itt nem vizsgált tényezőnek is hatása van. A magánhangzó-minőségek ezúttal is változatos mintázatokat adnak (7. ábra). A post hoc tesztek alapján csak az /a:/-környezetben volt eltérés, mégpedig az első (2,20) és a harmadik (1,73) szótagbeli, valamint az első és a negyedik (1,83) szótagi ferdeségi értékek között. Ebben az esetben az alacsonyabb ferdeségérték a frázis vége felé fortiszabb mássalhangzóképzésre utal. A magánhangzó-környezet tekintetében az /u/ esetében itt is több szótagban előfordul a többitől való szignifikáns eltérés (/a:/ vs. /u/ és /v/ vs. /u/ az első két szótagban, /i/ vs. /u/ a 2. és 3. szótagban).

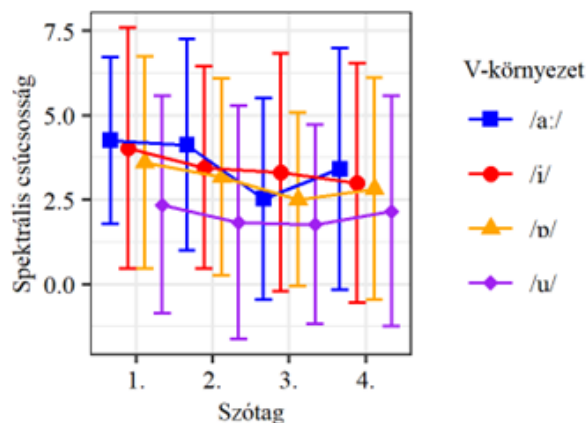


7. ábra. A spektrális ferdeség alakulása a szótagpozíció és a magánhangzó-környezet függvényében (átlag és szórás).

3.7. Spektrális csúcsosság

A csúcsosság értéke a spektrális eloszlás formáját jellemzi, mégpedig oly módon, hogy a pozitív csúcsosságérték jelzi azt, ha az eloszlás csúcsosabb, mint a normál eloszlás – a /p/ esetében erre számítottunk. A kérdés itt is az, hogy van-e

hatása a szótagpozíciónak és/vagy a kontextust képező magánhangzó minőségének a csúcosság értékére. A legjobb modellnek az bizonyult, ahol mindkét faktor szerepelt, beszélőnkénti és ismétléseknkénti random konstanssal, random meredekség nélkül. A két faktor interakciója nem bírt szignifikáns hatással, a két főhatást így külön elemeztük. A szótagpozíciónak ($F(3, 534,86) = 6,596$, $p < 0,001$) és a magánhangzó-minőségnek ($F(3, 194,02) = 9,104$, $p < 0,001$) a hatása is szignifikánsnak adódott. A marginális hatásnagyság ezúttal is elhanyagolható ($r_m^2 = 0,064$), a kondicionális pedig közepes ($r_c^2 = 0,575$), vagyis a vizsgált tényezők nem magyarázzák teljes körűen az adatokat. A post hoc tesztek közül ismét csak az /a:/-környezetben volt szignifikáns eltérés, mégpedig az első (4,26) és a harmadik (2,53) szótagban mért értékek között (8. ábra). Ebben az esetben az alacsonyabb csúcosságérték a frázis vége felé fortisabb mássalhangzóképzésre utal. A spektrális ferdeséghez hasonlóan ismét kisebb az /u/-s környezet többbitől való eltérése, de az /a:/-stől az 1. és 2., az /i/-stől a 3. szótagban szignifikáns eltérést mutat.



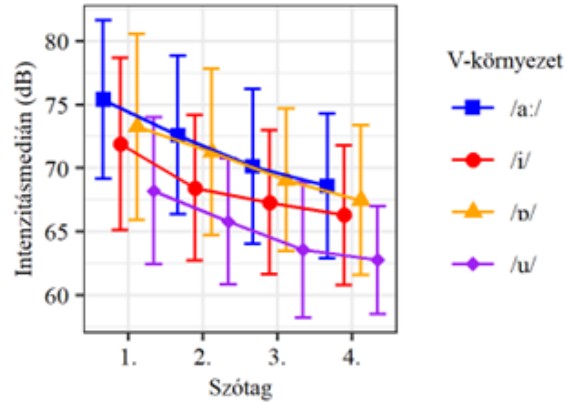
8. ábra. A spektrális csúcosság alakulása a szótagpozíció és a magánhangzó-környezet függvényében (átlag és szórás).

3.8. A zárfeloldás intenzitása

A zárfeloldás intenzitásának mediánjára vonatkozóan a legjobb modellnek az bizonyult, ahol mindkét faktor szerepel, a beszélőnkénti és ismétlésekenkénti random konstanssal, random meredekség nélkül. A két faktor interakciója ismét nem bizonyult szignifikánsan meghatározónak az eredményekben, míg a két faktor hatása külön-külön szignifikáns volt (a szótagpozícióra $F(3, 604, 86) = 213,582$, $p < 0,001$; a magánhangzó-minőségre $F(3, 201, 87) = 91,410$, $p < 0,001$). A marginális hatásnagyság alacsony ($r_m^2 = 0,256$), de elmondható, hogy az adatok mintegy negyedét magyarázzák az elemzett főhatások. A kondicionális hatásnagyság emellett jelentősen magasabb, azaz a vizsgált tényezők (beszélő, ismétlés) hatásai nagymértékben magyarázzák az adatokat ($r_c^2 = 0,860$). A post hoc tesztek alapján minden magánhangzó-minőség esetén szignifikáns eltérés van az első és a második, az első és a harmadik, az első és a negyedik szótagbeli /p/-k intenzitása között. Az /a:/-s és az /v/-s környezetben ezen túlmenően szignifikáns a különbség a második és a harmadik, valamint a második és negyedik szótag között is; az /i/ és az /u/ esetében pedig a második és a negyedik szótag között. Mindez azt jelenti, hogy az első szótag legintenzívebb zárfeloldásától kezdődően fokozatos intenzitáscsökkenés figyelhető meg a frázis utolsó (két) szótagjáig (9. ábra). Az azonos szótagban a magánhangzó-környezetek alapján ismét nagyobb különbségek mutatkoztak: az /a:/ minden, az /v/ több szótagban is szignifikáns eltérést mutatott az /i/-től és az /u/-tól. Az /i/-/u/ ugyancsak minden szótagban szignifikáns különbséget mutatott. Az /a:/-/v/ környezetek között csak az első szótagban kaptunk szignifikáns eltérést.

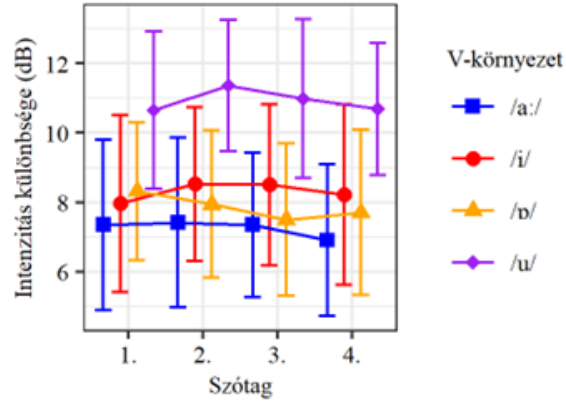
3.9. A mássalhangzó intenzitásának a követő magánhangzó legnagyobb intenzitásától való eltérése

A mássalhangzó intenzitásának a követő magánhangzó legnagyobb intenzitásától való eltérését a legjobban az a modell jellemezte, amelyben a magánhangzó-minőség és a szótagpozíció interakciója szerepelt, random konstanssal a beszélőre és az ismétlésre. Csak a magánhangzó-minőségre lett szignifikáns az eredmény ($F(3, 196, 12) = 63,787$, $p < 0,001$), a marginális hatásnagyság alacsony



9. ábra. A zárfeloldás intenzitásmédiánjának alakulása a szótagpozíció és a magánhangzó-környezet függvényében (átlag és szórás).

($r_m^2 = 0,264$), a kondicionális pedig közepes ($r_c^2 = 0,491$). A post hoc tesztek nem mutattak eltérést egyazon magánhangzós környezetben a szótagpozíciók között; a magánhangzó-környezetek között azonban az /u/ az összes szótagban magasabb értékeket adott a többi magánhangzóhoz képest. A többi magánhangzó környezetében mért értékek között nem volt eltérés (10. ábra).



10. ábra. A zárfeloldás intenzitásmédiánjának eltérése a követő magánhangzó intenzitásától a szótagpozíció és a magánhangzó-környezet függvényében (átlag és szórás).

4. Következtetések

A prozódiai erősítés az artikulációs gesztusok térbeli és/vagy időbeli expansziója, amely prozódiaileg kiemelkedő helyeken, például nagyobb prozódiai egységek szélein vagy hangsúlyos szótagokon történik. Ezeknek az artikulációs módosulásoknak akusztikai következményei is vannak/lehetnek, még ha nem is jellemezhető egyenes arányossággal az artikulációs gesztusbeli eltérés és az akusztikai szerkezetben jelentkező különbség közötti viszony (Stevens, 1989). A magyar mássalhangzókat tudomásunk szerint ez idáig nem elemezték a prozódiai erősítés szempontjából, így csak más nyelvekre vonatkozó vizsgálatok eredményei alapján feltételezhetjük, hogy a magyarban is találhatunk eltéréseket a frázisbeli pozíció (prozódiai tartomány eleje, vége; hangsúlyos helyzet vs. tartomány belseji, hangsúlytalan helyzet) függvényében a mássalhangzók valamely jellemzőiben. A szakirodalomban olvasható elemzések olyan nyelvek mássalhangzóit vizsgálták, amelyek a magyartól többek között a prozódiai szerkezet jellemzőiben (például a szóhangsúly helye, a szónál tágabb tartományhoz kötődő hangsúly funkciója és akusztikai manifestációja) és/vagy a mássalhangzókészletüket tekintve is eltérnek, így előzetesen nem tudtunk hipotéziseket megfogalmazni.

A vizsgálatban önálló megnyilatkozást alkotó $pV_1pV_1pV_1pV_1$ szerkezetű állszavak anyagában (amelyet korábban a magánhangzók prozódiai erősítésének elemzése céljából rögzítettük) elemeztük a bilabiális zöngétlen felpattanó mássalhangzókat. A független változók a szótagpozíció (a szótag sorszáma 1–4.) és (korábbi eredmények alapján) a szomszédos magánhangzók minősége (/i/, /u/, /ɒ/ és /a:/) voltak. Ezek függvényében elemeztük a /p/-ket a zárfelpattanás megléte/hiánya és száma, az időtartam, a VOT, a spektrális momentumok és az intenzitás (a zárfeloldás intenzitása, valamint ennek a követő magánhangzó legintenzívebb szakaszának intenzitásához viszonyított eltérése) szempontjából.

Általános, nyelvfüggetlen aerodinamikai elvek alapján nem vártuk a zárfelpattanás elmaradását az első szótagban, a további szótagokban azonban igen. (Megjegyezzük, hogy a frázisbeli helyzet függvényében a zárfelpattanás elmara-

dására vagy többszöröződésére vonatkozó elemzést a nemzetközi szakirodalomban sem találtunk.) Adataink alapján a zár-felpattanások megléte, illetve száma utal a prozódiai erősítésre az első és hangsúlyos szótagban.

A VOT esetében a szótagpozíció és a magánhangzó-minőség interakciója szignifikánsan meghatározó volt, és a modell nagymértékben magyarázta az adatokat. Az első szótagi helyzetben rövidebb VOT-t mértünk a szó belseji előfordulásokhoz képest. A post hoc tesztek alapján az /i/-s kontextusban az első és a második, valamint az első és a negyedik szótag között; az /u/-s kontextusban pedig az első szótagbeli /p/ VOT-értéke tért el az összes többi szótagbeli értéktől. Feltehető, hogy az /i/ és az /u/ a nyelvéllésfoka miatt alkalmasabb a magasabb nyomás hosszabb megtartására (másképp közelítve: kevésbé alkalmas a gyorsabb nyomáskiegyenlítésre), ezért a felső nyelvéllésű magánhangzók környezetében jobban hat a prominencia. A holland zöngés explozívák esetében ugyancsak VOT-rövidülést figyeltek meg a hangsúlyos szótagokban és a fráziskezdetben a frázis belsejéhez képest (Cho & McQueen, 2005).

A spektrum frekvenciaeloszlása kapcsán Tabain et al. (2016) az angol /k/-t vizsgálva a hangsúlyos szótagban alacsonyabb, Vicens (2010) pedig a grúz bilabialis zárhangokban szintén alacsonyabb átlagos frekvenciát mért a magasabb prozódiai szinteken megvalósuló prominens pozíciókban, így a prozódiai erősítés jele ezekben feltehetően az alacsonyabb CoG. Megjegyzendő ugyanakkor, hogy ilyen jellegű tendenciákat csak az alsó nyelvéllésű magánhangzók esetében találtunk a jelen anyagban, amelyek előtt a koartikulációs hatás (a zár-feloldás tágabb ajaknyílás felé tart, mint a felső nyelvéllésű magánhangzók előtt) magyarázza az alacsonyabb CoG-értékeket is. Az alsó nyelvéllésű magánhangzók esetében ráadásul az ajaknyílás mértéke kevésbé korlátozott (nagyobb a mozgástér), mint a zártak esetén, amelyeknek az ajaknyílását (és az ebből adódó, a megelőző más-salhangzóban mérhető CoG-értéket) korlátok közé szorítja a középső nyelvéllésű magánhangzó-minőségekbe való átcsapás lehetősége.

A fenti VOT-adatokkal összevetve a CoG-értékeket feltehetjük, hogy a prozódiai erősítés a felső nyelvéllésű magánhangzók szomszédságában inkább az időzítés paraméterében, az alsó nyelvéllésűakéban pedig inkább a spektrális

mintázatban fejeződik ki, ami természetesen összefügg a magánhangzók artikulációs jellemzőiből adódó aerodinamikai hatásokkal. Másrészt ugyanakkor a prozódiai határ közeledtének kulcsolására is szolgálhat az alsó nyelvvállású magánhangzók környezetében a magasabb spektrális súlypont, amit a frázisvégen (illetve az /a:/ esetben a frázis vége felé) láttunk; valamint a felső nyelvvállású magánhangzós kontextusban a hosszabb VOT. Az itt vizsgált egységek mindössze négy szótagosak voltak, de lehetséges, hogy hosszabb frázisok esetében nagyobb mértékben tér(né)nek el a frázis szélén és a frázis belsejében jelentkező mássalhangzók akusztikai jellemzői.

A zárfeloldás intenzitásadatait a korábbi, grúz nyelvre végzett vizsgálat adataihoz (Vicenik, 2010) nem tudjuk egy az egyben hasonlítani, mivel ott bilabiális zöngétlen zárhangból csak aspirált ejtésűt tartalmaz a mássalhangzókészlet. Mégis, a vizsgált tendenciák bizonyos fókig egybecsengenek a magyaréval, hiszen a magasabb rendű prozódiai helyzetnek (a magyar esetében az első szótagi és egyben hangsúlyos helyzetnek) erősítő hatása volt a zárfeloldás intenzitására.

A szakirodalom alapján feltehető, hogy a VOT értéke és a felpattanási zörej intenzitása összefügg (Sundara (2005) idézi Pickett (1999) felvetését), mivel a szájüregi zár időtartamának aerodinamikai következményéről van szó. A magasabb VOT-értékkel jellemezhető zárhangok hosszabb záridőtartamot is mutatnak, ennél fogva a szájüregi nyomás felépülésének intenzívebb felpattanási zöreje a következménye. Mindez természetesen ismét a nagyobb VOT-nak a feszeőbb ejtéssel való összefüggésére is utal. A mi adataink ugyanakkor ezt az összefüggést nem mutatják, mert bár az intenzitásmedián a frázisban folyamatosan csökken, a VOT (magánhangzónként változó mértékben és léptékben) inkább növekszik. Az egyes magánhangzók környezetében a frázis vége felé a CoG ugyancsak magasabb, a ferdeség és a csúcosság pedig alacsonyabb, ami szintén fortisabb ejtésre utal. Így tehát az intenzitás csökkenése mellett más akusztikai paraméterek inkább erősítést mutatnak a frázis vége felé.

Az intenzitásadatok csökkenési tendenciája lehet annak a sokak által általánosnak tartott és fiziológiai alapokra visszavezetett folyamatnak a része, hogy a megnyilatkozás kezdetétől a végéig fokozatos gyengülés (ereszkedés) történik,

amely megnyilvánul az artikulációs gesztusokban, a frekvenciában, az intenzitásban és az időzítésben egyaránt, mind szegmentális, mind szupraszegmentális tekintetben (vö. pl. Vayra & Fowler, 1992). Mindez a megnyilatkozásban előre haladván a szubglottális nyomás fokozatos csökkenésével függhet össze, amely a (beszéd)légzés fiziológiai sajátosságaiból ered (Cohen et al., 1982). Ez a fiziológiai magyarázat ugyanakkor azt sugallja, hogy a beszédműködés olyan következményéről van szó, amely nem áll a beszélő kontrollja alatt. Ezzel szemben számos példát ismerünk arra, hogy a légzési és a gégeizmok a beszélő kontrollálásával működnek (ilyen egyebek között a magyar kérdő intonáció, azaz a frázisvégi frekvenciaemelkedés és -csökkenés). Ennélfogva több szerző is azt a magyarázatot támogatja, hogy a frázis végéhez közeledve jelentkező artikulációs-akusztikai módosulások (például a frázisvégi nyúlás) aktív folyamat eredményei, és az a szerepük, hogy ráirányítsák a hallgató figyelmét a következő prozódiai határra, ezzel is kulcsot adva a beszédfeldolgozáshoz (vö. pl. Byrd, 2000; Cho, 2006; Cho et al., 2011). Ezt az elképzelést támogatja az a tény is, hogy például az említett frázisvégi nyújtás esetében – habár univerzális jelenségnek tűnik – eltérések találhatók a nyelvek között a mértékét és a kiterjedését tekintve (vö. Kohári, 2018 összefoglalását).

Amint láttuk, az alkalmazott magánhangzó-környezetek függvényében is kirajzolódtak mintázatok, amelyek további kérdéseket is felvetnek. Az /u/ környezetében majdnem minden paraméter, az /i/-kontextusban a VOT és az intenzitás tekintetében tértek el a /p/ akusztikai tulajdonságai a többi környezetben mértéktől. Mivel a /p/ kisebb koartikulációs hatással bír a környező magánhangzókra, ezért e magánhangzók szomszédságában a megemelt nyelvhelyzetnek nagyobb mértékű hatása lehet a szájüreg méretére, így a szupraglottális nyomás magasabb lehet, mint az alsóbb nyelvállásfokok esetében. Ez a hatás az /u/ hátul képzett volta következtében (az /i/-éhez képest is) kisebb üregméret miatt tovább fokozódhat. Ennek a nyomásviszonynak köszönhető feltehetőleg a Bernoulli-hatásnak a felpattanások számát megnövelő hatása. Ugyancsak emiatt lassabban egyenlítődik ki a zárfeloldás után a külső légnyomás és a belső (intraorális) nyomás, ezért a VOT is hosszabb lehet. A zár mögött közvetlenül

felgyűlő levegő mennyisége és nyomása pedig az intenzitásértékekre volt hatással. A spektrális momentumokban kisebb, de (az első kettőben erősebben) meg-lévő eltérés hátterében ezen okok együttesen játszhatnak szerepet. Mindennek vizsgálata további, ezt a kérdést célzó kutatások feladata, okai között szerepelhet a magánhangzó célkonfigurációjához közeli állapotban megtartandó nyelvforma.

Elemzésünk egy más céllal felvett hanganyagon természetesen korlátozottan adhatott választ a magyar mássalhangzók prozódiai erősítésével kapcsolatban felmerülő tudományos kérdésekre. A rendelkezésre álló szekvenciák viszonylag rövidek voltak, és mindösszesen egy mássalhangzó elemzését tették lehetővé, ráadásul e mássalhangzó zöngétlensége miatt bizonyos szempontokat (például a teljes mássalhangzó-időtartamot az első szótagban) nem tudtunk vizsgálni. Mindemellett a szakirodalom (Tabain et al., 2016) alapján azt is számításba kell vennünk, hogy a /p/ az ilyen szempontú vizsgálatra az artikulációs sajátos-ságaiból adódóan nem ideális választás (még ha az anyag eredeti vizsgálati célja szerint az volt is), mivel az akadály mögötti teljes üreg gerjesztődik a zár-felpattanás során létrejövő zörej által. Másrészt a mássalhangzó-észlelés szempontjából szerepet játszhat az a tény is, hogy a bilabiális zár vizuálisan egyértelműsíti a képzési jellemzőket (az ajakpozícióból és az izomtevékenység látható jeleiből a képzés helye, módja, a zöngésség és az időtartam is egyértelműen azonosítható), így nincs szükség akusztikai erősítésre. Mindezen érvényességi korlátok és kérdések tudatában mégis arra utalnak az elemzési eredmények, hogy a prozódiai erősítés jelenségét a magyar mássalhangzók tekintetében is érdemes (tovább) vizsgálni, erre a célra összeállított kísérleti anyagon, hiszen bizonyos, más nyelvekre is jellemző tendenciák itt is megmutatkoztak.

Köszönetnyilvánítás

Köszönjük az adatfelvételben és -feldolgozásban nyújtott technikai segítséget Krepsz Valériának. A tanulmány a Bolyai János Kutatási Ösztöndíj támogatásával (D. A., Cs. T. G.), az Innovációs és Technológiai Minisztérium ÚNKP-21-3 (J. K.) és ÚNKP-21-5 (D. A., Cs. T. G.) kódszámú Új Nemzeti

Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott támogatásával, valamint az Innovációs és Technológiai Minisztérium Nemzeti Kutatási, Fejlesztési és Innovációs Alapból nyújtott támogatásával (ELTE TKP2020-IKA-06 sz. támogatói okirat) és az FK128814-es pályázat számára nyújtott támogatásával készült. Köszönjük az anonim lektorok hasznos megjegyzéseit, javaslatait.

Hivatkozások

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723. doi:10.1109/TAC.1974.1100705.
- van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in dutch initial plosives: the role of prevoicing. *Journal of Phonetics*, 32, 455–491.
- Bartoń, K. (2022). Mumin: Multi-model inference. r package version 1.46.0. <https://cran.r-project.org/web/packages/MuMIn/index.html>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer. version 6.1.14. <http://www.praat.org/>.
- Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57, 3–16.
- Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English. *Journal of the Acoustical Society of America*, 117, 3867–3878.
- Cho, T. (2006). Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. *Laboratory Phonology*, 8, 519–548.

- Cho, T. (2015). Language effects on timing at the segmental and suprasegmental levels. In M. Redford (Ed.), *The handbook of speech production* (pp. 505–529). Hoboken, NJ: Wiley-Blackwell.
- Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, 39, 344–361.
- Cho, T., & McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33, 121–157.
- Cho, T., McQueen, J., & Cox, A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35, 210–243.
- Chodroff, E., & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *Journal of the Acoustical Society of America*, 136, 2762–2772.
- Cohen, A., Collier, R., & t' Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch? *Phonetica*, 39, 254–273.
- Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.
- Deme, A., Bartók, M., Csapó, T. G., Grácsi, T. E., Juhász, K., & Markó, A. (2021). Acoustic and articulatory vowel variation as quality shift and increased variance in anticipatory and carryover vowel-to-vowel coarticulation. In M. Tiede, D. H. Whalen, & V. Gracco (Eds.), *Proceedings of the 12th International Seminar on Speech Production* (pp. 32–35). New Haven, CT, Egyesült Államok: Haskins Press.
- Deme, A., Bartók, M., Csapó, T. G., Grácsi, T. E., Juhász, K., & Markó, A. (2022). A magánhangzók centralizációja és produkciós homogenitása az előrefelé és hátrafelé ható magánhangzók közti koartikulációban – artikulációs

- és akusztikai adatok. *Általános Nyelvészeti Tanulmányok XXXIV*, (pp. 15–49).
- Deme, A., Bartók, M., Csapó, T. G., Grácz, T. E., Juhász, K., & Markó, A. (kézirat). The effect of pitch-accent on the acoustic and articulatory variability of vowels: a real-word EMA study.
- Deme, A., Bartók, M., Grácz, T. E., Csapó, T. G., Juhász, K., & Markó, A. (megj.). A koartikulációs ellenállás és agresszió hangsúlyos helyzetű magánhangzókban: artikulációs és akusztikai adatok. *Nyelvtudományi Közlemények*, 118. doi:10.15776/NyK/2022.118.
- Deme, A., Bartók, M., Grácz, T. E., Csapó, T. G., & Markó, A. (2019a). A mondathangsúly hatása a magánhangzók megvalósulásának változatosságára. *Nyelvtudományi Közlemények*, 115, 199–232.
- Deme, A., Bartók, M., Grácz, T. E., Csapó, T. G., & Markó, A. (2019b). V-to-V coarticulation induced acoustic and articulatory variability of vowels: The effect of pitch-accent. In G. Kubin, & Z. Kačič (Eds.), *The 20th Annual Conference of the International Speech Communication Association (Interspeech 2019)* (pp. 3317–3321). Graz, Austria: International Speech Communication Association (ISCA).
- Edwards, J. E., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, 89, 369–382.
- Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, 84, 115–123.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101, 3728–3740.

- Gósy, M. (1997). Semleges magánhangzók a magyar beszédben. *Magyar Nyelvőr*, 121, 9–19.
- Gósy, M. (2000). A [p, t, k] mássalhangzók zöngékezdesi ideje. *Magyar Nyelvőr*, 124, 195–204.
- Gósy, M., & Beke, A. (2010). Magánhangzó-időtartamok a spontán beszédben. *Magyar Nyelvőr*, 134, 140–165.
- Gósy, M., & Krepsz, V. (2016). A magánhangzó nyúlása a szótagszám és a mondatpozíció függvényében. *Beszéd kutatás*, 2016, 59–88.
- Gósy, M., & Ringen, C. O. (2009). Everything you always wanted to know about VOT in Hungarian. (előadás.). In *International Conference on the Structure of Hungarian 2009*. Budapest. URL: https://www.researchgate.net/publication/253974237_Everything_You_Always_Wanted_to_Know_About_VOT_in_Hungarian.
- Grácz, T. E., & Kohári, A. (2012). A zöngékezdesi idő egy módszertani kérdés függvényében. In A. Markó (Ed.), *Beszéd tudomány. Az anyanyelv-elsajátítástól a zöngékezdesi időig* (pp. 228–248). Budapest: ELTE Bölcsészettudományi Kar és MTA Nyelvtudományi Intézet.
- Harrison, E., & Pius, R. (2021). R for health data science. (electronic version of the healthyr book published by chapman & hall/crc. https://argoshare.is.ed.ac.uk/healthyr_book/).
- Ip, M. H. K., & Cutler, A. (2022). Juncture prosody across languages: Similar production but dissimilar perception. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 13, 1–49.
- de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97, 491–504.

- Kassai, I. (1979). *Időtartam és kvantitás a magyar nyelvben. Nyelvtudományi Értekezések 112*. Budapest: Akadémiai Kiadó.
- Keating, P. A., Cho, T., Fougeron, C., & Hsu, C. (2004). Domain-initial strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Laboratory phonology, Vol. IV: Phonetic interpretation* (pp. 145–163). Cambridge, UK: Cambridge University Press.
- Kohári, A. (2018). *Időzítési mintázatok a magyar beszédben*. Budapest: ELTE Eötvös Kiadó.
- Kohári, A., & Markó, A. (2015). A megnyilatkozás végének jelzése: temporális szerkezet és zöngeminőség felolvasásban. *Beszédkutató*, 23, 35–52.
- Kovács, M. (2001). *Tendenciák és szabályszerűségek a magánhangzó-időtartamok produkciójában és percepciójában*. Ph.D. thesis Debrecen. URL: <https://dea.lib.unideb.hu/dea/bitstream/handle/2437/79756/ertekezes.pdf?sequence=6&isAllowed=y>.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. doi:10.18637/jss.v082.i13.
- Lenth, R. (2020). Emmeans: Estimated marginal means, aka least-squares means. <https://CRAN.R-project.org/package=emmeans>. R package version 1.4.4.
- Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H & H theory. In W. Hardcastle, & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Dordrecht: Kluwer Academic Publishers.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384–422.
- Mády, K., Bombien, L., & Reichel, U. (2008). Is Hungarian losing the vowel quantity distinction? In R. Sock, S. Fuchs, & Y. Laprie (Eds.), *Proceedings of 8th*

- International Seminar on Speech Production* (pp. 449–453). Strasbourg: INRIA. URL: https://phon.nytud.hu/mady/pub/mady_bombien_reichel.pdf.
- Mády, K., Reichel, U., & Szalontai, Á. (2017). A prozódiai prominencia (nem-)jelölése a németben és a magyarban. In *Általános Nyelvészeti Tanulmányok XXIX* (pp. 77–98).
- Magdics, K. (1965). *A magyar beszédhangok akusztikai szerkezete. Nyelvtudományi Értekezések 49*. Budapest: Akadémiai Kiadó.
- Magdics, K. (1966). A magyar beszédhangok időtartama. *Nyelvtudományi Közlemények*, 68, 125–139.
- Markó, A., Bartók, M., Csapó, T. G., Deme, A., & Grácsi, T. E. (2019). The effect of focal accent on vowels in Hungarian: articulatory and acoustic data. In *Proceedings of the 19th ICPhS* (pp. 2715–2719).
- Markó, A., Bartók, M., Grácsi, T. E., Deme, A., & Csapó, T. G. (2018). Prominence effects on Hungarian vowels: A pilot study. In *9th International Conference on Speech Prosody* (pp. 868–872).
- Markó, A., Juhász, K., Bartók, M., Csapó, T. G., Grácsi, T. E., & Deme, A. (2022). Magyar magánhangzók artikulációs és akusztikai jellemzői a fonetikai pozíció függvényében álszavakban. In *Általános Nyelvészeti Tanulmányok XXXIV* (pp. 51–79).
- Neuberger, T., & Beke, A. (2017). A zárfelpattanás spektrális jegyei a hosszúsági oppozíció függvényében. *Beszéd kutatás*, 25, 7–23.
- Pickett, J. M. (1999). *The Acoustics of Speech Communication*. Boston: Allyn and Bacon.
- Pierrehumbert, J. (1999). Prosody and intonation. In R. A. Wilson, & F. Keil (Eds.), *The MIT encyclopedia of cognitive sciences* (pp. 479–482). Cambridge, MA: MIT Press.

- Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives*. *Cambridge NA Report NA2009/06* volume 2009. Cambridge: University of Cambridge.
- R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.
- Reichel, U. D. (2012). Perma and balloon: Tools for string alignment and text processing. In *Proceedings of Interspeech 2012* (p. 346).
- Schiel, F. (1999). Automatic phonetic transcription of nonprompted speech. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 607–610).
- Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D. H. Spieler, & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4–31). PsychologyPress.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3–45.
- Sundara, M. (2005). Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *Journal of the Acoustical Society of America*, 118, 1026–1037.
- Szalontai, Á., Wagner, P., Mády, K., & Windmann, A. (2016). Teasing apart lexical stress and sentence accent in Hungarian and German. In C. Draxler, & F. Kleber (Eds.), *Tagungsband 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum (P&P 12)* (pp. 216–219). München: Ludwig-Maximilians-Universität.
- Tabain, M., Breen, G., Butcher, A., Jukes, A., & Beare, R. (2016). Stress effects on stop bursts in five languages. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7, 1–23.

- Vayra, M., & Fowler, C. (1992). Declination of supralaryngeal gestures in spoken italian. *Phonetica*, 49, 48–60.
- Vicenik, C. (2010). An acoustic study of Georgian stop consonants. *Journal of the International Phonetic Association*, 40, 59–92.
- White, L. (2002). *English Speech Timing: A Domain and Locus Approach*. Edinburgh: The University of Edinburgh. URL: <https://era.ed.ac.uk/handle/1842/23256> PhD Thesis.
- White, L., Benavides-Valera, S., & Mády, K. (2020). Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues? *Journal of Phonetics*, 81. URL: <https://doi.org/10.1016/j.wocn.2020.100982>.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Az önjavítási jelenségek magyar nyelvű spontán, baráti társalgásokban

Tar Cintia¹

¹*SZTE BTK Általános Nyelvészeti Tanszék*

Abstract

In spontaneous, everyday conversations, we can observe regularities and patterns that are repeatedly produced by the speakers during the conversation (Iványi, 2001). Such regularities can also be observed in the repair mechanism. The study seeks to answer whether the ten repair operations described by Schegloff (2013), namely, replacing, inserting, parenthesizing, deleting, aborting, recycling, searching, sequence-jumping, reformatting, and reordering, appear in Hungarian spontaneous conversations, and if so, how; and whether or not variations can be observed in each operation compared to Schegloff's definitions. In addition, the study also examines the frequency of each repair operation. I researched this using spontaneous, casual everyday conversations. The study revealed that replacing, inserting, parenthesizing, deleting, aborting, recycling, searching and sequence-jumping in Hungarian correspond to the English observations. There were, however, some discrepancies, which are the following: reformatting did not appear in my corpus. Furthermore, in the English *delicate* type of searching, the speaker deliberately delays a potentially offensive term (Schegloff, 2013), whereas in my example, the speaker searches for a word that is appropriate to the situation and does not produce an offensive, unpleasant term. In sequence-jumping, the current speaker interrupts the turn-constructional-unit by producing an inappropriate unit, thereby interrupting the action and exiting the sequence (Schegloff, 2013). In contrast, in my study, the turn of the current speaker is interrupted by his partners, the original speaker produces a response reaction to this, and thus he himself abandons the previous turn and exits the sequence. In the case of the reordering, it was suggested that the reordered form is an inherent recycling. However, in this case, the inserting required a change in word order, since the inserted element was in a focal position, and therefore the affix had to be reordered after the verb. The frequency order of repair operations is as follows: recycling, searching, replacing, inserting, aborting, parenthesizing, sequence-jumping, deleting, and reordering.

Email address: tarcinti05@gmail.com (Tar Cintia)

1. Bevezetés

1.1. A kutatás elméleti háttere

A konverzációelemzés empirikus módon, hangfelvételek és videóanyagok rögzítésének segítségével vizsgálja a társalgások – különös tekintettel a spontán társalgásokra – struktúráját, hogy hogyan alakítják a beszélgetőpartnerek a beszélgetéseket. A hangsúly itt az alakításon van, hiszen a konverzációelemzés nem kész-egészként tekint a társalgásokra, hanem annak a folyamatát kutatja, hogy a beszélgetőpartnerek hogyan hozzák létre lépésről lépésre a beszélgetést, és ezáltal produkálják újra és újra a szituációkat (Iványi, 2001). A konverzációelemzés szerint a javítási jelenségek a társalgások olyan természetes velejárói, amelyek megszakítják a beszéd cselekvés folytonosságát azért, hogy kezeljenek egy, a produkció, az észlelés vagy a megértés során felmerülő lehetséges hibát, problémát (Schegloff et al., 1977). A javítás biztosítja, hogy ezeknél a felmerülő problémáknál a beszélgetés ne fagyjon meg, hanem zavartanul folytatódjon tovább, hogy az interszubjektivitást, vagyis a valóság közös értelmezését fentartsa vagy helyreállítsa. Azokat a pontokat, ahol a beszéd során fellépnek ezek a problémák, a konverzációelemzés problémaforrásnak (*trouble-source*) vagy javítandónak (*repairable*) nevezi (Schegloff et al., 1977; Kitzinger, 2013). Habár a javítás a beszéd mint cselekvés természetes velejárója, a különböző nyelvekben különbözőképpen működhet, hiszen befolyásolja a javításra alkalmas lehetséges nyelvtani források készlete (Schegloff, 2007; Kitzinger, 2013).

A javítási folyamatban megkülönböztetünk javításkezdeményezést (*repair initiation*) és a javítás eredményét (*repair outcome*) (Schegloff et al., 1977). A javítást kezdeményezheti, illetve végrehajthatja a problémaforrást produkáló beszélő és a beszélgetőpartner is. Amikor a problémaforrást produkáló beszélő hajtja végre a javítás kezdeményezését és magát a javítást is, önkezdeményezett önjavításról beszélünk (Schegloff et al., 1977). Az önkezdeményezett önjavítás esetén az aktuális beszélő megszakítja a beszéd folytonosságát azon a ponton, ahol javítandónak véli a mondanivalóját, és valamilyen javítási jelenséggel javítja azt.

Az önkezdemenyezett önjavítások típusait Schegloff (2013) írja le összefoglalóan. Eszerint tíz különböző, úgynevezett javítási műveletet (*repair operation*) különböztethetünk meg. Ez a tíz javítási művelet a csere (*replacing*), a beszúrás (*inserting*), a zárójelezés (*parenthesizing*), a kihagyás (*deleting*), a törlés (*aborting*), az ismétlés (*recycling*), a keresés (*searching*), a szekvenciaugrás (*sequence-jumping*), az újraformázás (*reformatting*) és az újrendezés (*reordering*) (Schegloff, 2013).

A csere esetén a beszélő helyettesít egy már kimondott szót vagy egy nagyobb szerkezeti egységet egy másik elemmel, miközben megtartja azt az érzetet, hogy ugyanazon megnyilatkozáshoz tartozik a kicserélt rész (Schegloff, 2007). A javításban részt vevő kifejezések (*replacement terms*) lehetnek például antonimák, szinonimák vagy alternatív formulák (pl.: *told-said*) (Kitzinger, 2013). Szavak cseréje esetén a beszélő ugyanolyan szintaktikai szerepet betöltő szóra cseréli a problémaforrást alkotó szegmentumot, így a kicserélt beszédszakasz maga is felveszi azokat a toldalékokat, amelyek a cserélendő szón is megvannak.

A beszúrás felfüggeszti a folyamatban lévő fordulókonstruktív egységet (*turn-constructive unit*, a továbbiakban TCU), visszaugrik egy korábbi pontra, beszúr valamilyen egyszerű vagy összetett elemet, és megismétli a korábban elhangzott részt (Wilkinson & Weatherall, 2011; Németh, 2014). A beszúráshoz tehát mindig társul ismétlés. Az ismétlést ebben az esetben a konverzációelemzés nem elemzi külön javítási műveletként, mert a beszúráshoz inherens módon kapcsolódik a korábbi elem ismétlése (Wilkinson & Weatherall, 2011). Kitzinger (2013) példái alapján megállapítható, hogy a forduló általában beszúrás nélkül is grammatikus és érthető lenne, de a beszélő a beszúrt elem nélkül problémásnak tekinti a fordulót. A beszúrás fő funkciója tehát nem egy nyilvánvaló hiba javítása, hanem egy, az aktuális beszélő által problematikusnak vélt szakasz módosítása, finomra hangolása (*fine-tuning*).

A zárójelezés mint javítási művelet produkálásakor a beszélő „zárójelesen” beszúr egy megjegyzést, hozzáfűznivalót a fordulóba, amelyet fontosnak vél közölni a beszélgetőpartnereivel. A hangsúly itt a beszúráson van, hiszen ez a javítási művelet a beszúrás nagyobb terjedelmű formájának tekinthető. A be-

szúrás műveletével ellentétben azonban a beszélő egy egész szintaktikai szerkezetet ékel be a TCU-ba. Ez a jelenség leginkább a történetmesélés sajátossága, melynek célja, hogy a beszélő meggyőződjön arról, hogy beszélgetőpartnerei tudják követni a történetét, vagyis, hogy fenntartsa vagy helyreállítsa az interszubjektivitást mint a társas interakció egyik alapkövét. Ez a jelenség leginkább TCU-ba beágyazódva jelenik meg (Lipóczi, 2020).

A zárójelezéshez – ahogyan a beszúráshoz is – társulhat ismétlés. A beszélő gyakran megismétli a zárójeles megjegyzés előtti egységet (Kitzinger, 2013). Ez azonban a beszúrással ellentétben nem inherens módon kapcsolódik a javításhoz, nem törvényszerű, hanem inkább a mondanivaló követhetőségét biztosítja keretező funkciót ellátva. Ilyenkor az ismétlés nem tekinthető javítási műveletnek. A zárójelezés gyakran akkor jelenik meg, amikor a beszélő nem biztos abban, hogy elmesélte-e már az adott történetet, és megerősítést vár. Ilyen például a *nem tudom, meséltem-e már* formula, amely gyakran kérdés formájában jelenik meg: *vagy ezt meséltem már nektek?* (Lipóczi, 2020). Előfordulhat az is, hogy egy, a történethez nem szorosan kapcsolódó, de háttérinformációként szolgáló megjegyzést szúr be a beszélő annak érdekében, hogy fenntartsa vagy helyreállítsa a valóság közös értelmezését (Lipóczi, 2020).

A kihagyás esetén a beszélő töröl egy vagy több elemet, amelyet már teljesen vagy részlegesen produkált. A cserével ellentétben itt a beszélő úgy javít, hogy nem egy másik elemmel helyettesíti, hanem teljesen kihagyja, úgy tesz, mintha el sem hangzott volna az adott elem (Schegloff, 2013). A javítandó elem részleges kimondása alatt azt értem, hogy a beszélő már majdnem kimondta a szót, kifejezést, de még nem ment teljesen végbe az artikuláció. Ez a részleges kimondás együtt jár a szó elvágásával (*cut-off*), amely a legtöbb esetben a javításkezdeményezést jelzi. Ez az ún. megszakításos javításkezdeményezés (Németh, 2014). A szó elvágása mellett inherens ismétlés is kapcsolódhat a kihagyáshoz (Schegloff, 2013).

A törlés kapcsolatban áll a kihagyással annak alapján, hogy mindkét esetben egy korábbi egység elhagyása történik. Törlés alkalmazásakor azonban a beszélő az egész korábbi egységet, legtöbbször az egész TCU-t elhagyja (Scheg-

loff, 2013). Mivel egy fordulókonstruktív egységgel minimum egy cselekvést végre kell hajtania a beszélőnek (Iványi, 2001), így magát a cselekvés korábbi verzióját törli, és azt fogalmazza meg másképpen, vagyis kicseréli egy másik megfogalmazással. Ennek alapján a törlés sokkal szorosabb kapcsolatban áll a cserével, mint a kihagyással.

Ismétlés végrehajtásakor a beszélő újraprodukál egy beszédrészt, amely rendszerint közvetlenül azelőtt hangzott el. Ez a beszédrész lehet lexikális vagy kvázilexikális elem (Rieger, 2003; Németh, 2014), de lehet egy több összetevős, TCU-nál kisebb egység is (Schegloff, 2013). Fontos, hogy a konverzációelemzés csak akkor tekinti az ismétlést javítási műveletnek, ha az valamilyen produkcióbeli, észlelésbeli vagy megértésbeli problémát kezel (Kitzinger, 2013; Németh, 2014). Ennek értelmében tehát nem tekinthető javításnak a nyomatékosító funkciót betöltő ismétlés (pl.: *nagyon nagyon*).

Az ismétlés mint javítási művelet gyakran időnyerési stratégiaként funkcionál. Ilyenkor a beszélőnek nagyobb mentális erőfeszítést kell tennie, hogy előhívja a soron következő egységet, az ismétléssel pedig időt nyer, amíg sikerül előhívnia a keresett szót, kifejezést a mentális lexikonból (Gósy, 2005). Amikor a beszélő ismétlésekkel, szünetekkel és egyéb megakadásjelenségekkel késlelteti a soron következő kifejezést, szókeresésről beszélünk. Ahogy a világ nyelveiben, megfigyelhető a magyarban is, hogy ezek a késleltető elemek általában tartalmas szavak előtt jelennek meg, különösen olyanok előtt, amelyeket nem mindennap használ a beszélő. Megfigyelhető az is, hogy sokkal gyakoribb a funkciószó-ismétlés, mint a tartalmas szó ismétlése (Gyarmathy, 2009). Ennek oka, hogy a tartalmas szavak nyitott, míg a funkciószók zárt szóosztályba tartoznak. Emellett pedig a tartalmas szavak között sokkal gyakoribbak a ritkán használatos kifejezések, ami még tovább növeli az előhívás nehézségének mértékét (Fox et al., 2009). Mivel a magyarban a fráziskezdő elemek gyakran funkciószók, amelyek közvetlenül megelőzik a tartalmas szót, így a beszélők, hogy késleltessék a tartalmas szó produkálását, megismételhetik a tartalmas szó előtti elemet, vagyis a funkciószót (Németh, 2020). Mindez azt támasztja alá, hogy az ismétlés, különösen a funkciószók ismétlése időnyerésre is szolgálhat.

Az ismétlés másik lehetséges funkciója az átfedéskezelés (Schegloff, 1987). Az előbbivel ellentétben itt nem egy kifejezés előhívása, hanem az egyszerre beszélés, vagyis az átfedés (*overlap*) jelenti a problémát. Ilyenkor az aktuális beszélő magánál akarja tartani a szót, de belép egy másik résztvevő, ami veszélyezteteti az aktuális beszélő fordulójának végrehajtását. Schegloff (2013) ezt a jelenséget *surviving turn*-nek nevezi, amely az ismétlést végrehajtó beszélő fordulójának „túlélését”, „életben tartását” biztosítja (Schegloff, 2013).

A keresés szorosan összefügg az ismétléssel. A korai konverzációelemző tanulmányok kizárólag a szókeresésre korlátozták a keresést (Schegloff, 2013), ma már azonban egy nagyobb, szélesebb skálára terjesztették ki. A keresésnek két altípusát különböztetjük meg (Schegloff, 2013). Az egyik altípusba azok a jelenségek tartoznak, amikor valaminek a pontos megnevezése, például egy bizonyos személy neve, egy helyszín, egy tárgy pontos megnevezése vagy akár egy szak-kifejezés előhívása okozhatja a problémát. Ez az ún. *precise* típusú keresés (Schegloff, 2013; Kitzinger, 2013). A *precise* típusú keresést hezitációs jelenségek és esetleges ismétlések kísérhetik.

A keresés másik fajtája az ún. *delicate* típusú keresés. A *delicate* típusú keresés nem szókeresés, tehát nem egy pontos kifejezés előhívása okozza a problémát. A beszélő tudatosan késleltet egy kényes, többnyire tabu kifejezést, amelyet valamilyen oknál fogva nem illendő kimondani az adott kontextusban (Schegloff, 2013; Kitzinger, 2013). Ezek lehetnek sértő megjegyzések, betegségekkel, szexualitással kapcsolatos kifejezések, trágár kifejezések és a politikailag inkorrekt kifejezések (lásd még: Ullmann (1967) tabu-tipológiája). A beszélő ebben az esetben tisztában van azzal, hogy kényes kifejezést fog mondani, ezért hezitációs jelenségeket produkál, mint például a szünetek és az *öö*-zés, amellyel kifejezi a kifejezéshez való viszonyát, vagyis, hogy kényesnek, potenciálisan sértőnek tartja (Lerner, 2013).

A szekvenciaugrás több szempontból hasonló a törléshez, mégis külön kell kezelnünk. Amíg a törlés megtartja az eredeti cselekvést másképpen megfogalmazva, addig a szekvenciaugrás úgy szakítja meg a folyamatban lévő TCU-t, hogy egy egészen más, a fordulóhoz egyáltalán nem odaillő TCU-ba kezd a be-

szélő, vagyis teljesen elhagyja a folyamatban lévő cselekvést, és ezzel megszakítja az előző szekvenciát (Schegloff, 2013).

Az újrendezés egy olyan művelet, amely ugyancsak a törléssel áll kapcsolatban. Az újrendezést az újraformázáshoz hasonlóan grammatikai dimenzióban lehet megragadni. A beszélő a folyamatban lévő TCU-jának a szórendjét nem ítéli grammatikusnak, ezért újraprodukálja, azaz újrendezi az egész TCU-t vagy annak egy részét az általa helyesnek vélt szórenddel (Schegloff, 2013). Ez főképp olyan nyelvekre jellemző, amelyeknek kötött szórendjük van, mint például az angol vagy a német (Kitzinger, 2013).

Az újraformázás szintén hasonlóképpen működik, mint a törlés. Az újraformázás azonban a törléssel ellentétben csupán grammatikai dimenziókban értenődő (Schegloff, 2013), például amikor a beszélő egy kijelentést kérdéssé formáz át.

1.2. A kutatás motivációja és célja

A javítási jelenségek angol nyelvben való vizsgálata a konverzációelemzés elméleti keretén belül igen széleskörű szakirodalommal rendelkezik. A magyar nyelvre vonatkozóan a pszicholingvisztika (pl.: Gósy, 2004Gósy, 2004) és a fonetika (pl.: Horváth, 2014) gyakran kutatott területe, azonban a konverzációelemzés elméleti keretén belül még nem született olyan tanulmány, amely a Schegloff (2013) által jellemzett tíz, önkezdemenyezett önjavító javítási művelet mindegyikét vizsgálta volna a magyar nyelvben. Eddig ebben az elméleti keretben még csak négyet, a cserét, a beszúrás, az ismétlést és a törlést vizsgálták (Németh, 2017).

Kutatásom célja a Schegloff (2013) által leírt tíz javítási művelet, vagyis a csere, a beszúrás, a zárójelezés, a kihagyás, a törlés, az ismétlés, a keresés, a szekvenciaugrás, az újraformázás és az újrendezés magyar nyelvű kvalitatív és kvantitatív vizsgálata a konverzációelemzés elméleti keretén belül.

A kutatási kérdések a következők:

- (i) Megjelenik-e mind a tíz javítási művelet a magyar nyelvű korpuszokban?

- (ii) Ha megjelenik mind a tíz művelet, milyen azonosságok és különbségek figyelhetők meg az angol nyelvű előfordulásokhoz képest?
- (iii) Milyen gyakoriságot mutatnak az egyes javítási műveletek?

A hipotézisek szerint:

- (i) A magyar nyelvű saját korpuszban előfordul mind a tíz, Schegloff (2013) által leírt javítási művelet.
- (ii) A javítási műveletek többnyire azonosságokat mutatnak. Az azonosságok a javítási jelenségek természetes, univerzális jellegét erősítik (Fox et al., 2009), azonban azt is szem előtt kell tartanunk, hogy a javítási jelenségeket és műveleteket befolyásolhatják a különböző nyelvek grammatikai tulajdonságai, így feltehetőleg lesznek bizonyos eltérések, amelyek a szerkezeti különbségekből fakadhatnak.
- (iii) Mivel a magyar nyelvre vonatkozóan nem született még olyan tanulmány, amely mind a tíz javítási jelenséget vizsgálta volna, így nincsenek számadatok sem a gyakoriságukról. Németh (2012) vizsgálatában azt találta, hogy a magyarban többször fordul elő ismétlés, mint csere, így ez alapján feltételezem, hogy az ismétlések száma megelőzi a cserék számát.

2. Módszertan

A kutatás saját gyűjtésű hanganyagok alapján készült. A Kocsma-korpusz baráti összejövetelekről, kocsmai beszélgetésekről készült hangfelvételeket tartalmaz. Jelen tanulmányhoz a korpuszból 4 óra 18 percnyi hanganyagot használtam fel. Ez 8 felvételt jelent, amelyek közül a legrövidebb 9 perc 23 másodperc, a leghosszabb pedig 47 perc 35 másodperc, így átlagot számítva 32 perc 11 másodperces egy hangfájl a feldolgozott anyagok közül.

A felvételek több különböző időpontban készültek eltérő számú résztvevővel. A korpusz tartalmaz kétfős, háromfős és ötfős társalgást. Az 1. táblázat tartalmazza a társalgások adatait.

1. táblázat. A társalgások adatai

	Résztvevők száma	és neme	A társalgás hossza
KO_200916-14	2	1 nő 1 férfi	9 min 3 s
KO_200919-21	5	2 nő 3 férfi	15 min 49 s
KO_200922-23	3	2 nő 1 férfi	26 min 57 s
KO_200922-22	3	2 nő 1 férfi	29 min 49 s
KO_200916-17	5	3 nő 2 férfi	38 min 40 s
KO_201027-30	5	3 nő 2 férfi	42 min 54 s
KO_201027-31	5	3 nő 2 férfi	46 min 25 s
KO_201213-33	3	1 nő 2 férfi	47 min 35 s

A kutatásban 8 adatközlő vett részt, akik egy nyelví, köznyelvet beszélő sze-
gedi és budapesti lakosok, 3 nő és 5 férfi. Életkorukat tekintve 21 és 29 év
közöttiek, akik az anonimitás megőrzése érdekében nem az igazi nevükkel, ha-
nem fedőnevekkel szerepelnek, ahogyan azok a személyeket is, akik nem vettek
rész a kutatásban, de elhangzik a nevük a felvételek valamelyikében. Az adat-
közlők (2. táblázat) között iskolai végzettségük szerint fele-fele arányban vannak
felsőfokú tanulmányokat folytató és felsőfokú végzettséggel rendelkezők, illetve
két további adatközlő érettségivel rendelkezik; foglalkozásuk változatos.

Az adatközlők mindegyike a baráti társaság tagja, akik napi kapcsolatban
vannak egymással. A társaságban van két testvérpár, illetve egy pár is.

A hangrögzítés fázisában igyekeztem minimalizálni a megfigyelői paradoxon
(Labov, 1979) adatközlőkre gyakorolt hatását és maximalizálni a korpusz spon-

2. táblázat. Adatközlők

név	nem	kor	település	iskolai végzettség	foglalkozás
Etus	nő	26	megyeszékhely	egyetem	HR-es, grafikus
Feri	férfi	27	megyeszékhely	érettségi	ügyfélszolgálati munkatárs
Gábor	férfi	29	megyeszékhely	egyetem	HR-es
János	férfi	27	főváros	szakközépiskola	szakács
Jutka	nő	26	megyeszékhely	érettségi	egyetemi hallgató
Laci	férfi	24	megyeszékhely	egyetem	fordító
Lenke	nő	21	megyeszékhely	érettségi	egyetemi hallgató
Taki	férfi	21	megyeszékhely	érettségi	egyetemi hallgató

tán, természetes jellegét. A megfigyelői paradoxon minimalizálása érdekében olyan adatközlőkkel dolgoztam, akik jól ismerik egymást; a felvételeket nem stúdiókörnyezetben, hanem az általuk megszokott környezetben és élethelyzetben rögzítettem; illetve a társalgások nem irányítottak, hanem egytől egyig úgy haladtak előre, ahogy a beszélők alakították. Ezen tényezők csökkentették az adatközlők szorongását, feszélyezettségét, és növelték a természetesség mértékét. Így kellő számú adatot tudtam kiszűrni a felvételekből, és a korpusz ezáltal alkalmas lett a javítási jelenségek vizsgálatára.

Minden adatközlő beleegyező nyilatkozatban hozzájárult a kutatáshoz. A felvételek készítése előtt szóban jeleztem az adatközlők felé, hogy egy bizonyos ponton rögzíteni fogom az általuk mondottakat. A felvétel indításának és befejezésének a pontos idejét nem tudták. Ezzel biztosítottam, hogy etikusan járjak el az adatgyűjtés során, miközben a beszélgetés természetessége nem sérült. Így a korpuszom spontán, természetes társalgásokat tartalmaz.

A hangfelvételek elkészülte után a konverzációelemzés átírási konvencióinak megfelelő leíratot készítettem (lásd: Függelék). Ezután következett maga az elemzés, amely során az egyes beszédrészeket fordulókra, majd fordulókstruktúrák egységekre bontottam. Megvizsgáltam a forduló lehetséges lezárási pontjait, vagyis a váltásreleváns helyeket; a szekvenciális rendezettséget, vagyis a több fordulóból álló cselekvésláncolatok sorozatát. A kiemelt példákban meg-

mértem a beszédsszüneteket. Kutatásom során a legnagyobb hangsúlyt a javítási műveletek azonosítására, azok környezetének és a társalgás menetére gyakorolt hatásának vizsgálatára fordítottam. Az egyes javítási műveleteken kvalitatív elemzést végeztem, majd ezeket számszerűsítve kvantitatív módon is megvizsgáltam a javítási műveletek gyakoriságát az egész korpuszban.

3. Eredmények

3.1. A kvalitatív vizsgálat eredményei

3.1.1. Csere

Az (1) példában Etus és Lenke az irónia és az antiirónia közötti különbségről beszélgetnek.

(1) KO_200922-22

- 1 Etus: tarthatnál Gábornak kurzust [belőle.]
- 2 Lenke: [ja. ÖÖ] öhm viszont ugye el kell
- 3 különíteni az irónia meg az antiiróniát. .hh mer aaz antiirónia
- 4 az az amikor (.) hh ÖÖ HÚ, Etus meg megcsináltál ötvenmillió rajzot
- 5 EGY nap? JÓ LUSTA vagy. ez antiirónia. tehát hogy amikor
- 6 **pozitív [dolgokat] =**
- 7 Etus: [ü h ü m]
- 8 Lenke: **= vagy NA, negatív dolgokat mondok**, de igazából úgy értem
- 9 hogy ez pozitív.

Lenke a 6. sorban elhangzó *pozitív dolgokat* egységet *negatív dolgokat* egységre javítja (8. sor). Lenke egyből érzékeli, hogy felmerült egy hiba, amely javításra szorul. A beszélő olyannyira gyorsan észleli a problémát, a *vagy na* kifejezéssel egyértelművé teszi a beszélgetőpartnerre felé, hogy éppen egy felmerülő hibát javít, funkciója megegyezik azzal, mintha azt mondta volna, *jaj, rosszul mondtam*. Ez a javításkezdeményezés. Ebben az esetben Lenke ellentétes elemeket, antonimákat cserél ki egymással. A *dolgokat* szó ismétlése itt nem ismétlő funkcióval bír, csupán azért jelenhet meg a javításban is, mert egy

szorosan összekapcsolódó jelzős szintagmát alkot az előtte lévő szóval. Ezt a szoros kapcsolatot az is alátámasztja, hogy *a pozitív dolgokat* és *a negatív dolgokat* esetében mindkét szintagma ugyanazt a mondatrészi szerepet tölti be.

A (2) példában Feri cigarettával kínálja Lenkét.

(2) KO_200922-23

- 1 Feri: nem kérsz egy (0.5) **dobozosat, (.) egy polgárit?**
- 2 Lenke: ÖÖ DEE, köszönöm!

A fenti példát azért tekintem cserének, mert a baráti társaságban konvencionálisan a dobozos cigarettát *polgári cigarettának* nevezik. A *dobozosat* és a *polgárit* szavaknak tehát ugyanaz a referenciája, így szinonimának számítanak ebben a kontextusban. A *dobozosat* és a *polgárit* kifejezések tárgyragos alakban vannak, ugyanazt a szintaktikai szerepet töltik be, ez még inkább erősíti azt, hogy cseréről van szó.

3.1.2. Beszúrás

A (3) példában a beszúrás javítási műveletére láthatunk példákat:

(3) KO_200922-22

- 1 Etus: de mit csinálsz amúgy?
- 2 Lenke: dolgozok.
- 3 Etus: de mit?
- 4 Lenke: annotálok.
- 5 Etus: de hogy?
- 6 Lenke: .hh **egy fél- igazából egy félkész** öö munkát csinálok (.) meg,
- 7 amit (.) már valaki megcsinált, csak nem jól, (1.2) és ööm
- 8 Etus: de nem is hallgatsz [semmit.]
- 9 Lenke: [m á r] meghallgattam. nagyjából
- 10 emlékszem rá. MEG öö **fő- egyébként hallgatni főleg**
- 11 az iróniánál kell, ott meg öö (.) öö írásban is ff felismerem
- 12 valamennyire hogy ez lehet hogy ironia lesz. [így a] =
- 13 Etus: [üüm.]
- 14 Lenke: = szöveggörnyezetből.

Az 1. sorban Feri egy új fordulóba kezd. A 4. sorban a zárójelezés javítási művelettel pontosít: *tudod, amikor izé .hh a a az a nyomozó vagy rendőr faszival volt az a balhé.* Feri a történet mesélésekor többször megakad, bizonytalan és ő maga sem tudja még pontosan, hogyan formálja meg a mondanivalóját. Mivel a beszéd tervezése és kivitelezése egyszerre zajlik, ennek következtében zavar lép fel a cselekvés produkálásában (Gósy, 2005; Fabulya, 2007). A spontán társalgás úgy épül fel, hogy a beszélgetőpartner a válaszreakciójával kinyilvánítja, hogy hogyan értelmezte az előző fordulót (Heritage, 1984). Lenke válaszreakciója egymást követő, többszörösen kifejezett egyetértés, amellyel kifejezi, hogy megértette, hogy a sorozat mely pontjáról beszél Feri. A válaszreakcióból vissza-következtetve láthatjuk, hogy Feri 1–3. sorban végrehajtott cselekvése során az interszubsztitívitás sérült, és ezt sikeresen helyreállította a 4–5. sorban lévő javítással. Lenke megerősítése erre enged következtetni.

3.1.4. Kihagyás

Az (5) példában Feri arról beszél, hogy nem követi a koronavírus járvánnyal kapcsolatos híreket, mert nem szeretné stresszelni magát rajta. Emellett pedig azt gondolja, hogy már átesett a betegségen, és így kevesebb esélye van újra elkapni a vírust.

(5) KO_200922-22

- 1 Feri: nem tom, én nem is követem annyira ezt az egészet,
- 2 meer nem akarom idegesíteni magamat rajta meg
- 3 (.) meg mondom én üh annak betitulálom hogy
- 4 **én szerint- én** már túl is estem rajta.

Feri kihagyás javítási műveletet hajt végre (4. sor): a beszélő a *szerintem* szót kezdi el mondani, de végül „elvágja”, részlegesen produkálja, majd megismétli az előtte lévő *én* személyes névmást. Ez összhangban van a kihagyás fenti meghatározásával.

3.1.5. Törlés

A (6) példa jól mutatja a törlés működését:

(6) KO_201213-033

- 1 Lenke: **ketten nem olyan (0.5) öö (0.7) hát az szokatlan lenne**, ha (.)
2 a szójátékot ketten kéne.

Lenke belekezd egy fordulóba, amelynek folytonossága megszakad az *olyan* szó után a 0.5 másodperces szünettel, az *öő*-zéssel és a 0.7 másodperces szünettel. Lenke törli a megakadásjelenségek előtti beszédrészt: *ketten nem olyan*, vagyis a folyamatban lévő TCU-t és ezáltal a TCU-val végrehajtott cselekvés korábbi verzióját is. Magát a cselekvést megtartja, de végül másképpen fejezi ki a mondanivalóját.

3.1.6. Ismétlés

Az ismétlés két típusa az időnyerésre alkalmas ismétlés és az átfedéskezelő ismétlés. A (7) példában időnyerésre alkalmas ismétlés történik, amely során Etus a borfesztiválhoz fűződő emlékeit osztja meg beszélgetőtársaival:

(7) KO_200922-22

- 1 Etus: egyébként a ((torokköszörülés)) nekem a ((köhögés)) borfeszt az
2 mindig **ilyen öö ilyen számadás** az életemről, ami mindig van (.)
3 kivéve most.
4 Lenke: igen akkor most [mikor fogsz] =
5 Etus: **[hoggy így öö]**
6 Lenke: = számot adni az életemről?
7 Feri: hát, jövőre, hogyha [mázlink van]
8 Etus: [h á t ö ö]
9 Feri: ((nevetés)) (1.0) de akkor duplán kell (.) számot adni.
10 (1.5)
11 Etus: **de hoggy így, (.) hoggy így, (.)** igen, hogy ez egy olyan időszak,
12 egy alkalom amin én mindig vagyok, és akkor nem tudom,
13 átgondoltam hogy hmm, tavalgy ilyenkor, amikor a borfesztivál volt,
14 akkor ezt meg ezt csináltam, ezzel foglalkoztam, ez érdekelt, satöbbi.

Etus az *ilyen* funkciósót ismétli meg (2. sor). Feltehetőleg az ismétlés itt késleltetésre szolgál, hogy kellő ideje legyen előhívni a *számadás* tartalmat

szót. Ehhez hozzájárul még az *öö*-zés is, amellyel még több időt nyer a beszélő. Habár Schegloff (2013) a halasztó szerepű ismétléseket nem tekinti javításnak, mégis ide sorolható a jelenség az alapján, hogy ebben az esetben is megszakad a forduló progresszivitása, hogy a beszélő kezeljen egy produkcióbeli problémát (vö. Schegloff et al., 1977).

A 11. sorban ugyancsak funkciószó ismétlése történik, azonban nem egy szót, hanem két összetartozó funkciósót ismétel: *hogy így*. Az előző példával ellentétben viszont nem egy konkrét szó előhívása okozhatja a problémát, hanem egy összetett szerkezet megformálása. A progresszió megakadásához az is hozzájárul, hogy Etus kétszeri próbálkozással kísérli meg átvenni vagy visszavenni a szót, amit a beszélgetőpartnerei a beszélőváltás szabályaival összhangban vettek át tőle. Ez azt jelenti, hogy Etus nem jelöl ki más beszélőt, de záró intonációs kontúrt alkalmaz és szintaktikailag is lezártnak tekinthető a fordulója. Korábban (5. sor) Etus ugyanezt a szerkezetet produkálja, így nem csak kétszer, de háromszor hangzik el a beszédszakasz, mire sikerül megformálnia a fordulót (11–14. sor).

Az ismétlés másik típusa az átfedéskezelő ismétlés, amely során az átfedés veszélyezteti az aktuális beszélő fordulójának végrehajtását.

A (8) példában ugyancsak a Doktor House sorozatról beszélget Feri és Lenke.

[illegible]

3.1.7. Keresés

91

(9) KO_200922-22

- 1 Lenke: öö ilyen hosszú, (.) fehér haja van, (0.5) szakállá,
2 Feri: ühm
3 Lenke: és öö ilyen hon- hosszú fehér ruhában szokott öö (.) tanítani, (.) öö
4 régebben mezítláb, de [most **ilyen**] =
5 Feri: [((nevetés))]
6 Lenke: = höhő **nhem tudom ilyen (.) ilyen ilyen öö (.) üm: (.) kunkori cipőt**

Lenke különböző megakadásjelenségek sorozatát produkálja (6. sor vö. Laczkó, 2014). A *nhem tudom ilyen* egység után egy mikropauzát, az *ilyen* kifejezés kétszeri ismétlését, egyéb hezitációs jelenségeket: *öö*, *üm:*, illetve további mikropauzákat figyelhetünk meg. Maga a javítási művelet első része, a problémaforrás már korábban (4. sor) megjelenik, így az ilyen kifejezés valójában háromszor ismétlődik. Ezek a megakadások a beszélő mondanivalója megfogalmazásának nehézségeit, a beszélő bizonytalanságát fejezhetik ki (Vukov Raffai, 2016), amelyhez erősen hozzájárul a *nhem tudom* is. Végül nagy mentális erőfeszítés árán Lenkének sikerül előhívnia a kunkori cipő kifejezést. Vannak olyan esetek, amikor a nagy erőfeszítések ellenére sem sikerül megtalálni a keresett szót. Ez lehet azért, mert a keresett szóra nincsen megfelelő kifejezés az adott nyelvben, vagy teljesen hiányzik a beszélő szókincséből, vagy csak egyszerűen hiába ismeri a szót, kifejezést, az adott helyzetben nem képes előhívni a mentális lexikonból. Ilyenkor gyakran körülírást alkalmaz a beszélő (Iványi, 2002). Habár a konverzációelemzés nem bocsátkozik találgatásokba, mi lett volna a soron következő egység, feltehetőleg itt nem maga a *kunkori cipő* előhívása okozza a problémát, hanem az annak megfelelő pontos kifejezés, azonban ilyen nagy számú megakadásjelenség után a beszélő kénytelen valami mással helyettesíteni, hiszen könnyen elveszítheti a fordulóhoz való jogát.

Az (10) társalgásrészletben Jutka, Lenke és Laci egy közös ismerősük nem megfelelő viselkedéséről beszélgetnek.

(10) KO_200917-17

- 1 Laci: amúgy, (.) öö [velem (.) is a szokásosnál kevésbé] =
2 Lenke: [mondjuk csak két emberrel fogott kezét.]
3 Laci: = kedves.
4 Jutka: öö igen azt [mondta is Laci] =
5 Laci: [meg kell hagyni.]
6 Jutka: = hogy milyen izé volt hét (.) fön.
7 Laci: jaja
8 Jutka: én nem nem nagyon vet (.) tem észre
9 Jutka: de nem is figyeltem rá nagyon.
10 Laci: hm **kicsit** (.) **segg** volt.

A beszélgetőpartnerek alapvetően egy kényes témát boncolgatnak. Laci fordulójában (10. sor) a *kicsit* kifejezést egy szünet követi, egy mikropauza, amellyel a beszélő megszakítja a folyamatban lévő fordulóját. Ezután a *segg* diszpreferált, sértő kifejezést produkálja. Ezek együttese azt erősíti, hogy *delicate* típusú keresés történik.

A korpuszban megjelent egy hasonló, de nem teljesen azonos jelenség, amikor Lenke egy másik résztvevő díszpárnájára tesz negatív megjegyzést.

(11) KO_200922-22

- 1 Lenke: amúgy nagyon, **(1.0) rendkívül** (.) **SAJÁTOS** ez a párna.

Lenke ebben az esetben igyekszik elkerülni a kényes kifejezést. A *nagyon* szó kimondásakor egyértelművé teszi, hogy folytatni kívánja a már megkezdett, folyamatban lévő TCU-t. Az 1 másodperces szünet produkálása után még mindig nem tud előhívni egy olyan kifejezést, amely megfelelő lenne az adott szituációban, így a *nagyon* szót a *rendkívül* határozóval fokozza. A *rendkívül* után még egy szünetet, ezúttal egy rövidebb mikropauzát produkál, amely után megtalálja a megfelelő szót. Mivel a *sajátos* kifejezés nem tekinthető sértőnek, inkább semlegesnek, tehát nem produkál kényes kifejezést, így itt sokkal inkább a potenciálisan sértő kifejezés elkerülése, és annak egy, a szituációnak megfelelő, kevésbé sértő formájának keresése történik. Ez szorosan összefügg Leech (1983) jóvá-

hagyás maximájával, miszerint az interakcióban minimalizáljuk a partnerünk leszólását, és maximalizáljuk a partnerünk dicséretét.

3.1.8. Szekvenciaugrás

A (12) társalgásrészletben Jutka elmeséli, hogy vett le pénzt a bankkártyájáról, amikor hirtelen zajt hall.

(12) KO_200916-17

- 1 Jutka: én meg, direkt kerültem hogy vegyek le pénzt,
- 2 és aztán **kide- azt hittem jön a cica.**
- 3 Lenke: egyszer már elment.
- 4 Jutka: ja. szóval igen aztán kiderült hogy volt nálam pénz.

Jutka egy történetmesélésbe kezd, amikor egy külső tényező, egy hirtelen zaj észlelése révén megszakítja a folyamatban lévő fordulóját egy oda nem illő cselekvéssel: *kide- azt hittem jön a cica*.. A beszélő fordulójának megszakításával elhagyja az eredeti cselekvést, és azáltal, hogy Lenke erre válaszreakciót produkál, kilép az eredeti szekvenciából. Ezután pedig visszatér az eredeti fordulójához.

A korpuszban egy más, de kétségtelenül hasonló jelenség is történik. Feri, Etus és Lenke a borkészítésről beszélnek. Feri elmeséli, hogy a nővére és a párja Albertirsán éltek, és ott készítettek bort. Etusnak és Lenkének az Albertirsa településnevet meghallva eszébe jut a 2010 környékén, interneten felkapott Albertirsa fociinduló, amit énekelni kezdenek.

(13) KO_200922-22

- 1 Feri: nekem izé nővérem párja mikor öö Albertirsán laktak,
- 2 Etus: AL[BERTIRSAA]
- 3 Lenke: [ALBERTIRSAA]
- 4 Feri: [i g e h e n] (.) igen.
- 5 (1.0)
- 6 Feri: így izé, ő csinált bort, így együtt szüreteltük így családdal.

Feri belekezd egy fordulóba, egy történetmesélésbe (1. sor). Egy fordulókonstruktív egységet produkál, amit csak részben tud végrehajtani, erre a

záró intonációs kontúr elmaradása enged következtetni. A TCU befejezését Etus és Lenke énekléssel való belépése akadályozza meg. A két beszélő nem váltásreleváns helyen lép be, sértve ezzel a beszélőváltás normatív szabályait. Ezen a ponton azonban Etus és Lenke nem akarja átvenni a szót, csak egyfelől kinyilvánítják a településhez kapcsolódó élményüket, másfelől pedig színesítik Feri történetét. Itt tehát megszakad a cselekvés progresszivitása, azonban a szekvenciaugrás schegloffi leírásával ellentétben nem az eredeti beszélő, hanem beszélgetőpartnerei szakítják meg azt egy oda nem illő cselekvéssel, vagyis az énekléssel. Feri az éneklésre nevetve helyesel, vagyis válaszreakciót produkál, amivel maga is elhagyja az eredeti fordulót, majd a válaszreakciója után csak egy 1 másodperces szünet után tér vissza rá. Ez alapján a szekvenciaugrással rokon jelenségnek tekintem a fenti műveletet.

3.1.9. Újrarendezés

Az újrarendezés az angol nyelvű meghatározástól eltérően valósult meg a korpuszban. Ezt a műveletet először Schegloff (2013) nyomán Sidnell példáján keresztül mutatom be, majd ezt összevetem a saját korpuszomból vett példával.

(14) Sidnell, 2006

1 Que: But do you **get alway– d’ you always get** called on?

Az angol nyelvben az *always* gyakoriságot kifejező határozó grammatikusan az alany után, de még az ige előtt áll. A fenti példában Que a forduló első verziójában nem ezt a szórendet használja, hanem felcseréli az igét és az *always* kifejezést. A szó elvágása, az *alway–* után egyből javítja a problémát.

A fentivel nem azonos, de hasonló jelenség történik az alábbi, saját korpuszban található példában:

(15) KO_201027-30

1 Lenke: **kiesett** a (1.5) magától **esett ki**,

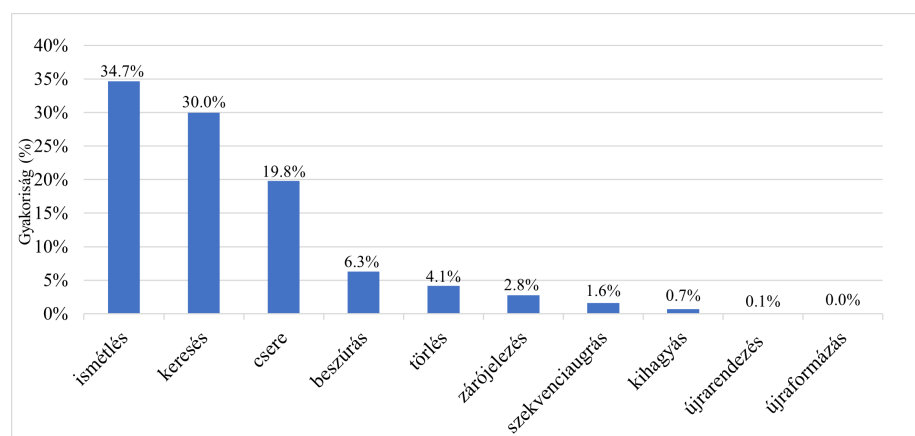
A fenti példában Lenke megszakítja a fordulóját egy 1.5 másodperces szünettel, majd beszúr egy elemet: *magától*, ezután pedig újratermészt a *kiesett*

elemet ezúttal felcserélve az igekötőt és az igét. Itt felvetődik, hogy a beszúrás révén inherens ismétlés történik. Annak alapján, hogy a megismételt egységet más formában produkálta a beszélő, a csere lehetősége is felmerül. Itt azonban a beszúrás megköveteli a szórendi változtatást, hiszen a *magától* fókuszpozícióba kerül, emiatt az igekötőt át kellett rendezni az ige mögé. Mivel a fenti példa összecseng az újrendezés definíciójával, miszerint a beszélő módosítja a nyelvtani egységek sorrendjét, ennek alapján sorolom az újrendezés kategóriájába.

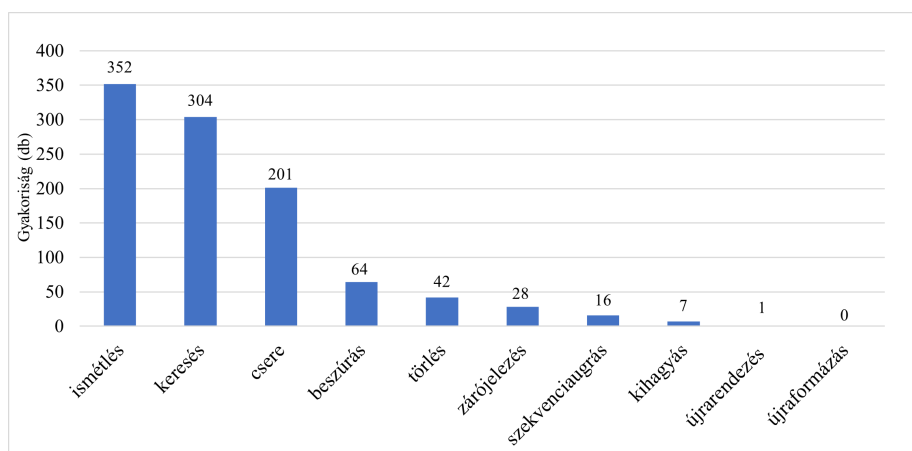
3.2. A kvantitatív vizsgálat eredményei

A magyar nyelvre vonatkozóan nem született még teljes, átfogó tanulmány a tíz javítási műveletről, ezért nincsenek számadataink az összes művelet előfordulásáról. Németh (2012) az ismétlés és a csere interakciós funkcióját vizsgálta a magyar nyelvben, amely során megállapította, hogy az ismétlés preferenciát mutat a cserével szemben, ennek alapján feltehetőleg gyakoribb művelet a magyar nyelvben.

Habár Kitzinger (2013) a cserére és a beszúrára gyakori javítási műveletként hivatkozik, ezt nem támasztják alá számadatok. A kvantitatív elemzés a tíz javítási művelet gyakoriságát mutatja be a magyar nyelvű saját korpuszban. A teljes korpusz 1016 javítási műveletet tartalmaz. A különböző típusok eltérő arányban jelentek meg a korpuszban (1. és 2. ábra).



1. ábra. A javítási műveletek gyakorisága (%)



2. ábra. A javítási műveletek gyakorisága (db)

A javítási műveletek gyakorisági sorrendje a következő: ismétlés, keresés, csere, beszúrás, törlés, zárójelezés, szekvenciaugrás, kihagyás és újrendezés; az újraformázás egyáltalán nem jelent meg a korpuszban.

A jelen korpuszban korábbi kutatási eredményekkel összhangban az ismétlés mutatta a legnagyobb gyakoriságot. Ezek leginkább tartalmas szavakat megelőző funkciószó-ismétlések voltak, amelyek időnyerésre alkalmas funkciót láttak el. Mivel a keresés egyik altípusa, a *precise* típusú keresés gyakran egybeveg az időnyerő funkciót betöltő ismétléssel, így a javítások egy részében egyszerre beszélhetünk ismétlésről és keresésről.

A vizsgálat során összehasonlítottam az ismétlés két típusának arányát, amely a következő: 15% átfedéskezelés és 85% időnyerés. Az ismétlés időnyerő funkciót betöltő típusa tehát jóval többször fordult elő a társalgásokban. Ennek oka, hogy a soron következő szavak, kifejezések előhívása során a beszélő gyakran akadályokba ütközik, nem képes, vagy csak nagyobb mentális erőfeszítés árán képes előhívni a keresett szót. A jelen vizsgálatban az újraindításokat, vagyis azokat az eseteket is ismétlésnek tekintettem, amikor a javításkezdeményezés még a szó felismerhetőségi pontja előtt történik (pl.: *minde mindegyik*) (Németh, 2017). Az ismétlések típusainak arányát befolyásolta a beszélők száma. Ötfős társalgások esetén az időnyerő funkciót ellátó ismétlések az ismétlések

számának 79%-át, míg az átfedéskezelő ismétlések a 21%-át tették ki. Háromfős társalgások esetén pedig az időnyerő funkciót betöltő ismétlés 91%-ot, az átfedéskezelő ismétlés pedig 9%-ot mutatott. Ez feltehetőleg annak köszönhető, hogy egy több résztvevős társalgásban lényegesen több átfedés van, sokkal inkább meg kell küzdeni a szó jogáért, így a beszélők igyekeznek átfedéskezelő ismétléssel biztosítani a fordulóhoz való jogukat (vö. Schegloff, 2013).

A keresés két típusa ugyancsak eltérő arányban jelent meg a korpuszban. A *precise* típusú keresés a keresések 97%-át, a *delicate* típusú keresés pedig a keresések 3%-át adta. A *precise* típusú keresés az időnyerő funkciót betöltő ismétléshez hasonlóan a beszédtervezési nehézségekkel van összefüggésben, így a nagyarányú előfordulása nem meglepő.

A csere esetén a beszélő általában a problémaforrással azonos TCU-ban hajtott végre a javítást, de előfordult olyan eset is, amikor a csere azonos fordulóban, de már váltásreleváns hely után valósult meg. A csere gyakran járt együtt a kereséssel az *izé* mint lexikális kitöltőelem révén, melynek során a beszélő helyettesít egy lexémát, amely átmenetileg nem elérhető a mentális lexikonból (Fabulya, 2007). A csere az egyes társalgásokban közel azonos arányban jelent meg, nem befolyásolta a beszélők száma.

A beszúrást esetén általában nem egy nyilvánvaló hiba javítása, hanem a forduló finomra hangolása történt. A beszélők általában akkor alkalmazták ezt a műveletet, amikor mondanivalójuk egy bizonyos szegmensét színesíteni, pontosítani, illetve enyhíteni kívánták. A zárójelezés leginkább történetmesélés során fordult elő, amellyel a beszélő biztosította a történet követhetőségét azáltal, hogy vagy egy résztvevők által nem ismert háttérinformációt szűrt be zárójeles megjegyzésként, vagy egy résztvevők által ismert információt idézett fel a követhetőség érdekében.

Mind a törlés, mind a kihagyás esetén egy korábbi egység törlése történt. A törlésnél a beszélők ugyanazon cselekvést produkáltak másképpen, pontosítva, gyakran a társas interakció preferenciarendszerét figyelembe véve. Ez általában egy egész fordulókonstruktív egységnyi beszédrészt érintett. A kihagyás ezzel szemben egy szót, összetett kifejezést vagy jelzős szerkezetet érintett.

A szekvenciugrást gyakran valamilyen külső, az aktuális forduló cselekvésétől független tényező váltotta ki, amelyet a beszélő azonnal közölni kívánt. Mivel a hanganyagok nem stúdiókörnyezetben, hanem kötetlen kocsmai beszélgetés során lettek rögzítve, így gyakori volt, hogy a beszélők öngyűjtőkéréssel, illetve dohányzásra invitálással szakították meg a fordulót, elhagyva ezzel az eredeti cselekvést és megszakítva a szekvenciát.

Az újrendezés az egész korpuszban csak egy előfordulást mutatott. Ennek oka feltehetőleg a magyar és az angol nyelv eltérő grammatikai szerkezete. Ez összecseng azzal, hogy a nyelvek grammatikai szerkezete befolyásolja a javításra alkalmas stratégiákat (Kitzinger, 2013). Az újraformázás egyáltalán nem fordult elő a korpuszban, amely ugyancsak grammatikai okokkal magyarázható.

4. Következtetések

A tanulmány célja magyar nyelvű spontán, baráti társalgásokban megjelenő javítási műveletek elemzése a konverzációelemzés elméleti keretén belül. A csere, a beszúrás, a zárójelezés, a kihagyás, a törlés, az ismétlés, a keresés, a szekvenciaugrás, az újraformázás és az újrendezés elemzése saját korpuszom alapján történt. A korpuszban ezek közül csak az újraformázás nem fordult elő; így az első hipotézisem, miszerint megjelenik mind a tíz, a konverzációelemzés által eddig feltárt javítási művelet, nem teljesült. A második hipotézisem szerint a javítási műveletek többnyire megegyeznek az angol szakirodalom meghatározásával, de lesznek olyan esetek, amelyek a nyelvek különböző grammatikájából fakadóan eltérést mutathatnak. Az eredmények szerint a magyar nyelvű korpuszban megjelent javítási műveletek többnyire azonosak voltak az angol nyelvű példákkal, de voltak olyan egyedi esetek, amelyek eltérést mutattak. Azonosságot mutatott a csere, a beszúrás, a zárójelezés, a kihagyás, a törlés, az ismétlés két típusa, a *precise* és a *delicate* típusú keresés és a szekvenciaugrás. A *delicate* típusú keresés (11-es példa), a szekvenciaugrás (13-as példa) megjelent a meghatározástól részben eltérő módon is. A 11-es példában bemutatott *delicate* típusú keresés abban tért el a meghatározástól, hogy a beszélő ahelyett, hogy

sértő kifejezést produkált volna, keresett egy, a szituációnak megfelelő semleges kifejezést. Így tehát a potenciálisan sértő kifejezés elkerülése, és annak kevésbé diszpreferált formájának keresése történt. A 13-as példában bemutatott szekvenciaugrás esetén a meghatározással ellentétben nem az eredeti beszélő, hanem beszélgetőpartnerei szakították meg a forduló progresszivitását egy oda nem illő cselekvéssel, vagyis az énekléssel. Ez azt eredményezte, hogy az eredeti beszélő válaszreakciót produkált, ezáltal ő maga hagyta el a fordulóval végrehajtott cselekvést, és lépett ki a szekvenciából. Az újrendezés a meghatározásnak megfelelő módon nem jelent meg, csak némi eltéréssel. A 15-ös példában bemutatott újrendezés esetén felmerült, hogy az újrendezett forma a beszúrás révén egy inherens ismétlés, illetve mivel a megismételt formában a beszélő az ígét és az igekötőt felcserélte, a csere lehetősége is. Azonban ebben az esetben a beszúrás megkövetelte a szórendi változtatást, hiszen a beszúrt elem fókuszpozícióba került, emiatt az igekötőt át kellett rendezni az ige mögé. Ezek az eltérések a műveletek új vonásait mutatták meg. Érdeemes lenne további vizsgálat alá vetni őket bizonyos nyelvspecifikus műveletek felismeréséhez.

A tíz javítási művelet gyakoriságának elemzése azt mutatta, hogy ezek gyakorisági sorrendje a következő volt a korpuszban: ismétlés, keresés, csere, beszúrás, törlés, zárójelezés, szekvenciaugrás, kihagyás, újrendezés. Ismétlés esetén jóval nagyobb számú előfordulást mutatott az időnyerő funkciót betöltő ismétlés, keresés esetén pedig a *precise* típusú keresés. Az időnyerésre alkalmas ismétlés gyakran együtt járt a *precise* típusú kereséssel. Az átfedéskezelő ismétlések számát befolyásolta a beszélők száma: több résztvevő esetén nagyobb előfordulást mutattak. Mivel a *delicate* típusú keresés a *precise* típusú kereséssel ellentétben kifejezetten a tabu, potenciálisan sértő kifejezések esetén történik, kisszámú előfordulásának oka egyrészt az lehet, hogy egy társalgásban viszonylag ritkán jelennek meg ezek a diszpreferált nyelvi jelenségek, másrészt pedig, hogy az adatközlők baráti, illetve rokoni kapcsolatban vannak egymással, így kevésbé feszélyezi őket a kényes kifejezések produkálása.

Az újrendezés csupán egy esetben jelent meg a korpuszban, amely grammatikai okokra vezethető vissza. Mivel a kihagyás és az újrendezés nagyon

alacsony előfordulást mutatott, illetve az újraformázás egyáltalán nem jelent meg, figyelembe véve azt, hogy mindhárom javítási művelet a törléshez hasonlóan működik, érdemes lenne ezeket a kategóriákat összevonni, amely által könnyebb, hatékonyabb lenne a javítási műveletek kategóriákba sorolása.

A jelen kutatásban hétköznapi, spontán, baráti társalgásokban megjelenő javítási műveleteket vizsgáltam a konverzációelemzés elméleti keretén belül. Mivel a korpusz csak kötetlen, hétköznapi beszélgetéseket tartalmaz, érdemes lenne megvizsgálni és összevetni más szituációban és környezetben rögzített, különböző demográfiai státuszú beszélők társalgásait is, amely során kimutathatóak lennének az esetleges interakciós sajátosságok.

Köszönetnyilvánítás

Köszönetet mondok témavezetőmnek, Dr. Németh Zsuzsannának és a lektoroknak, hogy értékes, építő jellegű megjegyzéseikkel hozzájárultak a tanulmány elkészüléséhez.

Hivatkozások

- Fabulya, M. (2007). Izé, hogyishívják, hogymondjam: Javítást kezdeményező lexikális kitöltőelemek. *Magyar Nyelvőr*, 131, 324–342.
- Fox, B., Wouk, F., Hayasi, M., Finckle, S., Tao, L., Sorjonen, M.-L., Laakso, M., & Hernandez, W. F. (2009). A cross-linguistic investigation of the site of initiation in same-turn self-repair. In J. Sidnell (Ed.), *Conversation Analysis: Comparative Perspectives* (pp. 60–103). Cambridge: Cambridge University.
- Gósy, M. (2004). A spontán magyar beszéd megakadásainak hallás alapú gyűjteménye. *Beszédkutatás 2004*, (pp. 6–186).
- Gósy, M. (2005). *Pszicholingvisztika*. Budapest: Osiris Kiadó.
- Gyarmathy, D. (2009). A beszélő bizonytalanságának jelzései: ismétlések és újraindítások. *Beszédkutatás*, 2009, 196–216.

- Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.
- Horváth, V. (2014). *Hezitációs jelenségek a magyar beszédben*. Budapest: ELTE Eötvös Kiadó.
- Iványi, Z. (2001). A nyelvészeti konverzációelemzés. *Magyar Nyelvőr*, 125, 74–93.
- Iványi, Z. (2002). A szókincs hiányosságainak problémája hétköznapi beszélgetésekben. *Nyelvtudományi Közlemények*, 99, 281.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the First Generation* (pp. 13–31). Amsterdam: John Benjamin.
- Kitzinger, C. (2013). Repair. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 229–256). Oxford: Wiley-Blackwell.
- Labov, W. (1979). A nyelv vizsgálata társadalmi összefüggésben. In C. Pléh, & T. Terestyéni (Eds.), *Beszédaktus – kommunikáció – interakció* (pp. 365–398). Budapest: Tömegkommunikációs Kutatóközpont.
- Laczkó, M. (2014). Hezitálássorozatok a spontán beszédben: gyakoriság, forma és funkció. *Magyar Nyelvőr*, 138, 315–327.
- Leech, G. (1983). *Principles of Pragmatics*. London: Longman.
- Lerner, G. (2013). On the place of hesitating in delicate formulations: a turn constructional infrastructure for collaborative indiscretion. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational Repair and Human Understanding* (pp. 95–134). Cambridge, UK: Cambridge University Press.
- Lipóczi, B. (2020). *Történetmesélés magyar nyelvű társalgásokban. Szakdolgozat*. Szeged: Szegedi Tudományegyetem Általános Nyelvészeti Tanszék.
- Németh, Z. (2012). Az ismétlés és a csere interakciós funkciói magyar nyelvű spontán társalgásokban. *Beszédkutatás*, (pp. 154–167).

- Németh, Z. (2014). A javítási műveletek jelöltségi hipotézise. *Jelentés és Nyelvhasználat*, 1, 29–54.
- Németh, Z. (2017). Négy javítási művelet interakciós funkciói magyar nyelvű társalgásokban. In T. Szécsényi, & E. N. T. (Eds.), *Stratégiák és struktúrák: tanulmányok Kenesei István 70. születésnapjára* (pp. 207–225). Szeged: JATEPress.
- Németh, Z. (2020). A nemlexikális öö hang interakciós szerepének elemzése magyar nyelvű társalgásokban. *Jelentés és Nyelvhasználat*, 7, 23–50.
- Rieger, C. (2003). Repetitions as self-repair strategies in english and german conversations. *Journal of Pragmatics*, 35, 47–69.
- Schegloff, E. A. (1987). Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, 50, 101–114.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University.
- Schegloff, E. A. (2013). Ten operations in self-initiated, same-turn repair. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversation Repair and Human Understanding* (pp. 41–70). Cambridge: Cambridge University Press.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361–382.
- Sidnell, J. (2006). Repair. In J. O. Ostman, & J. Verschueren (Eds.), *Handbook of Pragmatics*. Amsterdam/Philadelphia: John Benjamins.
- Ullmann, S. (1967). *Semantics: An Introduction to the Science of Meaning*. Oxford: Basil Blackwell.
- Vukov Raffai, É.. (2016). A diskurzusjelölő-választások életkori sajátosságai az *így, ilyen, hát, mondjuk, ugye* esetében. *Magyar Nyelvőr*, 140, 483–497.

Wilkinson, S., & Weatherall, A. (2011). Insertion repair. *Research on Language and Social Interaction*, 44, 65–91.

Függelék

A példákban használt átírási konvenciók (Jefferson, 2004)

(.)	mikropauza; 0.5 másodpercnél rövidebb szünet
(2.0)	megmért szünet (másodperc.tizedmásodperc)
[ja	a bal oldali szögletes zárójel az átfedés, vagyis az egyszerre beszélés kezdetét jelenti
[aha	
=	az egyenlőségjel összekapcsolja ugyanazon beszélő folyamatos beszédének részeit, ha azokat egy másik beszélő közbeszólása miatt az átírásban meg kell szakítani
ja]	a jobb oldali szögletes zárójel az átfedés végét jelzi; az átfedésben lévő egységek
aha]	mindig pontosan egymás alá kerülnek
.hh	hallható lélegzetvétel; jelezheti a beszédszándékot
nem,	a vessző folytatólagos intonációt jelez
nem	a pont eső intonációt jelez; ez nem feltétlen esik egybe a szerkezeti egység végével
NEM	a nagybetűs írás nyomatékos, hangsúlyos közlés jelez;
	a hangerő magasabb az átlagoshoz képest
nem?	a kérdőjel emelkedő intonációt jelez
neem	
nemm	a betűk duplikálása hangnyújtást/hangnyúlást jelez
ne-	a kötőjel a szó elvágását (cut-off) jelöli
((nevetés))	dupla zárójel egy olyan eseményt ír le, amit hanghatásokkal nem lehet érzékelteni
((köhögés))	

A dajkanyelv tagolódása 6 hónapos csecsemőkhöz szóló történetmesélésben

Kohári Anna¹, Harmati-Pap Veronika¹, Mády Katalin¹

¹*Nyelvtudományi Kutatóközpont, Általános és Magyar Nyelvészeti Intézet*

Abstract

The occurrence of pauses serves as an aid for children with speech parsing and, hence, language acquisition. Thus, understanding temporal patterns of infant-directed speech (IDS) is of crucial importance. However, there is only sparse data available in the literature on the frequency of pauses and their relationship with clause boundaries. In this study, we investigated the semi-spontaneous speech of 14 mothers – native speakers of Hungarian – directed to their 6-month-old infant. First, mothers were asked to tell a story based on pictures to an adult, then to their child. Although they had to incorporate prescribed sentences into their stories in the experiment, only the spontaneous parts of their speech were considered for the present work. The results have shown that the articulation rate of IDS was generally lower, and the clause and IPU durations were higher than in the case of adult-directed speech (ADS). The duration of pauses was largely affected by whether they occurred within or between clauses in both registers. Although the pause duration between clauses did not show substantial differences between the two registers, the pauses within clauses were found to be longer in IDS than in ADS. Moreover, pauses appeared more frequently at clause boundaries than within the clauses in both registers, but this difference was more pronounced in IDS. The results may drive attention to the observation that the inter- or intra-clause position of pauses affects their duration and frequency in infant-directed speech.

1. Bevezetés

A dajkanyelv, azaz a kisgyerekekhez szóló beszéd (IDS: *infant directed speech*) számtalan tulajdonságában eltér a felnőttekhez szóló beszédétől (ADS: *adult directed speech*). A gyermekekhez szóló beszédben lexikai és morfoszintaktikai szinten megjelenik a redundanciára való törekvés. A felnőttek csökkentett szókinccset használnak, és többször ismétlik ugyanazon szavakat a gyerekekkel történő kommunikáció során (Broen, 1972; Snow, 1972; Stern et al., 1983; Harmati-

Email addresses: kohari.anna@nytud.hu (Kohári Anna),
harmati.pap.veronika@nytud.hu (Harmati-Pap Veronika), mady@nytud.hu (Mády Katalin)

Pap et al., 2021). Akusztikai szinten a dajkanyelvnek kimutathatóan magasabb az alapfrekvenciája, tipikusan nagyobb az akusztikai magánhangzótere, és lassabb a tempója, mint a felnőttekhez szóló beszédnek (Saint-Georges et al., 2013; Martin et al., 2016). A dajkanyelv fluenciájáról, azaz a beszéd folytonosságáról, szünetek általi tagolódásáról, különösen ezeknek a tagmondathatárokhöz való viszonyáról viszont csak szórványos adatok állnak rendelkezésre. Mivel a tagolódás, például a szünetek megjelenése közvetlenül segíthet a gyermekeknek a beszédfolyamat egységekre, szavakra tagolni, és ezáltal elsajátítani a nyelvet (Johnson & Seidl, 2008; Ludusan et al., 2016; van Ommen et al., 2020), ezért kulcsfontosságú a dajkanyelv tagolódásának és időszerkezetének minél alaposabb megértése.

1.1. A dajkanyelv tagolódása

A dajkanyelv időzítésével kapcsolatban több nyelvre (pl. angol, holland, német, tagalog, tamil) kimutatták, hogy az artikulációs tempója alacsonyabb, mint a felnőttekhez szóló beszédé (Fernald & Simon, 1984; Narayan & McDermott, 2016). Ez a tempóbeli lassulás függetlennek tekinthető a beszédmódtól, ugyanis a dajkanyelv lassabb tempóját felolvasásban és spontán beszédben egyaránt kimutatták (Church et al., 2005; Martin et al., 2016). Abban azonban nincs egyetértés, hogy ez a lassabb tempó hogyan jön létre, vajon minden szótag nyúlik-e, vagy csak bizonyos szótagok. A felolvasott mondatok szótagjait és magánhangzóit szisztematikusan hosszabbnak mérték a gyermekekhez szóló beszédben, mint a felnőttekében, függetlenül a szó szóosztályától, illetve szófajától (Ko & Soderstrom, 2013; Kohári et al., 2022). Spontán beszédben viszont nem volt kimutatható, hogy tendenciózan minden szótag nyúlna, a frázisbelseji szótagok tempójáról ugyanis nem sikerült kimutatni, hogy a dajkanyelvben alacsonyabb lenne, mint a felnőttekhez szóló beszédben (Church et al., 2005; Martin et al., 2016). Ez azt eredményezheti, hogy a dajkanyelv bizonyos beszédegységeiben több olyan szótag jelenik meg, amelyek nyúlnak, míg más egységekben kevesebb ilyen hosszabb szótag fordulhat elő. A gyermekekhez szóló beszéd idő-

zítésében tehát jelentősége lehet annak, hogy milyen és mekkora szakaszokat veszünk figyelembe a tempó mérése során.

A dajkanyelvben igen gyakran a szünetek általi tagolást vették alapul az akusztikai mérésekhez (pl. Fernald & Simon, 1984; Church et al., 2005). A szünetek időtartamáról több nyelvben kimutatták, hogy általában hosszabbak dajkanyelvben, mint felnőttekhez szóló beszédben (amerikai angol: Stern et al., 1983; német Fernald & Simon, 1984; mandarin: Grieser & Kuhl, 1988; amerikai és brit angol, japán, francia, olasz: Fernald et al., 1989). Hozzá kell tennünk azonban, hogy előfordult olyan vizsgálat is, amelyben a szünetek időtartama nem tért el statisztikailag alátámaszthatóan a két regiszterben (Menn et al., 2022). Ebben a vizsgálatban a felnőttekhez és a gyermekekhez szóló beszéd tartalma jól körülhatárolt volt, hiszen a résztvevőknek meghatározott tárgyakról kellett beszélniük mindkét regiszterben. Mivel ezen tanulmány fókuszában nem a szünet akusztikai vizsgálata állt, így nem részletezték a mérések módszertanát, ugyanakkor a többi tanulmány leírása több kérdést is felvet. Azokban a tanulmányokban, amelyekben találtak eltérést a szünetek időtartamában a két regiszter között, közös, hogy a felvételek játék vagy szabad foglalkozás közben, spontán társalgás során készültek, továbbá nem vették figyelembe a 300 ms alatti szüneteket (Stern et al., 1983; Fernald & Simon, 1984; Grieser & Kuhl, 1988; Fernald et al., 1989). Más, a dajkanyelv tempójával és időszerkezetével kapcsolatos szakirodalomban viszont a megnyilatkozásokat 200 ms-os (pl. Martin et al., 2016), vagy 260 ms-os (Broen, 1972; Fisher & Tokura, 1996) szünetek alapján határolták el egymástól. A szünettel kapcsolatos általános szakirodalomban sincs arról egyetértés, hogy a szünetnek mennyi lenne a minimum időtartama, vagy egyáltalán szükséges-e ilyen határ megszabása (l. részletesen Trouvain & Werner, 2022). A rendelkezésünkre álló jelenlegi eszközökkel jó minőségű hangfelvételen megállapítható egy beszédhang hosszával összemérhető szünet határa is, amely lehetőséget ad a dajkanyelvi szünetezés még részletesebb feltárására.

Nemcsak a szünetek hosszáról, hanem a gyakoriságáról is felmerült, hogy eltérne a gyermekekhez és a felnőttekhez szóló beszédben. Több vizsgálatban arra a következtetésre jutottak, hogy a beszélők a dajkanyelvet gyakrabban tagolták

szünetekkel, mint a felnőttekhez szóló beszédet (Martin et al., 2016; Menn et al., 2022). Martin és munkatársai (2016) japán gyermekekhez szóló spontán beszédben azt találták, hogy a szünetek közti beszédszakaszok időtartama kisebb, mint a felnőttekhez szóló beszédben, miközben a dajkanyelv tempója alacsonyabb volt. Ebből arra következtettek, hogy a szünetek gyakoribbak a gyermekekhez szóló beszédben a másik regiszterhez képest. Egy vizsgálatban nemcsak a beszédszakaszok időtartamát, hanem a szótagjainak számát is kisebbnek találták a dajkanyelvben a felnőttekéhez képest (Grieser & Kuhl, 1988). Menn és munkatársai (2022) közvetlenül a szünetek percenkénti gyakoriságát elemezték német anyanyelvűek spontán beszédében, és eredményeik szintén megerősítették, hogy a szünetek előfordulása gyakoribb a dajkanyelvben, mint a felnőttek közötti interakcióban. Habár egy kivétellel csak közvetett adatok állnak rendelkezésre, a mérések egy irányba mutatnak a tekintetben, hogy az összes szünet gyakorisága eltér a két regiszterben.

A folyamatos beszéd azonban nemcsak a szünetek által tagolódik egységekre, hanem a szintaxis által megnyilatkozásokra, illetve tagmondatokra is osztható. A különféle tagolódásokat összevető elemzések azt mutatták, hogy a dajkanyelvben a szünetek jelentős része (90 fölötti arányban) a megnyilatkozások határán fordult elő, és csak töredékük megnyilatkozásokon belül. A felnőttekhez szóló beszédben ugyanakkor ez az arány jóval alacsonyabb (50%–68%) értékek között mozgott (Broen, 1972; Fernald & Simon, 1984; Fisher & Tokura, 1996). A megnyilatkozások meghatározása egészen eltérő volt a különböző tanulmányokban, míg az egyikben kizárólag a leíratot vették figyelembe prozódiai megvalósulások nélkül (Fernald & Simon, 1984), addig egy másik elemzésben a leírat mellett figyelembe vették a hosszú szüneteket is (Fisher & Tokura, 1996), míg a harmadikban egy percepciósi kísérletből kinyert megnyilatkozáshatárokat vizsgáltak (Broen, 1972). A mért szünet alsó határa az egyik vizsgálatban 300 ms (Fernald & Simon, 1984), a másik két tanulmányban pedig 260 ms volt (Broen, 1972; Fisher & Tokura, 1996). A megnyilatkozás és a szünetnagyság definiálásának problémájától eltekintve ugyanarra az eredményre jutottak, hogy a szünetek tipikusan a tagmondatoknál nagyobb egységek, egyszerűsítve megnyilatkozások

után jelentek meg. Az egyik elemzés arra is kiterjedt, hogy a megnyilatkozások után mennyire következetesen jelennek meg a szünetek (Fisher & Tokura, 1996). Az amerikai angol dajkanyelvben azt találták, hogy a megnyilatkozások 59%-át követte szünet, míg japánban ez az arány 69% volt. Mivel felnőtt-felnőtt interakciókat nem vizsgáltak, így nem egyértelmű, hogy a megnyilatkozások után bekövetkező szünetek gyakorisága mennyiben tekinthető a regiszter és mennyiben a vizsgált nyelv sajátosságának.

A szünetek gyakorisága mellett a szünetek időtartamát is megvizsgálták különböző beszédegységek határán. A megnyilatkozás mellett a szintaxis alapján jelölt tagmondatok és a prozódia alapján meghatározott frázisok határait is elemezték. Két amerikai angol édesanya gyerekükhöz szóló spontán beszédében azt találták, hogy a szünet a megnyilatkozáshatárokon hosszabb volt, mint a megnyilatkozásokon belül. Továbbá a tisztán szintaktikai alapon meghatározott tagmondathatárokon megjelenő szünetek is hosszabbak voltak, mint a szavak határainál megjelenőknek az időtartama (Soderstrom et al., 2008). Felnőttekhez szóló beszédet azonban ebben a tanulmányban sem vizsgáltak. Egy másik vizsgálatban japán édesanyák gyermekükhöz szóló spontán beszédében jelölték a prozódiai határokat (*intonational phrase: IP, accentual phrase: AP*) a több nyelvre kidolgozott ToBI-rendszer alapján, és összevetették a szünetek időtartamával (Ludusan et al., 2016). A dajkanyelvben a szünetek kimutathatóan hosszabbak voltak frázishatáron, mint nem frázishatáron, viszont a felnőttekhez szóló beszédben nem találtak ilyen eltérést, amikor a szünetidőtartamok alsó korlátjának a 300 ms-ot tekintették. Amikor azonban a 300 ms alatti szüneteket is bevonták a vizsgálatba, a felnőttekhez szóló beszédben is kimutatható volt a frázishatár hatása a szünetek időtartamára. Továbbá a szünetek időtartama nagyobb mértékben tért el a frázishatáron a nem frázishatárhoz képest a gyermekekhez szóló beszédben, mint a felnőttekhez szólóban. Ugyanezt az eredményt kapták az összes szünet figyelembevételével is. Habár a dajkanyelvben a szünetek megnyilatkozás-, tagmondat- és frázishatáron történő megjelenéséről meglehetősen kevés adat áll rendelkezésre, összefoglalva elmondható, hogy a megnyilatkozáshatár, tagmondathatár és frázishatár is egyértelműen jelölt a

szünet hossza által a gyermekekhez szóló beszédben. Továbbá a tagolás más mértékű hatást válthat ki a különböző regiszterekben, ugyanis nagyobb időtartambeli különbséget mutattak ki a frázishatáron és a nem frázishatáron lévő szünetek között a gyermekekhez szóló beszédben a felnőttekhez szólóhoz képest.

1.2. Magyarra vonatkozó dajkanyelvi eredmények

Az elmúlt években egyre nagyobb figyelem irányult a dajkanyelv magyar beszédbeli tulajdonságaira (pl. Gergely et al., 2017; Mády et al., 2018, 2021; Mády et al., 2022; Deme et al., 2019; Kohári et al., 2020a,b, 2022; Harmati-Pap et al., 2021). A morfológiai és szintaktikai elemzések megmutatták, hogy más nyelvekhez hasonlóan a magyar anyanyelvűek szintaktikai alapon meghatározott megnyilatkozásaiban (egyszerű mondatokban és alárendelt, összetett mondatokban) a szavak száma átlagosan kevesebb a gyermekekhez szóló beszédben, mint a felnőttekhez szólóban. Abban viszont nem tért el egymástól a két regiszter, hogy hány tagmondatot tartalmazott egy megnyilatkozás (Harmati-Pap et al., 2021). A magyar anyanyelvűek dajkanyelvét vizsgáló akusztikai elemzések azt is megmutatták, hogy a magyar anyanyelvű édesanyák – hasonlóan más anyanyelvűekhez – lassabban olvasnak fel mondatokat saját gyerekeiknek, mint más felnőtteknek (Kohári et al., 2019). Szintén felolvasott mondatokon azt is feltárták, hogy frázishatáron (IP, AP) gyakrabban jelenik meg szünet, mint nem frázishatáron, ugyanakkor a regiszterek között nem találtak különbségeket, és a szünetek időtartama sem tért el a gyermekekhez és a felnőttekhez szóló beszédben (Szalontai, 2018). Ez ellentmondani látszik annak a japán nyelvre kapott eredménynek, hogy nagyobb lenne a különbség a frázishatáron és a nem határon lévő szünetek között a dajkanyelvben. Ám ez utóbbi vizsgálat spontán beszédet elemzett, így elképzelhető, hogy a beszédmódból fakad az eltérés (Ludusan et al., 2016). Másfelől a magyar anyanyelvű édesanyák dajkanyelvét vizsgáló akusztikai méréseket a beszélők regiszterenként 3-3 felolvasott mondatán végezték (Szalontai, 2018), amely lehetővé tette a szerkezethatárok precíz összehasonlítását, de a szünetezési stratégiák elemzésére csak korlátozott lehetőséget biztosított. A

tagmondathatárok és a szünetezés összefüggéseit még nem vizsgálták magyar anyanyelvűek dajkanyelvében.

Jelen elemzés célja, hogy szisztematikusan megvizsgálja a dajkanyelvi beszéd tagolódásának időszerkezetét, a szünetek időtartamát és a gyakoriságát, összevetve a tagmondathatárokkal. A magyar beszédből származó nagymennyiségű adaton szeretnénk tanulmányozni azt, hogy a beszédszakaszok és a tagmondatok artikulációs tempója, beszédidőtartama és szótagszáma vajon eltér-e beszédregiszterenként, és a különböző tagolás eltérő eredményre vezet-e a regiszterek tekintetében. Továbbá elemezni szeretnénk a szünetek időtartamát, valamint gyakoriságát regiszterenként oly módon, hogy figyelembe vesszük a szünetek tagmondatokhoz képesti előfordulását.

2. Kísérleti személyek, anyag és módszer

Vizsgálatunkhoz *A korai nyelvfejlődés neuro-kognitív előjelzői* című projekt (NKFI-115385) során létrejött dajkanyelvi adatbázist használtuk fel. 14 egy-ermekes édesanya hangfelvételét választottuk ki az adatbázisból. Az édesanyákat a budapesti Magyar Honvédség Egészségügyi Központ Szülészeti-nőgyógyászati Osztályán toborozták szüléskor, de ettől eltekintve a mintavételezés véletlenszerű volt. Végzettségük tekintetében legalább középfokú végzettséggel rendelkeztek. Az édesanyák mindegyike Pest megyében vagy Budapesten lakott, nyelvi vagy beszédzavart nem diagnosztizáltak náluk. Átlagosan 31,57 évesek (24–41 év közöttiek) voltak a toborzáskor.

Az édesanyák gyermekükkel együtt a baba 6 hónapos korában az akkori MTA Természettudományi Kutatóközpont Kognitív Idegtudományi és Pszichológiai Intézetének laborjában stúdióköörülmények között vettek részt a következő kísérletben. A hangfelvételek elkészítéséhez szuperkardiod, kondenzátoros fejmikrofont használtak (Beyerdynamic TG H74c), és egy PC és egy M-Audio kétcsatornás USB-s külső hangkártya segítségével rögzítették 44,1 kHz-es mintavételezéssel, 16 biten digitalizálva. Egy képekkel teli, manókról szóló mesekönyv (1. ábra) történetét kellett elmesélniük úgy, hogy a mesekönyv lapjain szereplő

mondatokat is bele kellett szőniük a történetbe, a szöveg nélküli részeknél pedig saját szavaikkal kellett megfogalmazniuk a képen látható eseményeket. Jelen vizsgálatban a dajkanyelv tagolódásának vizsgálatához csak a szabadon elmondott, félszpontán szövegrészeket használtuk fel. A szünetek vizsgálatakor pedig csak azokat elemeztük, amelyek két spontánbeszéd-részletet választottak szét. A történet kétféle kondícióban hangzott el. Elsőként az édesanyák a kísérletet vezető felnőtt embernek (felnőttekhez szóló beszéd: AD) mondták el a történetet, majd a saját gyermeküknek (dajkanyelv: ID). Habár a gyermek 0, 6 és 18 hónapos korában is készültek felvételek, jelen vizsgálatban csak a 6 hónapos kori felvételeket elemeztük. Ebben az életkorban ugyanis már kimutatható volt német, francia és holland anyanyelvű babák esetében is, hogy a beszéd temporális sajátosságai, főként a szünetezés kulcsfontosságú szerepet tölt be a beszédfeldolgozásban, a frázisokra tagolásban (Johnson & Seidl, 2008; van Ommen et al., 2020).



1. ábra. Részlet a *Manóbújócska* című mesekönyvből.

A hanganyag szövegét kézzel lejegyeztük, jelöltük a beszédszakaszok és tagmondatok határait, és a bennük elhangzott szöveget. Manuálisan meghatároztuk a szünetek határait, és jelöltük a megakadásokat és a kitöltött szüneteket a *Praat 6.1.08* szoftverben (Boersma & Weenink, 2019). Egy script segítségével megmértük a beszédszakaszok, azaz a szünettől szünetig terjedő szakaszok időtartamát, megszámoltuk a bennük lévő magánhangzókat, és az így kapott szótagszámot elosztottuk a másodpercben mért beszédidőtartammal, hogy meg-

kapjuk az artikulációs tempót. A szünetek minimum időtartamának a 30 ms-ot tekintettük, és az így számolt adatokat közöljük, de kiszámoltuk úgy is a beszédszakaszok időtartamát, artikulációs tempóját és szótagszámát, hogy 300 ms-os volt a szünetek időtartamának a határa. Ez a mérismódszertan hasonló eredményekhez vezetett, mint mikor a szünetidőtartamnak az alsó korlátja alacsonyabb volt. A tagmondatok időtartamát, szótagszámát és artikulációs tempóját is kiszámoltuk, amely során szintén nem vettük figyelembe a szünetek időtartamát az egységek időtartama és az artikulációs tempójának kiszámítása során.

A tagmondathatárokat spontán beszédben az alapján határoztuk meg, hogy az adott rész egy alany-állítmányi viszonyal rendelkező szövegegység legyen (hasonlóan Soderstrom et al., 2008; Szaszák, 2008; Kohári, 2018). A spontán beszédre jellemző hiányos tagmondatokat, illetve a szervesen mondatrészeket ahhoz a tagmondathoz kapcsoltuk, amelyhez szünet nélkül illeszkedett. Amennyiben teljes tagmondatot nem alkotó szó vagy szavak előtt és után is szünet következett, a jelentésben hozzájuk tartozó tagmondatokhoz kapcsoltuk, amennyiben pedig a jelentés alapján nem volt egyértelműen eldönthető a hovatartozás, akkor időben a legközelebb esőhöz kapcsoltuk. Minden tagmondatot jelöltünk, függetlenül a tagmondat típusától. Ezt a megközelítést az is indokolta, hogy a tagmondat alárendelő vagy mellérendelő volta és a szünetek megjelenése között nem találtak közvetlen összefüggést a magyar beszédben (Krepsz et al., 2021).

Megmértük a szünetek időtartamát is. Mivel azonban a dajkanyelvi szakirodalomban igen gyakran csak a 300 ms feletti szünetidőtartamokat vizsgálták (Stern et al., 1983; Fernald & Simon, 1984; Grieser & Kuhl, 1988; Fernald et al., 1989), ezért kétféleképpen elemeztük az adatokat. Az egyik esetben a szünet alsó határának a 30 ms-ot tekintettük, amely hasonló mértékű a beszédhangoknak a spontán beszédben tipikusan előforduló legrövidebb időtartamrealizációihoz. A másik esetben csak a 300 ms feletti szünetek adatait nyertük ki. A 300 ms alatti szünetek vizsgálatát az indokolja, hogy nincs egységes álláspont arról, hogy hol is kellene meghúzni az alsó határt, ha egyáltalán szükséges ilyen minimumot bevezetni manuális mérések során (Trouvain & Werner, 2022). A kisebb

szünetek vizsgálatát az is motiválta, hogy nyelvenként is rendkívül eltérő lehet, hogy a szünetek hány százaléka fordul elő például 200 ms alatt (Campione & Véronis, 2002). A zöngétlen zárhangok, affrikáták zárszakaszának az idejét nem tekintettük szünetnek, szakaszhatáron egységesen 50 ms-ban határoztuk meg a zárszakaszok idejét (hasonlóan Markó, 2005).

A szünetek időtartama mellett a gyakoriságukat is megvizsgáltuk. Egy beszélő egy regiszterben történő beszédében előforduló szünetek számát elosztottuk a beszéd teljes idejével (hasonlóan Menn et al., 2022). Továbbá kiszámoltuk, hogy egy édesanya egy regiszterben történő beszédének szünetei milyen gyakran fordulnak elő tagmondathatáron, és milyen gyakran tagmondatok belsejében. A szünetek tagmondathatáron, illetve tagmondaton belüli előfordulási arányát százalékban adjuk meg a 3. fejezetben, ahol a 100% az összes szünetelőfordulás volt. Továbbá a történetmesélés szünetgyakoriságát 100 szótagra vetítve is közöljük. Meghatároztuk azt az arányt is, hogy egy beszélő egy regiszterében a tagmondatok hány százalékát követi szünet, illetve azt is, hogy a tagmondatok hány százalékát előzi meg szünet.

Mivel a kutatás célja az volt, hogy összehasonlítsuk a két regiszter tagmondatainak tempóját és néma szüneteinek időtartamát, ezért csak az olyan tagmondatokat és az azokban lévő néma szüneteket vettük bele a vizsgálatokba, amelyek nem tartalmaztak megakadásokat (téves újramezdések, hezitációk). Erre azért volt szükség, mert a megakadás befolyásolhatja a szünetek időtartamát (vö. Horváth, 2014), így mi csak olyan eseteket igyekeztünk vizsgálni a regiszterek összehasonlításához, amelyek esetében ez a külső tényező feltehetően kizárható. Tehát a mintavételezésnél a mesének csak a megakadás nélküli tagmondait, és azok szüneteit vettük figyelembe. A vizsgálataink kiterjedtek arra a kérdésre is, hogy milyen gyakoriak a szünetek a mese spontán részében, és milyen gyakran fordulnak elő tagmondathatáron. A megakadásoknak, illetve az azt követő és/vagy megelőző néma szüneteknek is lehet tagoló funkciója, ezért ezen kérdések esetében kétféleképpen számoltuk ki az eredményeket: megakadásokkal, illetve megakadások nélkül. Mivel a vizsgálat fókusza nem a tervezési nehézségek jelenségein, hanem a tagolás és a szintaktikai szervezés viszonyán van,

ezért a néma szünetek adatait részletesen ismertetjük a 3. fejezetben, és röviden kitérünk a megakadásokkal együtt számolt értékekre is. A tagmondathatárokon megvalósuló szünetek gyakoriságának vizsgálatakor néhány esetben (14 előfordulás) a tagmondathatáron a megakadás néma szünettel együtt realizálódott, a következő fejezet részletes bemutatásában ezeket az ún. kombinált szüneteket (Bóna, 2013) úgy vettük figyelembe, hogy a tagmondathatár szünettel jelölt. Ugyanakkor az elemzéseket ezen megakadásokkal együtt megvalósuló szünetek adatai nélkül is lefuttattuk, és ebben az esetben sem befolyásolták az eredményeket. Habár a meseolvasás rögzítése előtt az édesanyák áttanulmányozták a mesekönyvet, és az újszülött babájuknak is elmondták már a mesét korábban, a kísérleti elrendezés több édesanyának problémát okozott a tekintetben, hogy nehezen tudták a mondatokat beleszólni a képek alapján elmesélt történetbe, ezért mesénként sok megakadást produkáltak. Az eredetileg tervezett 20 édesanya helyett ezért csak 14-et vizsgáltunk meg, akiknek történetmondásában 8-nál kevesebb megakadás fordult elő, és így lényegében fogalmazási nehézségek nélkül tudták elmondani a történetet.

A kinyert függő változókra, amikor lehetséges volt, kevert modelleket alkalmaztunk az *R* 3.6.2 szoftverben (Team, 2022) az *lmerTest* csomag segítségével (Bates et al., 2015; Kuznetsova et al., 2017). A közölt modellek kiválasztásához a legbővebb modellből indultunk ki, és addig egyszerűsítettük, amíg a modell nem mutatott szignifikáns eltérést. Részletesen a 3. fejezetben tárgyaljuk, hogy a modellállítás során kapott legszűkebb modell milyen függő változókat, milyen interakciókat és random hatást tartalmazott. Irányelveként igyekeztünk random meredekséget (*slope*) használni a modellekben, amennyiben a modell nem konvergált, random konstanst (*intercept*) használtunk. A modellek *p*-értékeit a Satterthwaite approximáció segítségével nyertük ki (Kuznetsova et al., 2017), a különböző modelleket pedig az *anova()* függvénnyel hasonlítottuk össze (hasonlóan Winter, 2019). Független változóként faktor és numerikus változókat is használtunk a kevert modellekben (Bates et al., 2015). A függő változó típusának megfelelően lineáris kevert modelleket vagy általánosított változatukat használtuk. A beszédszakaszok szótagszáma esetében a Poisson-modell jelentős

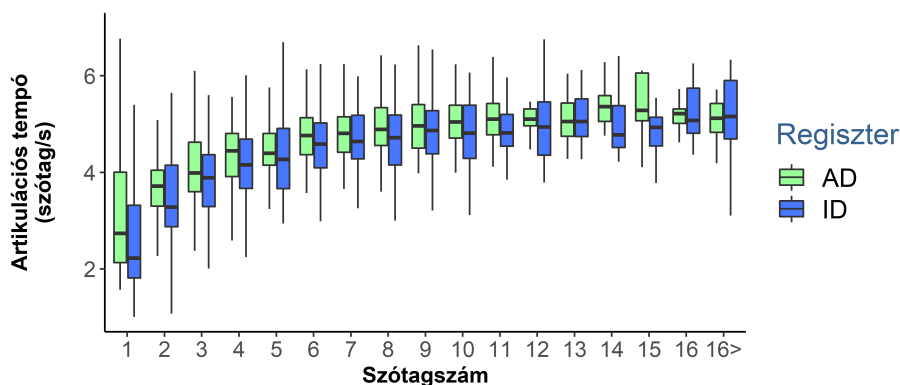
túlszóródása miatt (Payne et al., 2017) végül negatív binomiális regressziós kevert modellt alkalmaztunk. A szünetek időtartama esetében az erősen jobbra ferde eloszlás kezelésére a legjobb megközelítésnek az bizonyult, hogy vettük az értékek természetes logaritmusát, és lineáris kevert modellt futtattunk. Minden modell esetében ellenőriztük a reziduálisok eloszlását. A post hoc tesztek (Tukey-féle post hoc teszt) az *emmeans* csomag segítségével végeztük el (Lenth et al., 2019). A mixed modelleket és a post hoc tesztek eredményeit úgy közöljük, hogy mindig a felnőttekhez szóló beszédhez viszonyítjuk a dajkanyelvet. A kondicionális és marginális R^2 értékek becslését a *MuMIn* csomag segítségével határoztuk meg (Bartoń, 2022). A beszélőnként és regiszterenként számolt szünetgyakoriságok és előfordulási százalékos arányok esetében teszteltük a normális eloszlást, és az eredménynek megfelelően alkalmaztunk páros t -próbát vagy páros Wilcoxon-tesztet, és kiszámoltuk a próbának megfelelő hatásnagyságot is (Navarro, 2021; Mangiafico, 2022). A Cohen's D értékét úgy kapjuk meg, hogy a két csoport átlagát kivonjuk egymásból, és elosztjuk a különbségek szórásával, amelyet az *lsr* csomag *cohensD()* parancsának *method = "paired"* beállításával hajtottunk végre (Navarro, 2021). Amennyiben a mérőszám értéke 0,2 és 0,5 közé esik, a hatásnagyságot kicsinek tekintjük. 0,5 és 0,8 közötti érték esetén közepes, 0,8-nál nagyobb érték esetén nagy hatásnagyságról beszélhetünk (Mangiafico, 2022).

3. Eredmények

3.1. A beszédszakaszok és tagmondatok időszerkezete

Elsőként a szünettől szünetig terjedő szakaszok (másképpen beszédszakaszok) időzítését és szótagszámát vizsgáltuk meg. A beszédszakaszok artikulációs tempója átlagosan 4,60 szótag/s volt a felnőttekhez szóló beszédben (2. ábra), és alacsonyabb, 4,40 szótag/s volt a dajkanyelvben (a szórás AD: 1,13 szótag/s, ID: 1,18 szótag/s). A beszédszakasz artikulációs tempójának vizsgálatakor az a lineáris kevert modell bizonyult a legszűkebbnek, amelyben a regiszter és a beszédszakasz szótagszáma is szerepelt független változóként, interakciójuk

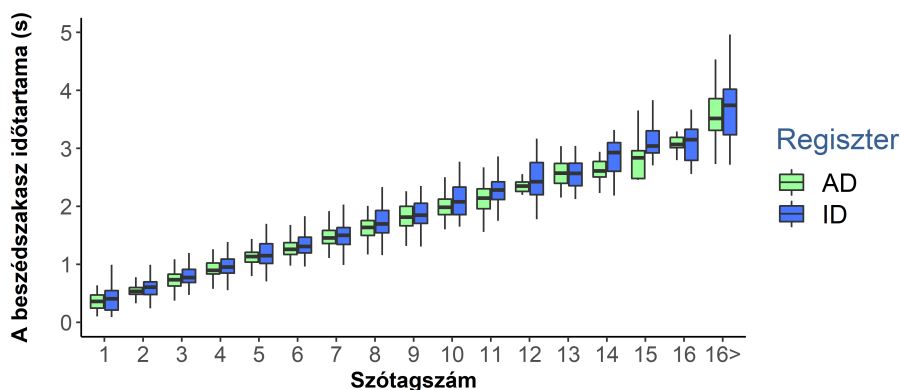
azonban nem. A modell a beszélőnkénti random konstanst tartalmazta. A lineáris kevert modell alapján a regiszter valóban hatással volt az eredményekre ($\beta = 0,17$, $SE = 0,05$, $t = 3,36$, $p < 0,001$). Továbbá a beszédszakaszokban minél kisebb volt a szótagszám, annál lassabb volt az artikulációs tempó ($\beta = 0,10$, $SE = 0,01$, $t = 17,33$, $p < 0,001$). A független és random hatások magyarázó ereje elfogadható mértékű volt (marginális $R^2 = 14,89\%$, kondicionális $R^2 = 24,72\%$), de az is látszott, hogy a beszédszakaszok tempója igen változatos volt attól függetlenül is, hogy ki és kihez beszél.



2. ábra. A beszédszakaszok artikulációs tempója a gyermekekhez és a felnőttekhez szóló beszédben a szótagszám függvényében. (A 16 fölötti szótagszámú beszédszakaszok egy csoportot alkotnak 16> jelöléssel.)

A beszéd tagolódásának alaposabb megértése céljából az alábbiakban részletezzük a szünettől szünetig terjedő szakaszok mért időtartamára és szótagszámára kapott eredményeket is. A beszédszakaszok időtartama átlagosan hosszabb volt dajkanyelvben (átlag: $1,47 \pm 0,84$ s), mint felnőttekhez szóló beszédben (átlag: $1,39 \pm 0,83$ s). A végső lineáris kevert modell a regisztert és a szótagszámot tartalmazta független változóként és beszélőnkénti random meredekséget a szótagszámra. Ahogyan az várható volt, a nagyobb szótagszám együtt járt a hosszabb beszédszakasszal ($\beta = 0,18$, $SE = 0,01$, $t = 44,83$, $p < 0,001$), de a regiszter hatása is kimutatható volt, a dajkanyelvben ugyanis a beszédszakaszok időtartama nagyobb volt, mint felnőttekhez szóló beszédben ($\beta = -0,07$, $SE =$

0,01, $t = -5,23$, $p < 0,001$). A szótagszám és a regiszter nem mutatott interakciót egymással, bármilyen modellt állítottunk fel. A marginális $R^2 = 88,95\%$, a kondicionális $R^2 90,96\%$ volt. A beszédszakaszok mért időtartamára tehát a szótagszámon túl a regiszter önmagában is kimutatható hatással volt (3. ábra), a szünettől szünetig terjedő szakaszok időtartama a gyermekekhez szóló beszédben hosszabbnak mutatkozott a felnőttekhez szóló beszédhez képest, amennyiben a modellben figyelembe vettük a beszédszakasz szótagszámát is.



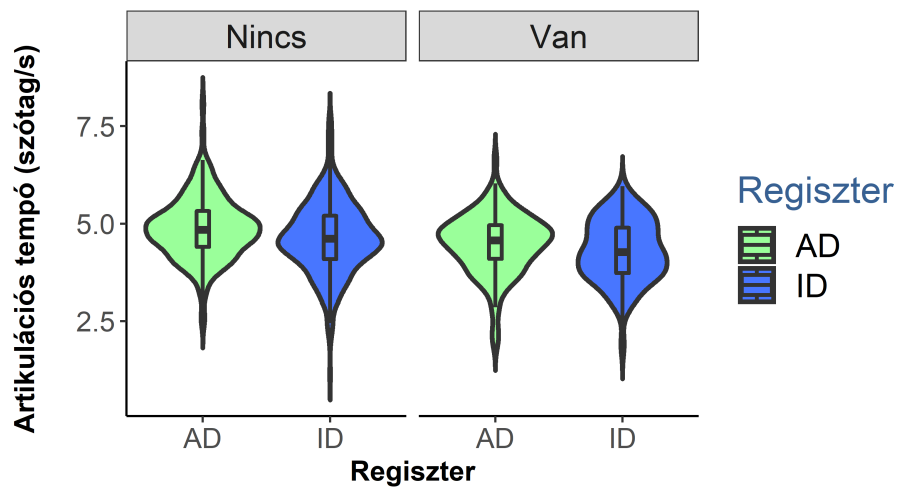
3. ábra. A beszédszakaszok időtartama regiszterenként a szótagszám függvényében. (A 16 fölötti szótagszámú beszédszakaszok egy csoportot alkotnak 16> jelöléssel.)

Felmerül a kérdés, hogy a szünettől szünetig terjedő szakaszok szótagszáma vajon eltér-e a gyermekekhez és a felnőttekhez szóló beszédben, ezért megvizsgáltuk függő változóként is. A dajkanyelvben átlagosan $6,59 \pm 4,28$ szótagból állt egy beszédszakasz, míg a felnőttekhez szóló beszédben átlagosan valamivel kevesebb, $6,56 \pm 4,31$ szótag alkotta. Az általánosított lineáris kevert modell (GLMM) a regisztert tartalmazta független változóként és a beszélőket random konstansként. A regiszter nem volt szignifikáns hatással a beszédszakaszok szótagszámára az általánosított kevert modellek szerint ($\beta = -0,01$, $SE = 0,03$, $z = -0,39$, $p = 0,70$). A modell magyarázó ereje rendkívül alacsony volt (marginális $R^2 < 0,1\%$ a kondicionális $R^2 = 3,52\%$). A beszédszakaszok szótagszámát tehát más tényezők erőteljesen befolyásolják, és a regiszter hatása nem volt kimutatható az adatokban. Összefoglalva a szünettől szünetig terje-

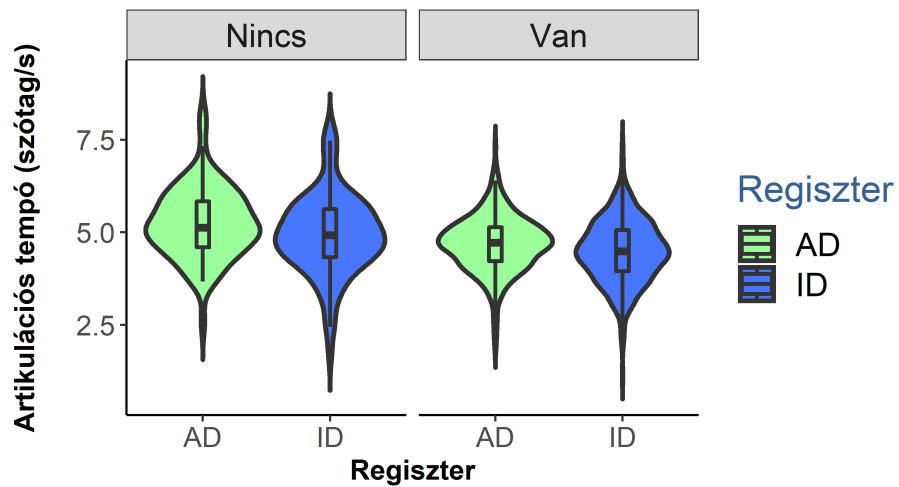
dő szakaszok akusztikai eredményeit, elmondható, hogy a gyermekekhez és a felnőttekhez szóló beszéd tehát hasonló szótagszámú beszédszakaszokra tagolódik, de a beszédszakaszok tempója lassabb dajkanyelvben. Mivel a tempó a dajkanyelvben lassabb, és hasonló szótagszámúak a szakaszok a két regiszterben, a beszédszakaszok időtartama is valamivel hosszabb a felnőttekhez szóló beszédhez képest.

Nemcsak a beszédszakaszok, hanem a tagmondatok időzítését is megvizsgáltuk. A tagmondatok artikulációs tempója átlagosan $4,79 \pm 0,81$ szótag/s volt a felnőttekhez szóló beszédben, és $4,56 \pm 0,89$ szótag/s volt a dajkanyelvben. A lineáris kevert modell független változói között szerepelt a szótagszám, a regiszter (ID vagy AD), valamint az, hogy a tagmondatot követte-e szünet (igen vagy nem), és hogy megjelent-e benne szünet (igen vagy nem). A modellre nem volt hatással, hogy a tagmondat előtt volt-e szünet vagy sem. A modell a beszélőnkénti random meredekséget is tartalmazta a regiszterre. Az eredmények alapján a regiszter valóban hatással volt az artikulációs tempóra ($\beta = 0,20$, $SE = 0,05$, $t = 4,36$, $p = 0,002$). A szótagszám és a követő szünet megjelenése, illetve hiánya, valamint a tagmondaton belüli szünetmegjelenés is befolyásolta a tempóértékeket (4. ábra), de mindkét regisztert azonos irányban és hasonló mértékben, mivel ezen tényezők között nem találtunk interakciót. Minél inkább kevesebb szótagból állt egy tagmondat, annál inkább jellemző volt, hogy alacsonyabb volt a tagmondat artikulációs tempója ($\beta = 0,07$, $SE = 0,01$, $t = 12,86$, $p < 0,001$). A szünet tagmondat utáni megjelenése együtt járt az alacsonyabb tempóértékekkel ($\beta = -0,52$, $SE = 0,06$, $t = -9,20$, $p < 0,001$), a tagmondaton belüli szünet megjelenése esetén a tagmondatoknak lassabb volt az artikulációs tempója, mint a szünet nélkül megvalósult tagmondatoknak ($\beta = -0,64$, $SE = 0,05$, $t = -12,19$, $p < 0,001$). A függő és random hatások magyarázó ereje elfogadható mértékűek voltak (marginális $R^2 = 17,00\%$, kondicionális $R^2 = 37,56\%$). A beszédszakaszok artikulációs tempójához hasonlóan a gyermekekhez szóló beszéd tagmondatai szintén lassabbak a felnőttekhez szóló beszédhez képest. Habár az artikulációs tempó kiszámításánál a definíció alapján nem vettük figyelembe a szünet időtartamát, mégis a tagmondaton megjelenő

szünet vagy az azt követő szünet hatással volt a tagmondat mért tempóértékére (4. és 5. ábra), de ez az összefüggés mindkét regiszterben ugyanúgy valósult meg.



4. ábra. A tagmondat artikulációs tempója, amikor szünet nélkül realizálódik (Nincs), és amikor szünettel valósul meg a tagmondat határain belül (Van).



5. ábra. A tagmondat artikulációs tempója, amikor a tagmondatot nem követi szünet (Nincs), és amikor a tagmondatot szünet követi (Van).

A tagmondatok szótagszámát és szünet nélküli időtartamát is megvizsgáltuk a beszédszakaszokhoz hasonlóan. A tagmondatok időtartama átlagosan nagyobb volt a dajkanyelvben (átlag: $1,89 \pm 1,06$ s), mint a felnőttekhez szóló beszédben (átlag: $1,78 \pm 1,06$ s). A lineáris kevert modell független változói között szerepelt a szótagszám, a regiszter (ID vagy AD), valamint az, hogy a tagmondatot követte-e szünet (igen vagy nem), és hogy megjelent-e benne szünet (igen vagy nem). Az artikulációs tempóhoz hasonlóan a modellhez nem adott hozzá az, hogy a tagmondatot megelőzte-e szünet vagy nem. A modell a beszélőnkénti random meredekséget is tartalmazta a regiszterre. Ahogyan az várható volt, a nagyobb szótagszám itt is együtt járt a hosszabb tagmondatokkal ($\beta = 0,19$, $SE = 0,01$, $t = 27,11$, $p < 0,001$). Emellett a regiszter hatása is kimutatható volt, a dajkanyelv tagmondatainak időtartama ugyanis tendenciózusan nagyobb volt, mint a felnőttekhez szóló beszédben realizálódó tagmondatoké ($\beta = -0,08$, $SE = 0,02$, $t = -4,79$, $p < 0,001$). A tagmondat után megjelenő szünet esetében a tagmondat időtartama nagyobb volt, mint amikor nem követte szünet ($\beta = 0,15$, $SE = 0,02$, $t = 6,21$, $p < 0,001$). A tagmondat időtartama szintén nagyobb volt, amikor a tagmondaton belül megjelent a szünet, mint mikor nem ($\beta = 0,31$, $SE = 0,02$, $t = 13,62$, $p < 0,001$). A modellben a marginális $R^2 = 87,11\%$, míg a kondicionális $R^2 = 91,10\%$ volt. A beszédszakaszokhoz hasonlóan a tagmondatok időtartama is hosszabb volt dajkanyelvben. Habár a szünetek megjelenése a tagmondatban és azt követően kimutathatóan befolyásolták a tagmondat időtartamának értékét, a különböző regiszterekben ugyanolyan irányban hatottak.

A tagmondatok szótagszáma hasonlóan alakult a beszédszakaszok szótagszámához a regiszter hatása szempontjából. Habár a dajkanyelv tagmondatainak szótagszáma átlagosan valamivel nagyobb volt ($8,48 \pm 4,50$) a felnőttekhez szóló beszédhez képest ($8,36 \pm 4,46$), az általánosított lineáris kevert modell (GLMM) nem mutatta ki a regiszter hatását a tagmondatonként számolt szótagok darabszámára. Az alkalmazott általánosított, beszélőnkénti random konstanssal felállított kevert modell három külön független változóként tartalmazta azt, hogy a szünet a tagmondat előtt megjelent-e vagy sem, illetve a szünet

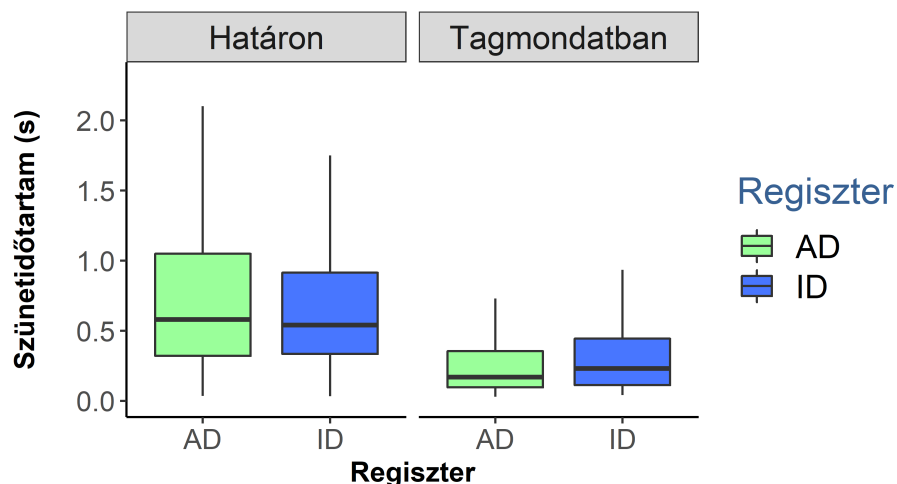
tagmondat után megjelent-e vagy sem, és hogy tagmondaton belül megjelent-e vagy sem. A szótagszám függő változóként szerepelt a modellben. Nagyobb volt a szótagok száma, amennyiben szünet realizálódott bármelyik pozícióban, mint amikor ugyanott nem jelent meg szünet (tagmondaton belül: $\beta = 0,49$, $SE = 0,03$, $z = 16,98$, $p < 0,001$; tagmondat előtt: $\beta = 0,14$, $SE = 0,04$, $z = 3,85$, $p < 0,001$; tagmondat után: $\beta = 0,17$, $SE = 0,04$, $z = 4,39$, $p < 0,001$). A marginális $R^2 = 21,20\%$, a kondicionális $R^2 = 22,90\%$ volt. A szótagok száma tehát sem beszédszakaszonként, sem tagmondatonként nem tért el kimutathatóan egymástól a két regiszterben, a tagmondat előtt, után és közben megjelenő szünetek viszont összefüggést mutattak a tagmondatokban lévő szótagok számával.

3.2. A szünetek időtartama és gyakorisága a kétféle regiszterben

A folyamatos beszéd tagolásában kulcsszerepet játszik a szünetek előfordulása, ezért megvizsgáltuk, hogy eltér-e gyakoriságuk, illetve időtartamuk a két regiszterben. A szünetek időtartamának vizsgálatakor figyelembe vettük a környező beszédszakaszok időtartamát, azaz hogy mekkora beszédfolyamat tagolnak két részre, továbbá a tagmondathoz való viszonyukat, hogy tagmondat-határon szerepelnek vagy egy tagmondatnyi egységen belül. A néma szünetek időtartamát logaritmikus transzformáció után vizsgáltuk lineáris kevert modellel, amelyben a független hatás a regiszter volt, továbbá a megelőző és a követő beszédszakasz időtartamának az összege, valamint a szünet tagmondathatárhoz való viszonya (tagmondathatáron szerepelt vs. tagmondaton belül valósult meg). A megelőző és a követő beszédszakasz időtartamát azért adtuk össze, és építettük bele a modellbe numerikus változóként, mert arra a kérdésre kerestük a választ, hogy ha nagyobb beszédszakaszokat választ szét egy szünet, akkor vajon hosszabban realizálódik-e? A modell tartalmazta a beszélőket random konstansként, továbbá a regiszter és a tagmondathatárhoz való viszony interakcióját és a beszédszakaszok összesített időtartamának interakcióját a tagmondathatárhoz való viszonnal is. Az eredmények megmutatták, hogy a tagmondathoz való viszony önmagában is hatással van a szünet időtartamára ($\beta = -1,71$, $SE =$

0,15, $t = -11,04$, $p < 0,001$). A környező beszédszakaszok összesített időtartama és a regiszter azonban csak interakcióban mutatott összefüggést a szünetek időtartamával. A post hoc teszt azt mutatta, hogy a regiszter a tagmondathoz való viszonytal olyan összefüggésben volt, hogy a tagmondaton belül előforduló szünet időtartama hosszabb volt dajkanyelvben, mint a felnőttekhez szóló beszédben ($\beta = -0,20$, $SE = 0,09$, $t = -2,32$, $p = 0,02$). A mért, nyers időtartamok alapján is látszik az eltérés (6. ábra), a nem tagmondaton belül előforduló szünetek időtartamának mediánja nagyobb volt dajkanyelvben, mint a felnőttekhez szóló beszédben (ID: $234,99 \pm 37,27$ ms, AD: $170,55 \pm 32,48$ ms). A tagmondatok közötti szünetidőtartam esetében azonban nem sikerült eltérést kimutatni a regiszterek között. A szünetidőtartamokra szintén interakcióban volt hatással a tagmondathoz való viszony a szünetet megelőző és követő beszédszakaszok összesített időtartamával ($\beta = 0,33$, $SE = 0,11$, $t = 2,91$, $p = 0,004$). A post hoc teszt azt mutatta, hogy a tagmondatokon belül megvalósult szünetek időtartama annál hosszabb volt, minél nagyobb volt a megelőző és követő beszédszakasz időtartamának összege, míg a tagmondatok határában megvalósuló szünetek esetében nem volt ilyen összefüggés (tagmondaton belül: $\beta = 0,17$, $SE = 0,04$, $t = 4,04$, $p < 0,001$, tagmondathatáron: $\beta = -0,01$, $SE = 0,03$, $t = -0,07$, $p = 0,94$). A marginális $R^2 = 25,19\%$, a kondicionális $R^2 = 28,64\%$ volt. Mivel a szakirodalomban (Stern et al., 1983; Fernald & Simon, 1984; Grieser & Kuhl, 1988; Fernald et al., 1989) igen gyakran csak a 300 ms feletti szünetek időtartamát vizsgálták, figyelmen kívül hagyva a többi szünetet, ezért ily módon is lefuttattuk a kevert modellt. Ez a leválogatás 684 db szünetet eredményezett. A regiszter viszont nem mutatott összefüggést a szünetek időtartamával, sem önmagában, sem a tagmondathatárral való interakcióban (regiszter: $\beta = 0,05$, $SE = 0,05$, $t = 1,15$, $p = 0,27$; interakcióban $\beta = -0,01$, $SE = 0,08$, $t = -0,13$, $p = 0,90$). A modell magyarázó ereje igen alacsony volt (marginális $R^2 = 7,50\%$, a kondicionális $R^2 = 15,37\%$). Bármilyen modellt állítottunk fel, a szünetek alsó és felső korlátjának figyelembevételével hasonló eredményt kaptunk. Összefoglalva a regiszterrel kapcsolatos eredményeket, az összes szünet figyelembevételekor a szünetek időtartamát nagymértékben meg-

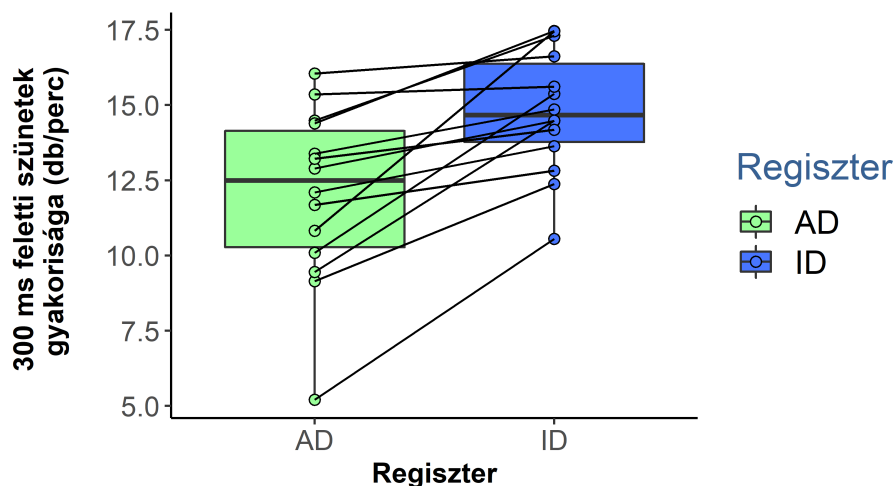
határozta, hogy tagmondathatáron vagy tagmondaton belül realizálódtak, és kimutatható volt, hogy a tagmondaton belül megvalósuló szünetek időtartama nagyobb volt a dajkanyelvben, mint a felnőttekhez szóló beszédben, miközben a regiszter a tagmondatok határán lévő szünetek időtartamára nem volt hatással.



6. ábra. A tagmondatok határán és a tagmondaton belül megjelenő szünetek mért időtartama (s) regiszterenként.

Azt is megvizsgáltuk, hogy a két regiszter spontán beszédében beszélőnként mennyi néma szünet található. A szünetek darabszámát úgy normalizáltuk, hogy elosztottuk a beszéd időtartamával, mivel a gyermekekhez tendenciaszerűen hosszabban beszéltek, mint a felnőttekhez. A gyermekekhez szóló beszédben valamivel gyakrabban fordultak elő szünetek, mint a felnőttekhez szóló beszédben (ID: $22,65 \pm 6,51$ db/perc, AD: $22,46 \pm 3,23$ db/perc), de a statisztikai próba nem támasztotta alá, hogy szignifikáns lenne az eltérés (páros Wilcoxon-teszt: $V(14) = 40$, $p = 0,46$). A szünetek gyakoriságát 100 szótagra vetítve is kiszámítottuk (ID: $12,20 \pm 3,02$, AD: $11,67 \pm 3,27$), a statisztikai próba eredményei alapján itt sem találtunk különbséget a két regiszterben (párosított t -próba: $t(13) = -1,68$, $p = 0,12$). Mivel a megakadásoknak, illetve az azt megelőző és követő szünetnek is lehet tagoló funkciója, ezért megvizsgáltuk, hogy a néma szünetek és a megakadások gyakorisága együttesen eltér-e a két

regiszterben, de ott is hasonló eredményeket kaptunk (percenkénti gyakoriság esetében: $V(14) = 47$, $p = 0,76$; 100 szótagra normalizált gyakoriság esetében: $t(13) = -1,07$, $p = 0,30$). Amikor viszont csak a 300 ms feletti néma szüneteket vettük figyelembe (7. ábra), akkor a statisztikai próba szignifikáns eltérést mutatott a két beszédmódban előforduló szünetgyakoriságok között (páros t -próba: $t(13) = -5,06$, $p < 0,001$; Cohen's D : 1,35). A dajkanyelvben beszélőnként statisztikailag is igazolhatóan nagyobb volt a 300 ms feletti szünetgyakoriság, mint a felnőttekhez szóló beszédben (ID: $14,80 \pm 2,05$ db/perc; AD: $12,02 \pm 2,90$ db/perc). Ugyanezt az összefüggést találtuk a 300 ms feletti néma szünetek 100 szótagra vetített gyakorisága esetében (ID: $7,85 \pm 1,02$; AD: $6,33 \pm 1,83$). A statisztikai próba ennél a mérőszámnál is szignifikáns eltérést jelzett (páros t -próba: $t(13) = -3,91$, $p = 0,002$; Cohen's D : 1,05).

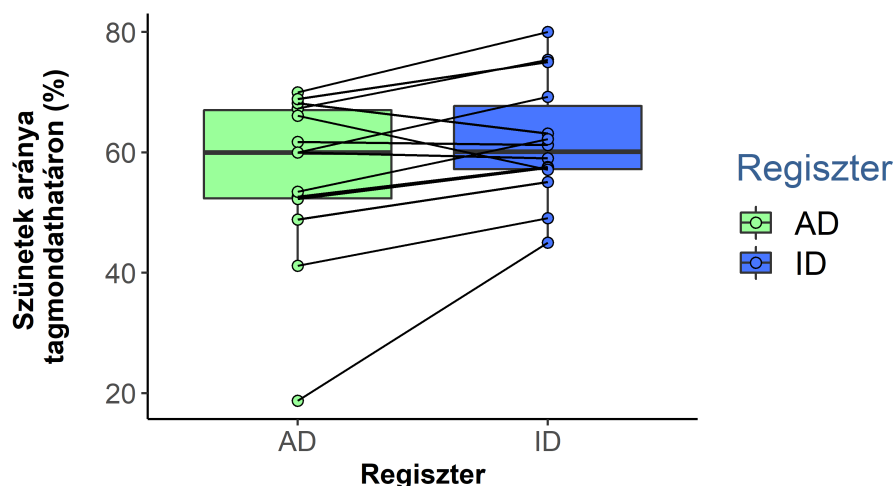


7. ábra. A 300 ms feletti szünetek gyakorisága beszélőnként a két regiszterben.

A szünetek gyakoriságával kapcsolatban azt is megvizsgáltuk, hogy a tagmondathatárok mennyire jelöltek a szünetek által, illetve a szünetek hány százaléka fordul elő tagmondathatáron vagy tagmondat belsejében, és hogy ezen gyakoriságok milyen összefüggést mutatnak a regiszterekkel. A tagmondatok gyakorisága, azaz a tagmondatok darabszáma elosztva az adott mese beszéd-időtartamával hasonlóan alakult a két regiszterben (ID: $22,06 \pm 3,15$ db/perc;

AD: $21,69 \pm 4,42$ db/perc). Nem volt statisztikailag kimutatható különbség a gyermekekhez és a felnőttekhez szóló beszéd között a beszélőnként párosított tagmondatgyakoriságban ($t(13) = -0,49$, $p = 0,63$). A szünetekkel történő határjelölést is megvizsgáltuk, és kiszámítottuk, hogy beszélőnként a különböző regiszterekben a tagmondatok hány százalékát követi szünet, illetve hány százalékát előzi meg szünet. A gyermekekhez szóló beszédben átlagosan a tagmondatok 86,10%-át követte szünet, míg a felnőttekhez szóló beszédben ez az arány 84,56% volt (szórás ID: 6,69; AD: 8,39). A két regiszter nem tért el lényegesen egymástól a statisztikai próba alapján ($t(13) = -0,70$, $p = 0,49$). A tagmondatot követő szünetekhez hasonló eredményeket kaptunk a tagmondatot megelőző szünetek esetében is. A tagmondatokat nagy százalékban megelőzte szünet mindkét regiszterben (ID: $85,76 \pm 6,67\%$; AD: $83,97 \pm 8,28\%$, és ez az arány a gyermekekhez és felnőttekhez szóló beszédben nem tért el szignifikánsan egymástól ($t(13) = -0,86$, $p = 0,40$). A tagmondatokat követő és megelőző szünetek százalékos aránya akkor sem tért el a két regiszterben, amikor a szünetnek 300 ms-os alsó határt szabtuk (megelőző szünet esetén: $t(13) = -0,14$, $p = 0,89$; követő szünet esetén: $t(13) = -0,15$, $p = 0,88$). A tagmondathatár és a szünetek előfordulásának elemzése során azt a kérdést is megvizsgáltuk, hogy egy beszélő egy meséjén belül a szünetek hány százaléka jelenik meg tagmondathatáron (8. ábra). A dajkanyelvben a szünetek nagyobb arányban jelentek meg tagmondathatáron, mint a felnőttekhez szóló beszédben (ID: $61,90 \pm 10,01\%$; AD: $56,39 \pm 13,85\%$). A regiszterek közötti eltérés statisztikailag is kimutatható volt ($t(13) = -2,50$, $p = 0,03$, Cohen's D : 0,67). Amikor csak a 300 ms feletti szüneteket vizsgáltuk, akkor azt az eredményt kaptuk, hogy a szünetek tagmondathatáron történő megjelenésének aránya a gyermekekhez szóló beszédben alacsonyabb volt, mint a felnőttekhez szóló beszédben (ID: $75,82 \pm 10,43\%$, AD: $81,63 \pm 11,81\%$). Az eltérés szintén szignifikáns volt a statisztikai próba alapján ($t(13) = 2,31$, $p = 0,04$, Cohen's D : 0,62). A tagmondathatárt tehát a beszélők mind a gyermekekhez szóló, mind a felnőttekhez szóló beszédben nagy arányban, következetesen jelölték szünetekkel, és a két regiszter nem tért el lényegesen egymástól ebben a tekintetben. Amikor azonban az összes szünetet

figyelembe véve megnéztük, hogy hány százalék valósul meg tagmondathatáron és hány százalék tagmondaton belül, akkor azt az eredményt kaptuk, hogy a szünetek nagyobb arányban jelennek meg a tagmondatok határán gyermekekhez szóló beszédben, mint felnőttekhez szóló beszédben.



8. ábra. A szünetek tagmondathatáron történő megjelenésének beszélőnkénti gyakorisága a két regiszterben

4. Következtetések

Jelen tanulmányban a dajkanyelv tagolódását és annak időzítésbeli sajátosságait, továbbá a szünetek időtartamát és gyakoriságát vizsgáltuk meg, összevetve a tagmondathatárokkal. Az eredményeink azt mutatták, hogy a szigorúan szintaxis alapján meghatározott tagmondathatárok a beszéd szakaszainak időzítésével hasonló módon függtek össze a gyermekekhez és a felnőttekhez szóló beszédben, a szünetek időtartama és gyakorisága ugyanakkor regiszterenként eltért a tagmondathatárok figyelembevételével. A korábbi szakirodalmi eredményekhez hasonlóan (pl. Fernald et al., 1989) az elemzett félspontán beszédben is kimutatható volt, hogy a dajkanyelv lassabb a felnőttekhez szóló beszédhez képest. A beszédegység nagysága pedig nem volt hatással a két regiszter közti

különbségre, mind a szünettől szünetig terjedő beszédszakaszok, mind a tagmondatok artikulációs tempója lassabb volt dajkanyelvben, mint felnőttekhez szóló beszédben. A beszédszakaszok és a tagmondatok szótagszáma viszont nem mutatott eltérést a regiszterek között. Mivel a két regiszterben nagyjából azonos szótagszámú beszédegységek voltak, és az artikulációs tempó alacsonyabb volt dajkanyelvben, így érthető, hogy a beszédszakaszok és a tagmondatok időtartama is valamivel hosszabb volt a gyermekekhez szóló beszédben, mint a felnőttekében. Ezek az eredmények ellentmondani látszanak azoknak a korábbi szakirodalmi eredményeknek, amelyek szerint a beszédszakaszok, megnyilatkozások időtartama rövidebb dajkanyelvben, mint felnőttekhez szóló beszédben (Fernald et al., 1989; Martin et al., 2016). Ugyanakkor fontos módszertani különbség, hogy ezekben az elemzésekben a beszélők a témamegjelölésen túl kötetlenül kommunikálhattak a gyermekükkel és a felnőttekkel, míg a jelen vizsgálatban a mondanivaló lényege képek által meghatározott volt, tehát a két regiszterben az átvinni kívánt információknak nagyon hasonlónak kellett lenniük. Egy vizsgálat kimutatta, hogy a szülő-gyerek interakcióban igen gyakran használt hangutánzó szavak előtt és után is gyakrabban használtak a szülők szüneteket, mint más szavak esetében (Laing et al., 2017). A szülő-gyermek szabad játék során történő interakcióban a felnőttek gyakran izolálnak egy-egy szót a beszéd többi részétől, amely elősegíti a szavak tanulását (Brent & Siskind, 2001). Ezek a jelenségek magyarázatul szolgálhatnak arra, hogy eredményeinkkel ellentétben más vizsgálatokban miért találták rövidebbeknek a beszédszakaszokat, megnyilatkozásokat időtartamukban és szótagszámukban is (Grieser & Kuhl, 1988; Fernald et al., 1989; Martin et al., 2016).

A beszédszakaszok és a tagmondatok tiszta beszédidejének vizsgálatán túl a szünetek időtartamát és gyakoriságát is elemeztük a dajkanyelvben. Az eredményeink sajátos összefüggést mutattak a különböző regiszterekben a szünetek időtartama és a tagmondathatárok között. A tagmondaton belül előforduló szünetek időtartama hosszabb volt a dajkanyelvben, mint a felnőttekhez szóló beszédben. A tagmondatok közötti szünetek időtartama azonban nem tért el lényegesen egymástól a két regiszterben. Mind a felnőttekhez szóló beszédben,

mind a gyermekekhez szólóban a tagmondathatáron lévő szünetek lényegesen hosszabbak voltak a tagmondathatáron belül megvalósuló szüneteknél, amely egybevág a korábbi szakirodalmi eredményekkel. Egyrészt a magyar beszédben, felnőttek közötti kommunikációban is kimutatható volt a tagmondathatár szünetidőtartamra gyakorolt hatása (Gyarmathy, 2017), másrészt két amerikai angol anyanyelvű édesanya gyermekükhöz szóló beszédében is hasonló összefüggést találtak, ugyanis a tagmondathatárokon megjelenő szüneteket hosszabbnak mérték, mint a nem tagmondathatáron előforduló szünetek időtartamát (Soderstrom et al., 2008). Habár ebben a tanulmányban a dajkanyelvet nem vetették össze a felnőttekhez szóló beszéddel, de egy másik vizsgálatban (Ludusan et al., 2016) a prozódiai határok és a szünet időtartamának kapcsolatát ebben a regiszterben is elemezték, és szintén hosszabbnak találták a frázishatárokon lévő szüneteket a frázison belül realizálódóknál. Ludusan és munkatársai (2016) azt is kimutatták, hogy a szünetek időtartama nagyobb mértékben tért el a frázishatáron a nem frázishatárhoz képest a gyermekekhez szóló beszédben, mint a felnőttekhez szóló beszédben. Ezt az összefüggést a jelen tanulmányban nem sikerült kimutatni, ugyanakkor hozzá kell tennünk, hogy ők prozódiai frázishatárokat vizsgáltak és nem tagmondathatárokat. A szakirodalomban a szünetek időtartamának regiszterek közötti különbségét gyakran a prozódiai vagy szintaktikai egységhatár vizsgálatának bevonása nélkül elemezték korábban (pl. Stern et al., 1983; Grieser & Kuhl, 1988; Fernald & Simon, 1984; Menn et al., 2022). Jelen tanulmány eredményei a korábbi frázishatárvizsgálatokkal együtt rávilágítanak arra (Ludusan et al., 2016), hogy a szünetek időtartamának modellezése nehezen képzelhető el prozódiai vagy szintaktikai egységhatárok vizsgálata nélkül. A szünetek időtartamának variabilitását még a tagmondathatárok figyelembevételével is csak korlátozott mértékben sikerült magyaráznunk modelljeinkkel, feltehetően a regiszterek vizsgálatakor további szünetidőtartamra ható tényezők bevonása lenne indokolt a jövőben.

A szünetek időtartama mellett a szünetek gyakoriságának elemzése során is találtunk eltéréseket a két regiszterben. Amikor az összes szünetelőfordulást figyelembe vettük függetlenül a szünet időtartamának hosszától, akkor a két

regiszter nem különbözött egymástól, amikor azonban csak a 300 ms feletti szünetek gyakoriságát vizsgáltuk, akkor a dajkanyelvben gyakoribbnak bizonyultak a szünetek a felnőttekhez szóló beszédhez képest, hasonlóan a korábbi szakirodalmi eredményekhez (Martin et al., 2016; Menn et al., 2022). Az eredményeinkből arra lehet következtetni, hogy nem általában a szünetek gyakorisága lenne nagyobb a dajkanyelvben, hanem bizonyos nagyságú és/vagy típusú szünetek előfordulása lenne gyakoribb. Ezért megvizsgáltuk azt is, hogy a szünetek hány százaléka fordul elő tagmondathatáron vagy tagmondat belsejében, és hogy ezen gyakoriságok milyen összefüggést mutatnak a regiszterekkel. A tagmondathatárt a beszélők mind a gyermekekhez szóló, mind a felnőttekhez szóló beszédben nagy arányban, következetesen jelölték szünetekkel, ebben a tekintetben a két regiszter nem tért el lényegesen egymástól. A szünetek ugyanakkor nagyobb arányban jelentek meg a szintaktikai alapon meghatározott tagmondathatáron dajkanyelvben, mint a felnőttekhez szóló beszédben, amikor az összes szünetet figyelembe vettük, de a tagmondathatárok jelöltsége nem volt egyértelműen kapcsolatba hozható a szünetek hosszával. A tagmondathatárokon megjelenő szünetarány-különbségek a két regiszterben hasonlóak a megnyilatkozásokhoz, illetve a tagmondatoknál kisebb, szintaktikailag egybetartozó egységekre kapott szakirodalmi eredményekhez (Broen, 1972; Fernald & Simon, 1984; Fisher & Tokura, 1996). A dajkanyelvi beszédprodukciónban kimutatott tagmondathatáron történő gyakoribb szünetmegjelenés a babák figyelmére is hatással lehet, és ezáltal segítheti a beszédfolyamat egységekre tagolását. Kimutatták ugyanis, hogy a babák szívesebben figyeltek olyan dajkanyelvi beszédre, amelyben a tagmondathatárokon jelentek meg a szünetek, mint olyan beszédre, amelyben a tagmondathatárokon belül fordultak elő (Hirsh-Pasek et al., 1987; Nelson et al., 1989). Habár a jelen vizsgálatban a szintaktikai alapon meghatározott tagmondathatárookra fókuszáltunk, a prozódiai frázishatárok tagolódása is összefüggéseket mutathat a szünetek időtartamával és gyakoriságával. A jövőben ezt a kérdést részletesebben is kívánjuk majd vizsgálni, hogy még átfogóbb képet kapjunk a dajkanyelvi tagolódás sajátosságairól.

Köszönetnyilvánítás

A vizsgálat *A korai nyelvfejlődés neuro-kognitív előjelzői* című projekt (NKFI-115385, vezető kutató: Winkler István) és *A dajkanyelv longitudinális vizsgálata multimodális módszerekkel* című projekt (NKFI-134775) keretében készült. Ezúton szeretnénk köszönetet mondani a kutatásban résztvevő gyerekeknek és szüleiknek, továbbá az adatgyűjtést végző munkatársaknak: Balázs Andreának, Haraszi Zsófiának, Horváth Blankának, Özéné Kende Líviának, Roschéné Farkas Juditnak, Radányi Annának, Szekeres Rebekának, Szerafin Ágnes Katának és Uhrin Beatrixnak.

Hivatkozások

- Bartoń, K. (2022). Mumin: Multi-model inference. <https://cran.r-project.org/web/packages/MuMIn/index.html>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer. <http://www.praat.org/>.
- Bóna, J. (2013). A beszédszünetek fonetikai sajátosságai a beszéd típus függvényében. *Beszéd kutatás 2013*, (pp. 60–75).
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, B33–B44.
- Broen, P. A. (1972). *The Verbal Environment of the Language-Learning Child. ASHA Monographs*, 17. Washington: American Speech and Hearing Association.
- Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel, & I. Marlien (Eds.), *Proceedings of the first international conference on speech prosody* (pp. 199–202). Aix-en-Provence, France: Institut de Phonétique d’Aix.

- Church, R., Bernhardt, B., Pichora-Fuller, K., & Shi, R. (2005). Infant-directed speech: Final syllable lengthening and rate of speech. *Canadian Acoustics*, 33, 13–19.
- Deme, A., Kohári, A., Mády, K., Reichel, U. D., & Szalontai, Á. (2019). A magánhangzós hosszúsági fonológiai kontraszt a dajkanyelvben a csecsemő életkorának függvényében. *Beszéd kutatás*, 27, 221–242.
- Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20, 104.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16, 477–501.
- Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, 67, 3192–3218.
- Gergely, A., Faragó, T., Galambos, Á., & Topál, J. (2017). Differential effects of speech situations on mothers' and fathers' infant-directed and dog-directed speech: An acoustic analysis. *Scientific Reports*, 7, 13739.
- Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, 24, 14–20.
- Gyarmathy, D. (2017). A néma szünetek funkciói a spontán beszédben. *Beszéd kutatás*, 25, 67–92.
- Harmati-Pap, V., Vadász, N., Kas, B., & Tóth, I. (2021). Anyai dajkanyelvi narratívák lexikai és szintaktikai jellemzőinek longitudinális vizsgálata. *Beszédtudomány – Speech Science*, 2, 207–242.

- Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269–286.
- Horváth, V. (2014). *Hezitációs jelenségek a magyar beszédben*. Budapest: ELTE Eötvös Kiadó.
- Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*, 13, 440–455.
- Ko, E. S., & Soderstrom, M. (2013). Additive effects of lengthening on the utterance-final word in child-directed speech. *Journal of Speech, Language, and Hearing Research*, 56, 364–371.
- Kohári, A. (2018). *Időzítés mintázatok a magyar beszédben*. Budapest: MTA.
- Kohári, A., Deme, A., Reichel, U. D., Á., S., & Mády, K. (2019). A dajkanyelv temporális jellemzői 4 és 8 hónapos csecsemőkhöz szóló beszédben. *Beszéd-kutatás*, 27, 243–258.
- Kohári, A., Deme, A., Reichel, U. D., Szalontai, Á., & Mády, K. (2020a). A beszédritmus időbeli dimenziójának jellegzetességei a dajkanyelvben. In J. Bóna, & V. Krepsz (Eds.), *Nyelvfejlődés csecsemőkortól kamaszkorig* (pp. 195–208). Budapest: ELTE Eötvös Kiadó.
- Kohári, A., Deme, A., Reichel, U. D., Szalontai, Á., & Mády, K. (2022). Tartalmas és funkciószavak időtartama csecsemők"oz szóló beszédben. In *XXVIII. Magyar Alkalmazott Nyelvészeti Kongresszus*.
- Kohári, A., Reichel, U. D., Markó, A., & Mády, K. (2020b). Breathiness in speech directed to 4-month-old infants. In M. Tiede, D. Whalen, & V. Gracco (Eds.), *12th International Seminar on Speech Production* (pp. 1–2). New Haven: Haskins Laboratories – Yale University.
- Krepsz, V., Huszár, A., & Mády, K. (2021). The relationship between boundary markers and audible inhalation in hungarian read speech. In O. Niebuhr,

- M. Zellers, K. Fisher, S. Asadi, A. and. Berger, M. K. Langedijk, R. and Nielsen, M. Rossi, & I. Valls-Ratés (Eds.), *Proc. 1st International Conference on Tone and Intonation (TAI)* (pp. 254–258). Sønderborg, Denmark: University of South Denmark.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26.
- Laing, C. E., Vihman, M., & Keren-Portnoy, T. (2017). How salient are onomatopoeia in the early input? a prosodic analysis of infant-directed speech. *Journal of Child Language*, 44, 1117–1139.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package 'emmeans'. estimated marginal means, aka least-squares means. <https://cran.r-project.org/web/packages/emmeans/index.html>.
- Ludusan, B., Cristia, A., Martin, A., Mazuka, R., & Dupoux, E. (2016). Learnability of prosodic boundaries: Is infant-directed speech easier? *The Journal of the Acoustical Society of America*, 140, 1239–1250.
- Mády, K., Gyuris, B., Gärtner, H.-M., Kohári, A., Szalontai, A., & Reichel, U. D. (2022). Perceived emotions in infant-directed narrative across time and speech acts. In S. Frota, M. Cruz, & M. Vigário (Eds.), *Proceedings 11th International Conference on Speech Prosody Lisbon* (pp. 590–594). Lisbon, Portugal: International Speech Communication Association (ISCA).
- Mády, K., Reichel, U., Szalontai, Á., Kohári, A., & Deme, A. (2018). Prosodic characteristics of infant-directed speech as a function of maternal parity. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th International Conference on Speech Prosody* (pp. 294–298). Poznań, Poland: International Speech Communication Association (ISCA).
- Mády, K., Reichel, U. D., Kohári, A., & Szalontai, Á. (2021). The role of accommodation in expressing emotions to newborn babies. In *1st International Conference on Tone and Intonation* (pp. 124–125).

- Mangiafico, S. (2022). rcompanion: Functions to support extension education program evaluation. <https://CRAN.R-project.org/package=rcompanion>.
- Markó, A. (2005). *A spontán beszéd néhány szuprasegmentális jellegzetessége. Monologikus és dialogikus szövegek összevetése, valamint a hűmmögés vizsgálata*. Ph.D. thesis ELTE Budapest.
- Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition*, 156, 52–59.
- Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, 251, 118991.
- Narayan, C. R., & McDermott, L. C. (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America*, 139, 1272–1281.
- Navarro, D. (2021). lsr: Companion to "learning statistics with r". <https://CRAN.R-project.org/package=lsr>.
- Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16, 55–68.
- van Ommen, S., Boll-Avetisyan, N., Larraza, S., Wellmann, C., Bijeljac-Babic, R., Höhle, B., & Nazzi, T. (2020). Language-specific prosodic acquisition: A comparison of phrase boundary perception by french- and german-learning infants. *Journal of Memory and Language*, 112, 104108.
- Payne, E. H., Hardin, J. W., Egede, L. E., Ramakrishnan, V., Selassie, A., & Gebregziabher, M. (2017). Approaches for dealing with various sources of overdispersion in modeling count data: Scale adjustment versus modeling. *Statistical methods in medical research*, 26, 1802–1823.

- Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M.-C., & Cohen, D. (2013). Motherese in interaction: at the cross-road of emotion and cognition? (a systematic review). *PloS one*, 8, e78103.
- Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 43, 549–565.
- Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35, 869–902.
- Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, 10, 1–15.
- Szalontai, Á. (2018). Prosodic boundaries in hungarian infant-directed speech. In *Proceedings Experimental and Theoretical Advances in Prosody (ETAP) 4*. URL: http://real.mtak.hu/89350/1/ETAP4_paper_53.pdf.
- Szaszák, G. (2008). *A szuprasegmentális jellemzők szerepe és felhasználása a gépi beszédfelismerésben*. Ph.D. thesis Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Kar Távközlési és Médiainformatikai Tanszék. Budapest. URL: <https://repositorium.omikk.bme.hu/bitstream/handle/10890/809/ertekezes.pdf?sequence=1&isAllowed=y>.
- Team, R. C. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.
- Trouvain, J., & Werner, R. (2022). A phonetic view on annotating speech pauses and pause-internal phonetic particles. In *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion: Konzepte, Probleme, Lösungen*, 55 (pp. 55–73). Narr Francke Attempto Verlag.

Winter, B. (2019). *Statistics for linguists: An introduction using R*. New York: Routledge.

A logopédiai terápia hatása az óvodások fonológiai tudatosságára

Zemán Andrea Anna¹

¹*ELTE BTK Nyelvtudományi Doktori Iskola*

Abstract

Phonological awareness is a predictive factor in successful reading acquisition, influenced by several factors. Many children in kindergarten receive speech and language therapy due to articulation or language disorders. The aim of the present study was to analyze how speech therapy affects phonological awareness and its three levels (rhyme, syllable, and phoneme awareness) of five- and six-year-old kindergarteners with or without speech sound disorders. Results show better achievement in all levels of phonological awareness in cases of those who received speech therapy development. Examining the interaction between speech sound disorders and speech therapy, we found that all children with speech articulation disorders aged five- and six- years, who received speech therapy performed better in all levels of phonological awareness than those, who have articulation disorders, but who had not received speech therapy. Among the groups of children with clear speech, children who received speech therapy development performed better in all levels of phonological awareness than those who did not receive speech therapy development. Overall, the best performance was achieved by those children, who did not have articulation disorder but nevertheless received speech therapy development. The weakest performance was achieved by those children who have speech sound disorders but did not receive speech therapy. The study's results suggest that it would be important for all children to receive speech therapy development, not just children with speech sound disorders, especially in kindergarten, the year before elementary school. This development can lay the foundation for successful written language acquisition.

Keywords: phonological awareness, rhyme, syllable, phoneme awareness, speech sound disorders, speech therapy development

1. Bevezetés

A mai modern társadalmak fejlődése az információs társadalommá válás felé tolódik. Ebben a társadalomban a boldogulás „záloga” a fejlett olvasási készség mellett a megfelelő szintű írás és számolás elsajátítása, mely az oktatás révén

Email address: zemanaa@staff.elte.hu (Zemán Andrea Anna)

valósul meg. Az emberek döntő része tanulás nélkül képes elsajátítani anyanyelvét. Ezzel szemben az olvasás – eltekintve néhány ritka kivételtől – tanulási folyamat révén jön létre (Cs. Czachesz, 1998).

Az olvasási készség kialakulásához mai tudásunk szerint az agy 18–20 területének összehangolt működése szükséges. Az olvasásnak legszorosabb kapcsolata a nyelvi rendszerrel van (Csépe, 2005). Az olvasást megalapozó képességek és készségek feltárását eddig már számos kutatás tűzte ki célul. A sikeres olvasás-elsajátítás egyik prediktív faktora a 4–8 év közötti időszakban a fonológiai tudatosság (Jordanidisz, 2009; Kiss et al., 2019; Juhász & Kálló, 2017), melynek nem megfelelő működése akadályát képezheti a sikeres olvasástanulásnak (Adamikné Jászó, 2006; Tóth & Csépe, 2009; Steklács, 2013).

A fonológiai tudatosság a metanyelvi tudatosság része, mely által képesek vagyunk a szótagokkal, beszédhangokkal tudatos és szándékos műveleteket végezni. A metanyelvi tudatosság a nyelvről, illetve annak funkciójáról való gondolkodást jelenti. Göncz (2003) úgy fogalmaz, hogy a metanyelvi tudatosság teszi lehetővé a nyelvnek a közlések megtételén és megértésén túl magáról a nyelvi rendszerről való gondolkodást is. Adamikné Jászó (2006) szerint „a nyelvi tudatosság olyan intuitív analízáló képesség, melynek segítségével megfigyeljük beszédünket, s alkotórészeire bontjuk azt. A valódi fonológiai tudatosság (vö. korai fonológiai tudatosság) tulajdonképpen a metanyelv kialakulását jelenti” (Adamikné Jászó, 2006, 153).

Fonológiai tudatosság alatt a szavak belső szerkezetéhez történő tudatos hozzáférést, a szavaknak az eltérő egységekre (szótagokra, hangokra) való bontásának képességét értjük (Csépe, 2006), mely segítségével képesek vagyunk a hangzó beszédet kisebb egységekre bontani és ezeken az egységeken különböző műveleteket végrehajtani. A fonológiai tudatosságon belül két fő szintet tudunk megkülönböztetni a fejlődésük alapján (Barbour et al., 2003; Goswami, 2003; idézi Jordanidisz, 2009, 2015): az egyik a fonológiai szint (ide tartozik a rím és a szótagtudatosság), a másik a fonémaszint (ami a fonématudatosságot foglalja magába, vagyis a fonémákkal/beszédhangokkal végzett műveleteket). A

szintek fejlődése időben eltérő: a fonéमतudatosság fejlődését a fonológiai szint fejlődése megelőzi (Jordanidisz, 2009).

A fonológiai tudatosság fejlődése a 4–5 éves időszakban robbanásszerű fejlődésnek indul (Juhász & Kálló, 2017). Ebben az életkorban a rímtudatosság tekinthető a későbbi hatékony olvasás- és írástanulás korai indikátorának több vizsgált nyelv esetén (Bradley & Bryant, 1983; De Jong & Van Der Leij, 2003). A szótagtudatosság fejlődését főleg a „mondókázás,” a mondókák tanítása/tanulása segíti, azonban a rímtudatosság fejlődését is elősegíti. A szótagszintű műveltség gyengeségét mutatták ki olvasási zavarral küzdő gyermekek esetében, aminek hátterében a fejletlen beszédritmus-észlelés állt (Goswami, 2002). A fonémaszint érése az olvasástanulás során és azzal együtt fejlődik (Goswami, 2003; Csépe, 2006). Iskoláskor kezdetén a gyerekek megtanulják, hogy minden kimondott szó hangok sorozatából áll, és ezeket a hangokat le is tudják írni, mivel minden egyes hang egy betűt jelöl (hang $\rightarrow$ betű megfeleltetés). Olvasás során pedig ennek a folyamatnak az inverze zajlik, vagyis a betű $\rightarrow$ hang megfeleltetés. Így érthető, hogy a fonémaszint fejlettsége jelentős mértékben meghatározza az olvasástanulás sikerét.

Jordanidisz (2010) 30 fő budapesti, magyar anyanyelvű, 4–6 éves óvodás gyermek fonológiai tudatosságának vizsgálati eredménye azt mutatta, hogy a szintek fejlődési sorrendjében először a szótagtudatosság, utolsónak pedig a fonémaszint fejlődött.

Jordanidisz doktori értekezésében (2015) a magyar anyanyelvű, 4–10 éves gyermekek fonológiai tudatosságának fejlődését mérte fel. Az alkalmazott vizsgálati eljárás a Fonológiai Tudatosság Teszt, a Phonological Awareness Skills Survey magyar adaptációja (Jordanidisz, 2017) volt. A teszt 15 képességterület fejlettségét méri fel. A vizsgálatba 450 fő, ép intellektusú óvodás és iskolás lett bevonva. Az 1. táblázatban az 5 és a 6 éves gyermekek teljesítményeinek mediánjai láthatók százalékos értékben (Jordanidisz doktori értekezésében a 4–10 éves korosztályban vizsgálta a fonológiai tudatosságot, melyet százalékos értékben mutatott be (2015, 77-78). Tekintettel arra, hogy jelen kutatásban az 5

és a 6 évesek fonológiai tudatosságát vizsgálom, ezért az 1. táblázat az eredeti táblázat felhasználásával, saját szerkesztéssel készült).

1. táblázat. A fonológiai tudatosság teszt feladataiban elért eredmények mediánjai az 5 és 6 éves korcsoportban (Jordanidisz, 2015, 77-78. alapján)

	A vizsgált terület	5 éves (%)	6 éves (%)
1	rímfelismerés	55	85
2	rímkérés 1	60	80
3	rímkérés 2	0	40
4	szótagszintézis 1	80	100
5	szótagszintézis 2	80	80
6	szótag-szegmentálás	100	100
7	szótagolás	70	80
8	szótagtörlés	30	80
9	fonématörlés	0	60
10	fonémaizolálás	10	90
11	fonémaszintézis 1	0	40
12	fonémaszintézis 2	0	20
13	fonémaszegmentálás	0	35
14	fonémaazonosítás	0	50
15	fonémamanipuláció	0	30

1.1. Beszédhangzavar hatása a fonológiai tudatosságra

Az elmúlt évtizedekben egyre több kutatás igyekezett feltárni a beszédhangzavar és a fonológiai tudatosság közötti kapcsolatot. Számos vizsgálat eredménye azt mutatta, hogy az óvodáskori beszédhangzavarok fokozott kockázatot jelenthetnek a fonológiai tudatosság nem megfelelő szintű fejlődésére és működésére, ezzel az olvasás és helyesírás sikeres elsajátítására (Bird et al., 1995; Blachman, 1997; Webster et al., 1997; Nathan et al., 2004; Raitano et al., 2004; Anthony et al., 2011; Hayiou-Thomas et al., 2017).

Preston et al. (2013) 25 fő beszédhangzavart mutató gyermek körében folytattak utánkövetéses vizsgálatot. Kutatásuk eredménye azt mutatta, hogy az

óvodáskori (4–5 év) beszédhanghibák kockázatot jelentenek, illetve előre jelezhetik a fonológiai tudatosság hosszútávú gyengeségét.

Cowan & Moran (1997) felmérésükben összehasonlították az artikulációs zavart mutató és az azt nem mutató óvodások, valamint kisiskolások (első és második osztályosok) fonológiai tudatosságát. A kutatásba 28 főt vontak be. A fonológiai tudatosságot 3 feladattal mérték fel: rímkeresés, fonéma-manipuláció és fonémaszámlálás. Az eredmények szignifikáns különbséget mutattak a két csoport teljesítménye között: az artikulációs zavarral küzdő gyermekek lényegesen több hibát követtek el a fonológiai tudatosság mindhárom feladatában, szemben az ép artikulációjú gyermekekkel. A kapott eredmények összhangban állnak korábbi, a témában folyt kutatások eredményeivel (Cowan & Moran, 1997). Ennek ellenére a szerzők leszögezik, hogy nem minden artikulációs problémával küzdő gyermek teljesít gyengébben a fonológiai tudatosságot mérő feladatokban.

Thomas & Sénéchal (1998) kutatásukba 80 fő, normál fejlődésű, 3–4 év közötti gyermekeket vontak be, melyben meghatároztak specifikus artikulációs mintákat, és tanulmányozták a gyermekek fonológiai tudatosságát ezekre a specifikus artikulációs mintákra nézve. Ezek a minták a(z angol nyelvterületen) gyakran hibásan ejtett szókezdő beszédhangok ([r], [l], [m]) voltak. Különösen a [r], de emellett a [l] hangokat rendszerint a [w]-vel cserélték. A vizsgálat kétirányú okozati összefüggést igazolt a fonéमतudatosság és az artikuláció között. Azok a gyerekek, akik képtelenek voltak pontosan képezni a [r] hangot, gyengébben teljesítettek a fonéमतudatosságnak a fonéma-manipulációt igénylő feladataiban. A szerzők szerint az artikuláció és a fonéमतudatosság közötti kapcsolat lehet közvetett is, melyet egy másik változó (pl. nonverbális intelligencia, szókincs nagysága, rövidtávú – fonológiai – memória) közvetít.

Nem minden vizsgálat támasztja alá azonban az artikulációs zavar és a fonológiai tudatosság közötti kapcsolat meglétét. Foy & Mann (2001) a 4–6 éves gyermekek körében végzett kutatásának eredménye alapján nem igazolódott szignifikáns összefüggés az artikulációs hibák és a fonéमतudatosságban nyújtott teljesítmény között.

1.2. A logopédiai ellátás

Az Európai Unió logopédiai szakmai szervezete (CPLOL) „meghatározása szerint a logopédus az a szakember, aki a kommunikációs zavarok megelőzésével, vizsgálatával, diagnosztikájával, terápiájával, kezelésével és ehhez kapcsolódó tudományos vizsgálatokkal foglalkozik” (Gereben et al., 2012, 28).

Magyarországon a logopédusok döntő többsége pedagógiai szakszolgálatok által foglalkoztatott szakemberek. A pedagógiai szakszolgálati intézmények működéséről szóló 15/2013 (II.26.) EMMI rendelet 25. § (1) alapján a logopédiai ellátás feladata a hangképzési, beszédbeli, valamint a beszélt és írott nyelvi készségek és képességek zavarainak, valamint a specifikus tanulási (olvasás-, írás és helyesírás-, számolás) zavaroknak a megelőzése, állapotfeltárása és terápiája. A rendelet 25. § (3) pontja értelmében a logopédusok elvégzik a „harmadik és ötödik életévüket betöltött gyermekek beszéd- és nyelvi fejlettségének szűrését [...] az ötéveskori szűrés elsősorban a beszédartikulációra, illetve az írott nyelvi (írás és olvasás) készségekre irányul,” és amennyiben szükséges, további vizsgálatok végzésére tesznek javaslatot. Ezen eredmények birtokában, akiknek a beszédartikulációja, valamint az olvasás- és íráskészség elmaradása indokoltta teszi, logopédiai ellátásba, terápiába vonják.

Hazánkban a logopédus szakemberek olyan komplex ellátást nyújtanak, melynek keretében az artikulációs, azaz a beszédhang-képzési- és kiejtési hibák javításán túl többek között a diszlexia prevenció, a részképességek- és nyelvi készségek fejlesztése is a foglalkozások részét képezik. Vagyis logopédiai fejlesztésben nem csak a beszédartikulációs zavart mutató gyermekek vesznek/vehetnek részt, hanem olyanok is, akiknek kiejtése tiszta.

2. A kutatás

2.1. A kutatás célja, hipotézisei

Áttekintve néhány – a témában született – szakirodalmat, azt látjuk, hogy a sikeres írás- és olvasástanulás mint írott nyelv egyik nélkülözhetetlen alap-

készsége a fonológiai tudatosság, melynek fejlődését számos tényező, például a beszédhangzavarok fennállása befolyásolhat.

Az artikulációs zavarral kapcsolatban a kutatások eredményei nem egységesek; vannak vizsgálatok, melyek alátámasztják, hogy a beszédartikulációs hibák megléte összefüggést mutat a fonológiai tudatosság (továbbiakban FT) tesztekben elért teljesítménnyel, míg mások nem találtak egyértelmű kapcsolatot a két változó között. Magyarországon eddig még egy kutatás sem vizsgálta az összefüggést (van-e, és ha igen, milyen irányú a kapcsolat) a FT-ra vonatkozóan az artikulációs zavar, vagyis a beszédhiba és a logopédiai terápiában való részesülés alapján az 5–6 éves óvodás gyermekek körében. A bemutatásra kerülő eredmények egy nagyobb ívű kutatás részei, melyek segítenek a vizsgálat további irányának meghatározásában.

Kutatásomban arra kerestem választ, hogy egyfelől a beszédhibás és a tiszta artikulációjú gyermekek FT-beli teljesítményében van-e különbség, másfelől a logopédiai ellátás befolyásolja-e a FT-teszt eredményeit.

- Felvetésem az volt, hogy a mintába került, logopédiai terápiában részesülő gyermekek jobb teljesítményt nyújtanak a FT minden vizsgált szintjén azokkal szemben, akik nem kapnak logopédiai fejlesztést.
- Azt feltételeztem továbbá, hogy a teljes mintában a tiszta artikulációjú gyermekek jobb eredményt érnek el a FT tesztben és a FT három szintjén (rím-, szótag-, fonéमतudatosság) a beszédhibás gyermekekkel szemben.
- A FT teszt mindhárom vizsgált területén a legjobb teljesítményt azok a tiszta artikulációjú gyermekek nyújtják, akik annak ellenére, hogy nem beszédhibások, logopédiai fejlesztésben vesznek részt, és leggyengébben az artikulációs zavarral küzdő, de a logopédiai terápiából kimaradó gyermekek teljesítenek.

2.2. Vizsgálati személyek, anyag, módszer

A keresztmetszeti, nem reprezentatív vizsgálatba Borsod-Abaúj-Zemplén megye több településéről összesen 103 fő 5 és 6 éves kor közötti gyermek került

bevonásra (átlagéletkor: 68,5 hónap). A családok rendszeresen használt nyelve magyar, az anamnesztikus adatok alapján a felmérésben részt vevő valamennyi gyermek tipikus fejlődésű. A vizsgálatból a sajátos nevelési igényű gyermekek kizárásra kerültek. A 2. táblázatban a kutatásban részt vevők megoszlása látható nemenként, továbbá a beszédhiba megléte, illetve a logopédiai ellátásban való részesülés szerint, emellett a teljes mintában való arányuk.

2. táblázat. A minta összetétele

	fő	teljes mintára (103 fő) vetítve
fiú	62	60,19%
lány	41	39,81%
beszédhiba	45	43,68%
kap(ott) logopédiai ellátást	48	46,60%

A kutatásban való részvételhez történő hozzájárulással egyidőben a szülők válaszoltak arra a kérdésre, hogy gyermekük kap, illetve korábban kapott-e logopédiai ellátást. Így a logopédiai terápia szerint kialakított csoportba bekerültek olyan gyermekek, akik beszédhibájuk miatt a vizsgálatkor is logopédiai terápiában részesültek, illetve akiknek korábban volt beszédhang-éjtési zavaruk, amire logopédiai ellátást kaptak, ám a vizsgálatkor már beszédük tiszta volt, valamint olyan gyermekek is, akik beszéde mindig tiszta/ép volt, azonban egyéb ok miatt részesültek logopédiai fejlesztésben. Arra vonatkozóan nincsenek adatok, hogy a gyermekek a logopédiai ellátást hol kapták meg: intézményrendszeren belül (pl. óvodában, pedagógiai szakszolgálatoknál stb.) vagy magánúton.

A felmérésben több fiú vett részt. A mintába került gyermekek közel 44%-a beszédhang-kiejtési zavarral küzdő (45 fő). A 103 gyermek nagyjából fele (48 fő) részesül, vagy korábban vett részt logopédiai intervencióban, 55 fő nem kapott logopédiai fejlesztést. (Az adatok alapján úgy tűnhet, hogy többen részesültek logopédiai ellátásban, mint ahányan beszédhibások. Megvizsgálva azonban életkoronként a beszédhiba és a logopédiai ellátásban részesülést, az látható, hogy a mintában szereplő 28 fő beszédartikulációs zavarral küzdő 5 éves közül csupán

24-en részesültek logopédiai terápiában; vagyis nem minden beszédhang-ejtési zavart mutató gyermek vett részt logopédiai terápiában. A 6 évesek esetében 17 fő volt beszédhibás, viszont 24 gyermek kapott logopédiai ellátást, ami azt jelenti, hogy a 6 évesek közül olyanok is kaptak logopédiai fejlesztést, akiknek beszéde tiszta. Nem volt lehetőség azonban a FT tesztben elért eredményeket életkoronként vizsgálni a beszédhiba és logopédiai ellátásban való részesülés szerinti csoportbontásban (5/6 év, beszédhiba van/nincs, logopédiai ellátást kap/nem kap), mivel az egyes csoportok létszáma jelentősen eltér egymástól (2–24 fő) és a kapott eredmények alapján téves következtetéseket vonhatnánk le).

A fonológiai tudatosságban elért eredmények összehasonlítására a következő csoportok lettek kialakítva a beszédhiba és a logopédiai ellátásban részesülés alapján (3. táblázat).

3. táblázat. A csoportok létszáma a beszédhiba és a logopédiai ellátásban való részesülés alapján

beszédhiba	logopédiai terápia	összesen (fő)
van	igen	34
	nem	11
nincs	igen	14
	nem	44

A gyermekek fonológiai tudatosságának vizsgálatára alkalmazott eljárás a NILD Tanulási Terápia program Fonológiai Tudatosság tesztje (lásd bővebben: Jordanidisz, 2009, 2015) volt. A teszt 15 feladatot tartalmaz, melyből 3 a rímtudatosságot tárja fel (feladattípusok pl. rímfelismerés, szó- és mondóka alapú rímkeresés). 5 feladat a szótagtudatosságot (feladattípusok pl. valódi- és ál-szavak szótagszintézise, szegmentálás, szótagolás, szótagtörlés szavakból) és 7 feladat a fonéमतudatosságot méri (feladattípusok pl. beszédhang azonosítása, hangokból történő szóalkotás, hangszegmentálás, hosszú hang azonosítása, hangmanipuláció: szavakon belüli hangcsere, szavak megfordítása/visszafelé mondása) (Jordanidisz, 2015). Az egyes feladatokban eltérő pontszámok ad-

hatók (maximum 5, illetve 10 pont). A tesztben elért egyéni eredmények feldolgozása során az összehasonlíthatóság érdekében a pontok átváltásra kerültek teljesítményszázalékra a következők alapján:

- elérhető maximum: 10 pont $\rightarrow$ 100%. Elért: 3 pont $\rightarrow$ 30%, elért: 2 pont $\rightarrow$ 20% stb.
- elérhető maximum: 5 pont $\rightarrow$ 100%. Elért: 3 pont $\rightarrow$ 60%, elért: 2 pont $\rightarrow$ 40% stb.

A feladatok fokozatosan nehezedőek. A teljes teszt felvétele a gyermekek teljesítményétől függően 10–25 percet vett igénybe.

A beszédhiba mint csoportosító tényező szerint kialakított két csoportba a FT teszt felvételekor elvégzett artikuláció/beszédhangejtés vizsgálata alapján kerültek a gyermekek (nincs/van $\rightarrow$ *nincs*: ha a beszéd tiszta, illetve *van*: a hibás artikulációval érintett hang(ok)/hangcsoport(ok) számától, vagyis a súlyosságtól függetlenül észlelhető beszédhiba). A beszédhangok ejtésének felmérésére a *Logopédiai vizsgálatok kézikönyvének* (Juhász, 2007) Az artikuláció vizsgálata (10. és 11. melléklete), valamint a *Szól-e?* szűrőeljárás az óvodáskori logopédiai ellátáshoz (Kas et al., 2012) teszt Beszédhangejtés vizsgálat szubtesztje lett használva. Mivel Borsod-Abaúj-Zemplén megye nem minden településén van lehetőség a *Szól-e?* szűrőeljárást alkalmazni, így, ahol a logopédusok szeptember–október hónapok során az ötéveskori szűrés során a *Szól-e?* tesztet vették fel, úgy a jelen kutatásban az artikuláció vizsgálatára a *Logopédiai vizsgálatok kézikönyvének* említett melléklete került alkalmazásra, és fordítva. Így elkerülhető volt, hogy a beszédhangejtési zavarok szűrése ugyanazzal a teszttel történjen. A beszédhibák jellegüket és a mintában való arányukat tekintve a következők voltak: hangcsere (paralália: 85,15%), hangtorzítás (diszlália: 11,56%), hangkihagyás (alália: 2,81%), hangátvetés (0,48%). A mintába bevont gyermekek egyikénél sem fordult elő beszédfolyamatosság zavara. Mivel relatíve kevés adatközlőtől vannak eredmények, ezért nem volt lehetőség a beszédhiba típusa, vagyis az artikulációs zavarral érintett hangok/hangcsoportok, vagy az artikulációs zavar súlyossága szerinti összevetésére és elemzésére.

A független változó a logopédiai ellátásban való részvétel volt, mely alapján két csoport került kialakításra (logopédiai terápiában részesül(t): *igen/nem*). Az alkalmazott statisztikai modell lineáris modell volt. Az eredmények feldolgozása és az adatelemzés az *SPSS 20* szoftverben történt. A független változó hatásának vizsgálata egyszempontos varianciaanalízis (ANOVA) módszerrel történt.

3. Eredmények

A kutatás eredményeinek közlése során először a teljes, 15 feladtból álló FT tesztben, majd az egyes szinteken (rím-, szótag- és fonéमतudatosság) nyújtott eredmények kerülnek bemutatásra (százalékos értékben), aszerint, hogy miként befolyásolja a mintába került gyermekek teljesítményét a logopédiai terapia, valamint hogyan változnak a FT tesztbeli teljesítmények a beszédhiba megléte és a logopédiai ellátásban részesülés alapján (a felmérésben a beszédhiba mint csoportosító tényező szerepel, nem független változó). Annak ellenére, hogy a jelen kutatásban vizsgált életkorban a fonéमतudatosság padlóhatást mutat, mégis fontos megvizsgálni ezt a szintet is, hogy lássuk, a beszédhangszintű műveletekre milyen hatással bír a logopédiai terapia, valamint a beszédhangejtési zavar. A leíró statisztika alapján a logopédiai terápiában való részvétel szerinti eredményeket a 4. táblázat mutatja be. Az 5. táblázatban a beszédhiba megléte szerint kialakított csoportok átlagos FT teljesítményei láthatók a logopédiai ellátásban való részvétel alapján.

4. táblázat. A logopédiai terapia hatása a fonológiai tudatosságra

logopédiai ellátás	összesen (fő)	teljes FT teszt		tudatosság szintje					
				rím-		szótag-		fonéma-	
		átlag	szórás	átlag	szórás	átlag	szórás	átlag	szórás
igen	48	42,31	15,00	50,35	21,25	70,50	16,06	18,72	17,09
nem	55	35,68	14,68	42,13	18,83	62,73	19,86	13,60	15,79

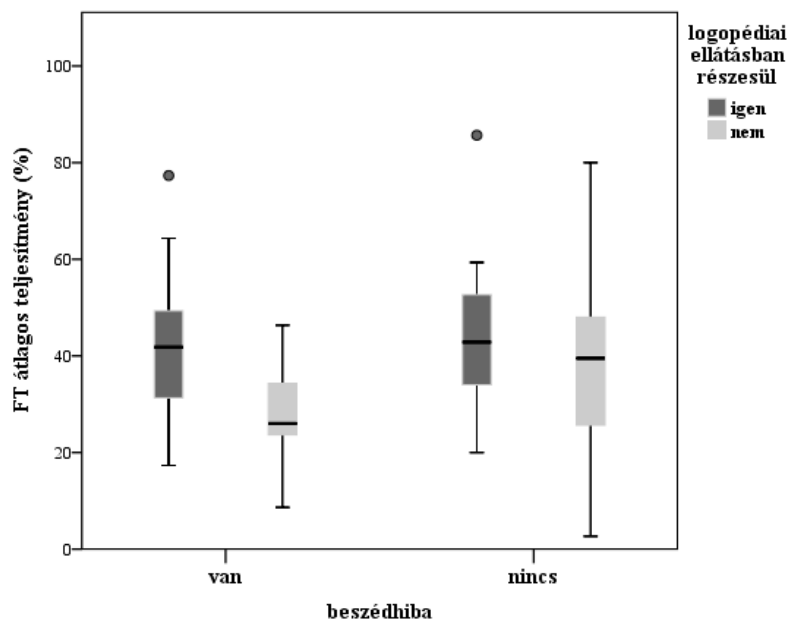
5. táblázat. A logopédiai terápia és a beszédhiba hatása a fonológiai tudatosságra

beszédhiba	logopédiai ellátás	összesen (fő)	teljes FT teszt		tudatosság szintje					
					rím-		szótag-		fonéma-	
			átlag	szórás	átlag	szórás	átlag	szórás	átlag	szórás
van	igen	34	41,46	14,32	48,33	21,26	69,65	16,32	18,38	15,58
	nem	11	27,85	10,32	38,18	17,34	51,64	14,94	6,43	9,38
nincs	igen	14	44,36	16,92	55,24	21,19	72,57	15,80	19,54	20,94
	nem	44	37,64	15,04	43,11	19,24	65,50	20,10	15,39	16,62

3.1. Teljes FT teszt

A logopédiai ellátásban való részesülés főhatásnak, vagyis a 103 fős mintára nézve a teljes FT tesztbeli eredményeket szignifikánsan befolyásoló tényezőnek mutatkozott ($F[1, 99] = 8,964$ $p = 0,03$). Azok a gyermekek nyújtottak jobb eredményt, akik logopédiai ellátásban vesznek részt.

A teljes FT tesztben elért teljesítményeket a logopédiai ellátásban részesülés és a beszédhiba alapján kialakított csoportokon vizsgálva azt látjuk, hogy legjobban azok a gyermekek teljesítettek, akiknek bár tiszta a beszédük, ennek ellenére mégis részt vesznek logopédiai fejlesztésben, leggyengébben pedig a logopédiai ellátásból kimaradó beszédhangejtési zavarral küzdő óvodások teljesítettek (1. ábra és 5. táblázat). Összességében a beszédhibás gyermekek átlagos teljesítménye alacsonyabb, mint a tiszta artikulációjú gyermekeké, azonban a csoportok teljesítménye közötti különbségek nem voltak szignifikánsak.

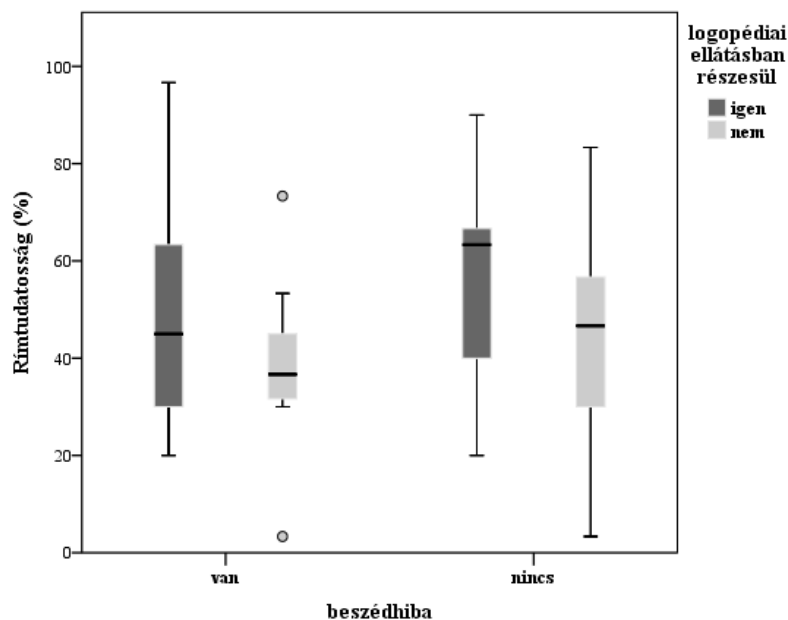


1. ábra. A logopédiai terápia hatása a fonológiai tudatosságra a beszédhiba megléte alapján

3.2. Rímtudatosság

A logopédiai terápiának, mint független változónak szignifikáns hatása igazolódott a rímtudatosság területén a teljes, 103 fős mintára nézve ($F[1,9] = 5,772$, $p = 0,018$), vagyis azok az 5 és 6 éves óvodások, akik logopédiai ellátásban részesülnek, magasabb átlagos eredményt értek el.

A beszédartikulációs zavart mutató 5 és 6 éves óvodások átlagos teljesítménye a rímtudatosság területén alacsonyabb, mint az ép beszédű társaiké. Ha a beszéd tisztasága szerint kialakított csoportokra vetítjük a logopédiai terápia hatását, azt látjuk, hogy a rímtudatosság feladatokban elért átlagos eredményekben eltérés mutatkozik a beszédhibás és a tiszta beszédhangejtésű gyermekek teljesítményei között, azonban a különbség nem szignifikáns. Legjobban a logopédiai fejlesztésben részesülő, tiszta beszédhangejtésű gyermekek teljesítettek, leggyengébben pedig a beszédhibás és logopédiai ellátásból kimaradóak (2. ábra és 5. táblázat).



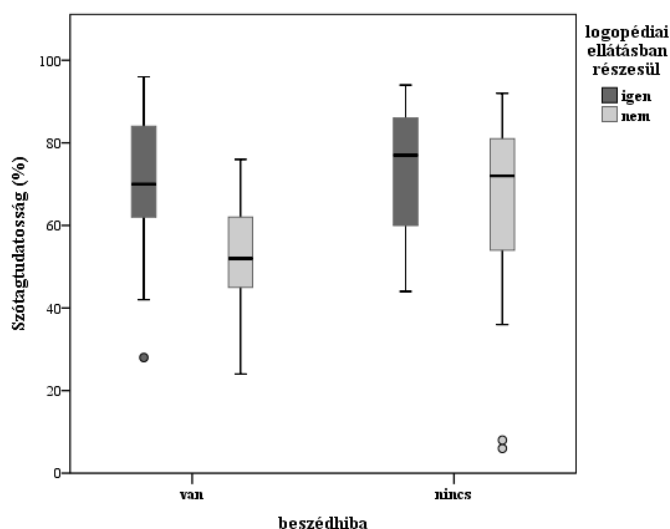
2. ábra. A logopédiai terápia hatása a rímtudatosságra a beszédhiba megléte alapján

3.3. Szótagtudatosság

A teljes mintára nézve a szótagtudatosság tesztekben elért eredményekre a logopédiai ellátásban való részvétel főhatásnak bizonyult a szótagtudatosság tekintetében ($F[1, 99] = 9,181, p = 0,003$), a kapott különbség szignifikáns. A logopédiai ellátásban való részvétel pozitív hatást gyakorolt a mintába került gyermekek szótagtudatosságára; magasabb átlagos eredményt értek el azokhoz képest, akik nem kapnak logopédiai fejlesztést. A beszédhibás gyermekek csoportjában a szótagtudatosságban elért átlagos teljesítmények kissé elmaradtak a tiszta beszédhangejtésű gyermekekhez képest, a különbség nem volt szignifikáns.

Ha a beszédartikulációs zavar megléte szerinti csoportbontásban nézzük a logopédiai fejlesztés hatását az elért eredményekre azt látjuk, hogy a beszédhangejtési zavart mutató gyermekek átlagos teljesítménye alacsonyabb a tiszta beszédű gyermekekhez képest. Hasonlóan az eddig bemutatott területekhez, a szótagtudatosság területén is a legjobb teljesítményt a tiszta beszédű és logopédiai fejlesztésben részesülő gyermekek érték el, a legrosszabb eredményt a

beszédhibás és logopédiai terápiát nem kapó óvodások nyújtották (3. ábra és 5. táblázat), a különbség tendenciajellegű.



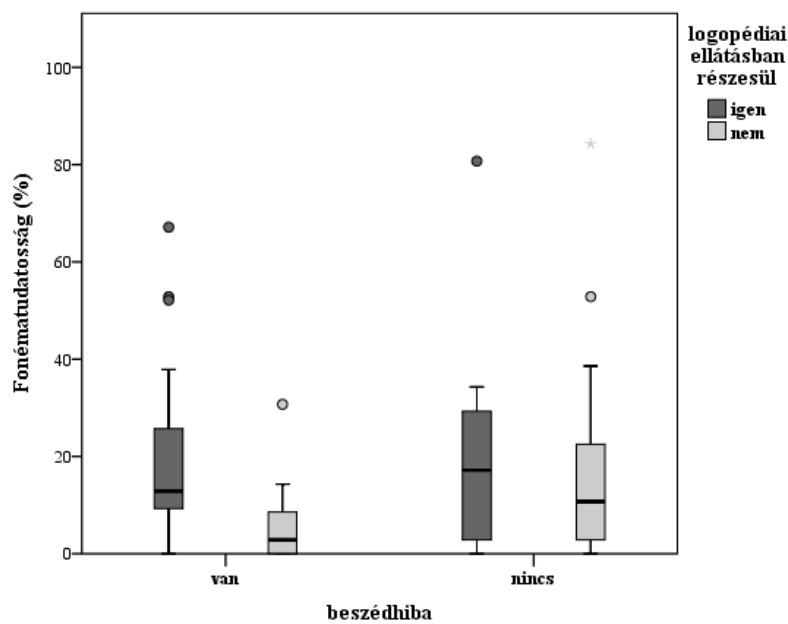
3. ábra. A logopédiai terápia hatása a szótagtudatosságra a beszédhiba megléte alapján

3.4. Fonéमतudatosság

Mint arról korábban már volt szó, a kutatásban vizsgált (5 és 6 év) életkorban a fonéमतudatosság padlóhatást mutat. A logopédiai ellátásban való részvétel szignifikáns különbségeket idézett elő a fonéमतudatosságot mérő feladatokban ($F[1, 99] = 4,522, p = 0,036$). Azok a mintában szereplő 5 és 6 éves gyermekek, akik részesülnek logopédiai fejlesztésben, magasabb átlagos eredményt nyújtottak azokhoz képest, akik nem vesznek részt logopédiai foglalkozásokon. A teljes mintában a beszédhibás gyermekek átlagos teljesítménye a fonéमतudatosságot mérő feladatokban elmaradást mutat ép beszédű társaik eredményeihez képest.

Ahogy az a 4. ábrán és 5. táblázatban látható, a beszédhiba megléte alapján kialakított csoportok teljesítményei között különbség van a logopédiai ellátásban részesülés szerint: a legjobb eredményt a nem beszédhibás és logopédiai foglalkozásokon részt vevő gyermekek nyújtották, míg a leggyengébb teljesítményt

a beszédartikulációs zavart mutató és logopédiai terápiát nem kapó óvodások érték el. A kapott különbségek nem voltak szignifikánsak.



4. ábra. A logopédiai terápia hatása a fonématusdátosságra a beszédhiba megléte alapján

4. Következtetések

A fonológiai tudatosság a sikeres olvasáselsajátítás egyik prediktív faktornak tekinthető, melyet számos tényező befolyásolhat. A jelen, nem reprezentatív kutatás célja az volt, hogy a logopédiai terápia hatását vizsgálja az óvodások fonológiai tudatosság egészére vonatkozóan, továbbá annak felmérése, hogy a beszédhangejtési zavar megléte befolyásolja-e a fonológiai tudatosság területein elért eredményeket.

A kutatás eredményei alapján a teljes mintára (103 fő) nézve a logopédiai terápiában részesülés szignifikáns különbséget mutatott valamennyi vizsgált területen. Azok a gyermekek, akik logopédiai foglalkozásokon vesznek részt, jobb eredményt értek el mind a teljes FT tesztben, mind a rím-, szótag- és fonématusdátosság területeken. Vagyis a kapott eredmények arra engednek következtetni,

hogy a logopédiai ellátásban való részesülés pozitív hatást gyakorolt a mintába került gyermekek fonológiai tudatosságára. A kutatás eredményei alapján az a feltevés, mely szerint a logopédiai terápiában részesülő gyermekek jobb teljesítményt nyújtanak a FT minden vizsgált szintjén azokkal szemben, akik nem kapnak logopédiai fejlesztést, igazolódott.

A beszédhangejtési zavarral küzdő gyermekek teljesítményei elmaradtak ép beszédű társaikhoz képest a teljes mintára nézve. Alacsonyabb átlagos teljesítményt nyújtottak ugyan a beszédhibás gyermekek a teljes FT tesztben, valamint a FT mindhárom szintjén tisztán beszélő társaikhoz képest, azonban a kapott különbségek nem voltak szignifikánsak. Így az a feltevés, mely szerint a tiszta artikulációjú gyermekek jobb teljesítményt érnek el a FT tesztben, a FT három szintjén (rím-, szótag-, fonéमतudatosság) a beszédhibás gyermekekhez viszonyítva, nem nyert igazolást. A látott különbségek tendenciajellegűek.

Az artikuláció tisztasága szerint kialakított csoportokra vetítve a logopédiai terápia hatását azt látjuk, hogy minden vizsgált területen azok a gyermekek teljesítettek legjobban, akik nem beszédhibások, ennek ellenére mégis logopédiai foglalkozásokon vesznek részt. A leggyengébb eredményt a beszédartikulációs zavart mutató és a logopédiai terápiából kimaradó gyermekek nyújtották. Azok az óvodások, akiknek beszéde tiszta – és talán épp emiatt nem is vesznek részt logopédiai fejlesztésben – alacsonyabb eredményt értek el azokkal szemben, akiknek beszéde nem tartalmaz hangzóhibát és mégis logopédiai ellátásban részesülnek. Vagyis a kutatás eredményei azt mutatják, hogy a nem beszédhibás gyermekeknek „jót tett” a logopédiai foglalkozásokon való részvétel a fonológiai tudatosság szempontjából. Az átlagos eredményekben elért legnagyobb különbségeket a beszédhibás gyermekek csoportjában látjuk: akik logopédiai terápiában részesülnek, jóval magasabb teljesítményt nyújtottak valamennyi vizsgált területen azokhoz képest, akik beszédartikulációs zavarral küzdenek, azonban mégsem kapnak logopédiai terápiát.

A beszédhiba mint csoportosító tényező mentén vizsgálva a logopédiai terápia hatását a fonológiai tudatosság területeire nézve a kutatás azt az eredményt hozta, hogy a mintába került gyermekek közül legjobban a tiszta beszédű, ám

mégis logopédiai fejlesztésben részt vevő gyermekek teljesítettek. Azonban a hipotézisem, mely szerint legjobban a nem beszédhibás, de logopédiai terápiában részesülő gyermekek teljesítenek, továbbá leggyengébben azok, akik beszédartikulációs zavart mutatnak és mégsem kapnak logopédiai terápiát, nem nyert bizonyítást, mivel az egyes csoportok eredményei közötti különbségek nem voltak szignifikánsak.

A fonológiai tudatosság fejlődése szintenként eltérő, a fonémaszint fejlődését a fonológiai szint (rím- és szótagtudatosság) fejlődése megelőzi. A jelen kutatás eredményei alátámasztják Jordanidisz (2010) vizsgálatának eredményét, mely szerint a magyar anyanyelvű 4–6 éves óvodások fonológiai tudatossága szintjeinek fejlődési sorrendjében először a szótagtudatosság fejlődik, míg utoljára a fonématudatosság. A jelen kutatásba bevont 103 gyermek átlagos teljesítménye a szótagtudatosság területén volt a legmagasabb, míg a leggyengébb eredményt a fonématudatosságban látjuk. Ahogy arról korábban már volt szó, a kutatásban vizsgált (5 és 6 év) életkorban a fonématudatosság ugyan padlóhatást mutat, ennek ellenére fontos azt vizsgálni, hogy a beszédhangszintű műveletekre hogyan hat a logopédiai fejlesztés, valamint a beszédartikulációs zavar megléte. Így a jelen, feltáró jellegű vizsgálatban elért eredmények kiindulópontjai lehetnek a további, a fonológiai tudatosságot vizsgáló kutatásoknak.

Logopédiai terápiát nem minden artikulációs zavarral küzdő, 5 és 6 éves gyermek kap/kapott, aminek oka lehet a szakemberhiány is (kevés a logopédus, emiatt – különösen a nagyvárosoktól távolabbi településeken – gyermekek maradhatnak ki az ellátásból). A mintában szereplő 28 fő 5 éves, beszédartikulációs zavart mutató gyermek közül csak 24 kap/kapott logopédiai terápiát, vagyis az 5 évesek közül sajnos többen kimaradnak a logopédiai ellátásból. A 6 évesek között 17 gyermek volt beszédhibás, ám 24 fő részesül(t) logopédiai fejlesztésben, ami azt jelenti, hogy a 6 évesek közül olyanok is kapnak/kaptak logopédiai fejlesztést, akiknek beszéde tiszta. Ez utóbbi háttérben több ok is meghúzódhat: korábban volt a gyermeknek beszédhibája, amire logopédiai ellátást kapott, vagy akár az is, hogy ugyan sosem volt artikulációs probléma,

azonban egyéb okok vezérelték a szülőket a gyermekük logopédiai fejlesztésben való részvétele érdekében (lásd az 1.2 alfejezetet).

Nem volt lehetőség azonban a FT tesztben elért eredményeket és intervenciók/kereszthatásokat vizsgálni életkor, beszédhiba és a logopédiai ellátásban részesülés szerint, (5/6 év; beszédhiba van/nincs; logopédiai ellátást kap/nem kap), mivel az egyes csoportok létszáma jelentősen eltérne egymástól (2–24 fő), és a kapott eredmények alapján téves következtetéseket vonhatnánk le. Fontos lenne azonban látni, hogy hogyan is alakulnak a gyermekek fonológiai tudatosságbeli teljesítményei életkoronként annak tükrében, hogy van-e beszédartikulációs probléma vagy sem, különösen az iskolába lépést megelőző időszakban. A kapott eredmények összevethetők lennének egy jelenleg zajló felméréssel, mely azt igyekszik feltárni, hogy az 5 és 6 éves óvodások fonológiai tudatosságát a kognitív folyamatok (pl. emlékezés, figyelem, beszédpercepció részfolyamatai: beszédhang-differenciálás, szeriális észlelés stb.) hogyan befolyásolják a beszédhiba megléte szerint kialakított csoportokban.

A kutatás eredményeinek értelmezése azonban több szempontból is óvatosságra int. A mintába került gyermekek logopédiai terápiában való részvételi ideje nem volt egységes. A jelen kutatásban a tesztek felvételére október–február hónapok között került sor. Magyarországon a logopédusok szeptember és október hónapok során mérik fel a gyermekek nyelvi- és beszédartikulációs állapotát, ami alapján kezdik meg szeptembertől a szűrést követően az ellátást azoknál a gyermekeknél, akiknél szükséges. Emiatt a jelen vizsgálatba a logopédiai terápiában részesülő gyermekek csoportjába bekerült az is, aki alig 1–2 hét óta kapott fejlesztést, de olyan gyermekek is, akik akár már 4–5 hónapja. Továbbá előfordulhatott az is (többnyire a 6 évesek esetében), hogy a beszédhiba miatt még a vizsgálat időpontjában is logopédiai ellátásban részesült (kb. 1 éve) a gyermek, de akár olyan is adódhatott, hogy korábban ugyan artikulációs zavarral küzdött, ám a kapott logopédiai terápia hatására a teszt felvételekor már tünetmentes (tiszta hangképzésű) volt. Emiatt nem ítéltethető meg objektíven a logopédiai terápiában való részvétel szerinti teljesítménykülönbség a csoportok között. Mivel összesen „csak” 103 gyermek került bevonásra, ez nem adott lehe-

tőséget további alcsoportok kialakítására (például, mióta részesülnek logopédiai ellátásban, heti hány órában, vagy mennyi az otthoni gyakorlás ideje stb.).

Meg gondolandó – megfelelő számú adatközlő esetén – a beszédhiba alapján kialakított csoportokon belül további megkülönböztetéseket tenni aszerint, milyen hangokat, illetve hangcsoportokat érint a beszédhiba (pl. zöngés-zöngétlen hangképzési zavar, paraláliák pl. r-l-j, s-zs-cs, sz-z-c, ty-gy-ny, k-t, g-d stb.). Ez idáig nagyon kevés olyan kutatás folyt, mely a hibásan ejtett hangok szerint vizsgálta volna az óvodáskorúak fonológiai tudatosságát. Ezáltal választ kaphatnánk arra, hogy van-e összefüggés a fonológiai tudatosság tesztben elért eredmények és a cserélt hangok típusa között.

További korlátot jelenthet a beszédhibás gyermekek teljesítményének értékelése a beszédhiba alapján kialakított csoportban, ugyanis épp a beszédartikulációs zavar miatt részesülnek logopédiai ellátásban. Így amikor a beszédhiba alapján kialakított csoportok teljesítményeit hasonlítjuk össze, az eredményeket befolyásolhatja a logopédiai terápia hatása is. Mivel a mintába került artikulációs zavarral küzdő óvodások egy része kap/kapott logopédiai terápiát, így nem tudjuk kizárni teljesen a beszédhiba alapján kialakított csoportban a logopédiai ellátás hatását is.

Összességében elmondható, hogy a jelen kutatás során igazolódott a logopédiai ellátás hatása a fonológiai tudatosságra és annak területeire nézve. A fonológiai tudatosság jelentősége az iskolába lépés előtt különösen fontos, hiszen annak megfelelő szintje megalapozza a sikeres írott nyelv elsajátítását. Jelentős információ birtokába jutottunk továbbá a felmérés eredményei által a nem beszédhibás gyermekek logopédiai fejlesztésben való részesülésének hasznosságát illetően a fonológiai tudatosság egészére, valamint annak szintjeire nézve.

A fonológiai tudatosság a sikeres olvasás- és írás elsajátítás egyik legfontosabb prediktív faktora. Ezért fontos lenne, hogy minden, nem csak a beszédhibás gyermekek kapjanak logopédiai ellátást, különösen a beiskolázás előtt álló óvodások. Továbbá a fejlesztés fókuszában – mint az olvasástanulás egyik fontos tényezője – a fonológiai tudatosság fejlesztése kellene, hogy álljon.

Hivatkozások

- Adamikné Jászó, A. (2006). *Az olvasás múltja és jelene*. Budapest: Trezor Kiadó.
- Anthony, J. L., Aghara, R. G., Dunkelberger, M. J., Anthony, T. I., Williams, J. M., & Zhang, Z. (2011). What factors place children with speech sound disorders at risk for reading problems? *American Journal of Speech-Language Pathology*, 20, 146–160.
- Barbour, K., Keafer, K., & Scott, K. (2003). *Sounds of Speech*. Norfolk, VA: Phonological Processing Activities. NILD.
- Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research*, 38, 446–462.
- Blachman, B. A. (1997). Early intervention and phonological awareness: A cautionary tale. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 409–430). Lawrence Erlbaum Associates Publishers.
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read. *Nature*, 301, 419–421. doi:<https://doi.org/10.1038/301419a0>.
- Cowan, W. E., & Moran, M. J. (1997). Phonological awareness skills of children with articulation disorders in kindergarten to third grade. *Journal of Children's Communication Development*, 18, 31–38.
- Cs. Czachesz, E. (1998). *Olvasás és pedagógia*. Szeged: Mozaik Oktatási Stúdió.
- Csépe, V. (2005). *Kognitív fejlődés-neuropszichológia*. Budapest: Gondolat Könyvkiadó Kft.
- Csépe, V. (2006). *Az olvasó agy*. Budapest: Akadémiai Kiadó.

- De Jong, P. F., & Van Der Leij, A. (2003). Developmental changes in the manifestation of phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, 95, 22–40.
- EMMI (2013). 15/2013. (II. 26.) EMMI rendelet a pedagógiai szakszolgálati intézmények működéséről. URL: <https://net.jogtar.hu/jogszabaly?docid=a1300015.emm>.
- Foy, J. G., & Mann, V. (2001). Does strength of phonological representations predict phonological awareness in preschool children? *Applied Psycholinguistics*, 22, 301–325.
- Gereben, F., Fehérné Kovács, Z., Kas, B., & Mészáros, A. (2012). Beszéd- és nyelvi zavart mutató (beszédfogvatékos) gyermekek, tanulók komplex vizsgálatának diagnosztikus protokollja. In A. Torda (Ed.), *Diagnosztikai kézikönyv*. Budapest: Educatio Nonprofit Kft. URL: http://paks.tmps.hu/userfiles/files/diagnosztikai_kezikonyv_3fejezet.pdf.
- Goswami, U. (2002). Phonology, reading development and dyslexia: a cross-linguistic perspective. *Annals of Dyslexia*, 52, 141–163.
- Goswami, U. (2003). Phonology, learning to read and dyslexia: A cross-linguistic analysis. In V. Csépe (Ed.), *Dyslexia: Different Brain, Different Behaviour* (pp. 1–40). Boston, MA: Springer. doi:https://doi.org/10.1007/978-1-4615-0139-8_1.
- Göncz, L. (2003). A metanyelvi képességek fejlődése egynyelvű és kétnyelvű gyermekeknél. *Alkalmazott Nyelvtudomány*, 2, 5–20.
- Hayiou-Thomas, M. E., Carroll, J. M., Leavett, R., Hulme, C., & Snowling, M. (2017). When does speech sound disorder matter for literacy? the role of disordered speech errors, co-occurring language impairment and family risk of dyslexia. *Journal of Child Psychology and Psychiatry*, 58, 197–205. URL: <https://acamh.onlinelibrary.wiley.com/doi/10.1111/jcpp.12648>.

- Jordanidisz, A. (2009). A fonológiai tudatosság fejlődése az olvasástanulás időszakában. *Anyanyelv-pedagógia*, 2. URL: <https://www.anyanyelv-pedagogia.hu/cikkek.php?id=222>.
- Jordanidisz, A. (2010). A gyermekek fonológiai tudatosságának fejlődése 4–6 éves korban. In T. Váradi (Ed.), *IV. Alkalmazott Nyelvészeti Doktorandusz Konferencia* (pp. 46–59). Budapest: MTA Nyelvtudományi Intézet.
- Jordanidisz, A. (2015). *Magyar anyanyelvű gyermekek fonológiai tudatosságának fejlődése 4 és 10 éves kor között*. Budapest: Eötvös Loránd Tudományegyetem. URL: <http://doktori.btk.elte.hu/lingv/jordanidiszagnes/diss.pdf> doktori disszertáció.
- Jordanidisz, A. (2017). Magyar anyanyelvű gyermekek fonológiai tudatosságának fejlődése 4 és 10 éves kor között. In J. Bóna (Ed.), *Új utak a gyermeknyelvi kutatásokban* (pp. 203–220). Budapest: ELTE Eötvös Kiadó.
- Juhász, A. (2007). *Logopédiai vizsgálatok kézikönyve*. Budapest: Logopédia Kiadó.
- Juhász, V., & Kálló, V. (2017). A szótagtudat fejlesztése óvodás-, illetve kisiskoláskorban. *Anyanyelv-pedagógia*, 10. URL: https://www.anyanyelv-pedagogia.hu/img/keptar/2017_3/Anyp_X_2017_3_3.pdf.
- Kas, B., Lőrík, J., Molnárné Bogáth, R., Szabóné Vékony, A., & Szatmáriné Mályi, N. (2012). *Szól-e? Szűrőeljárás az óvodai logopédiai ellátásban*. Budapest: Logotech Kiadó.
- Kiss, R., Mokri, D., & Csapó, B. (2019). A fonológiai tudatosság online mérése óvodás gyermekek körében. *Alkalmazott Pszichológia*, 19, 35–54.
- Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. (2004). The development of early literacy skills among children with speech difficulties: A test of the "critical age hypothesis". *Journal of Speech, Language, and Hearing Research*, 47, 377–391.

- Preston, J. L., Hull, M., & Edwards, M. L. (2013). Preschool speech error patterns predict articulation and phonological awareness outcomes in children with histories of speech sound disorders. *American Journal of Speech-Language Pathology*, 22, 173–184.
- Raitano, N. A., Pennington, B. F., Tunick, R. A., Boada, R., & Shriberg, L. D. (2004). Pre-literacy skills of subgroups of children with speech sound disorders. *Journal of Child Psychology and Psychiatry*, 45, 821–835.
- Steklács, J. (2013). *Olvasási stratégiák tanítása, tanulása és az olvasásra vonatkozó meggyőződés*. Budapest: Nemzedékek Tudása Tankönyvkiadó.
- Thomas, E. M., & Sénéchal, M. (1998). Articulation and phoneme awareness of 3-year-old children. *Applied Psycholinguistics*, 19, 363–391.
- Tóth, D., & Csépe, V. (2009). Az olvasás fejlődése kognitív idegtudományi nézőpontból. *Pszichológia*, 4, 357–375.
- Webster, P. E., Plante, A., & Couvillion, L. M. (1997). Phonological impairment and prereading: Update on a longitudinal study. *Journal of Learning Disabilities*, 30, 365–375.

Afáziás személyek lexikai hozzáféréseinek vizsgálata olvasás közben

Kis Orsolya^{1,2,4}, Steklács János^{3,4}, Jakab Katalin¹, Klivényi Péter¹

¹*SZTE Szent-Györgyi Albert Orvostudományi Kar, Szent-Györgyi Albert Klinikai Központ,
Neurológiai Klinika*

²*SZTE Neveléstudományi Doktori Iskola*

³*PTE BTK Nevelés- és Oktatásméleti Tanszék*

⁴*MTA SZTE Metakogníció Kutatócsoport*

Abstract

As a result of research results exploring the functioning of linguistic abilities, the way of thinking about language has been formed. Recent views are dominated by an approach to a unified architecture of cognitive functions that states that linguistic and cognitive processes are based on common neural networks (Ralph et al., 2017; Blumstein & Amso, 2013; Kunert et al., 2015; Tóth, 2018a,b; Molnár-Tóth, 2021). Accordingly, reading disorders are explored through a comprehensive exploration of language, reading, and cognitive processes. Exploring the characteristics of reading disorders associated with adult acquired language disorder, i.e. aphasia as a syndrome, is of paramount importance not only in the diagnostic process but also in the therapeutic process. In our study, we present the first results of our measurement tool developed for the exploration of adult acquired reading disorders. Lexical access plays a central role in reading processes (Perfetti, 1999; Perfetti & Stafura, 2014). In our research, we observed the main components of linguistic and reading abilities of aphasic individuals ($n = 19$). During data analysis, correlation tests were performed. Based on our first results, the tasks measuring lexical access showed a significant correlation with several language and reading components, and we also experienced strong effects according to the internal structure of the tasks. Based on all this, the tests of our reading test measuring lexical access seem to be able to differentiate well among the main characteristics of reading disorders associated with aphasia.

Keywords: aphasia, reading, lexical access, connectionism, lexicon

1. Bevezetés

A kognícióval kapcsolatos kutatási eredmények az olvasáskutatásra is hatással voltak, ennek köszönhetően az olvasással kapcsolatos vizsgálatokban is

Email addresses: kis.orsolya0809@gmail.com (Kis Orsolya), steklacs@gmail.com (Steklács János), katalin.v.jakab@gmail.com (Jakab Katalin), klivenyi.peter@med.u-szeged.hu (Klivényi Péter)

számos elméleti irányzat jelent meg. A kezdeti elméletek a kognitív működés folyamatait egymástól elkülönült, reprezentáció alapú rendszerként értelmezték. A konnekcionizmus megjelenésével azonban változott a kognícióról való általános gondolkodásmód, vagyis az újabb irányzatok nem elkülönült rendszerekből álló mechanizmusként, hanem inkább egymással kölcsönhatásban működő, közös neurális architektúrával rendelkező folyamatokból álló egészként írják le. Mindezek hatással voltak az olvasással kapcsolatos nézetekre is (Coltheart et al., 1993; Alvermann et al., 2013; Stafura & Perfetti, 2017). Napjainkban az olvasást tágabb kontextusban, a nyelvi, az ortografikus és a kognitív folyamatok kölcsönhatásaként definiálják, vagyis az olvasás egy olyan kognitív folyamat, amely a nyelvi produkcióhoz és feldolgozáshoz szorosan kapcsolódik, emiatt pedig a nyelvi folyamatokban megjelenő károsodás az olvasás mechanizmusát is érinti (Perfetti, 1999; Hallowell & Chapey, 2008; Papathanasiou & Coppens, 2013; Riley & Kendall, 2013; Perfetti & Stafura, 2014; Stafura & Perfetti, 2017).

Jelen tanulmányban célunk a felnőttkori szerzett olvasászavarok vizsgálatára kialakított mérőeszközünk első eredményeinek ismertetése. Közleményünk fókuszában az olvasási folyamatok egyik legmeghatározóbb funkciójára, a lexikai hozzáférés mérésére létrehozott vizsgálati anyagunk ismertetése, valamint hatékonyságának bemutatása áll. Korábbi, nem magyar anyanyelvű afáziás személyekkel végzett kutatási eredmények alapján feltételezzük, hogy a szóhosszúság és a szófaj eltérő mértékben hat a különböző olvasási funkciókra (Sung et al., 2011; Kaye & Cherney, 2016; DeDe, 2017; Webster et al., 2018; Ciaccio et al., 2020). Kutatásunk legfőbb célja feltárni azt, hogy magyar anyanyelvű afáziás személyeknél milyen eltérések detektálhatók a szóhosszúság, valamint a szófaj alapján, továbbá mindezek mely nyelvi és olvasási komponensekkel mutatnak összefüggést.

2. Elméleti háttér

2.1. Olvasás

Az olvasás a nyelvi folyamatokkal szorosan összefügg, ugyanis számos feldolgozási mechanizmus a nyelvi és az olvasási folyamatokban is egyaránt szerepet játszik. Az olvasás megfelelő működéséhez alapvetően három főbb feldolgozási rendszer, a nyelvi, az ortografikus, valamint az előzetes tudást és a megismerő folyamatokat is magába foglaló általános kognitív folyamatok megfelelő működése, továbbá ezek interakciója szükséges (Perfetti & Stafura, 2014; Stafura & Perfetti, 2017).

Az olvasáskutatás kezdeti modelljei a kognitív tudományokban uralkodó reprezentáció alapú irányzatot képviselték. Az egyik legelterjedtebb olvasást leíró modell a kettős útmodell (Coltheart et al., 1993), amely alapján a vizuális inputból érkező inger feldolgozása alapvetően két külön útvonalon mehet végbe: (1) az egyik a lexikai út, amely során az olvasás a szavakhoz tartozó lexikai-szemantikai információk elemzésén és feldolgozásán keresztül történik, (2) a másik a nem lexikális útvonal, ennek során az olvasás a graféma-fonéma (betű-hang) megfeleltetési szabályok megfelelő használata révén történik, vagyis ebben az esetben lexikai feldolgozása nem megy végbe. Ez a két, egymástól különálló útvonal a fonológiai pufferben egyesül, amely a szóprodukcióhoz szükséges szóalak kiválasztásáért felelős (Riley & Kendall, 2013). A konnekcionizmus térhódításának hatása megjelenik a kognícióval kapcsolatos paradigmákban. A napjainkban uralkodó modellek az olvasáshoz szükséges folyamatok átfogó leírását adják, ennek köszönhetően az olvasás tágabb értelmezése jelenik meg, ahol mind az adatvezérelt (alulról felfelé), mind a fogalomvezérelt (felülről lefelé) folyamatok, valamint a nyelvi, az alapvető kognitív és szociokulturális tényezők és ezek interakciói meghatározó szerepet kapnak (Alvermann et al., 2013; Stafura & Perfetti, 2017).

Egyik ilyen, az olvasási folyamatok komplex leírását adó modell Perfetti (1999) olvasást leíró keretrendszere, amelyben az olvasáshoz szükséges működés leírásakor nemcsak az ortografikus folyamatok, de az ehhez szükséges nyelvi és

általános kognitív folyamatok egyaránt szerepet kapnak. Perfetti (1999) alapján az olvasás az alacsonyabb szintű perceptuális folyamatoktól elkezdve egészen a magasabb szintű megértési és következtetési folyamatokig számos olyan mechanizmust és tudástípust felölel, amelyek nem csak az olvasás során használatosak. Az olvasás során a lexikon működése kiemelt jelentőséggel bír (Perfetti, 1999; Perfetti & Stafura, 2014; Stafura & Perfetti, 2017). A továbbiakban erről frunk bővebben.

2.2. A lexikon szerepe az olvasásban

A lexikai feldolgozás kiemelt jelentőséggel bír az olvasás megfelelő működésében, ugyanakkor ezt megelőzi néhány folyamat (Perfetti, 1999; Perfetti & Stafura, 2014). Az olvasás először szóazonosítással kezdődik, ekkor történik a vizuális inputból érkező, a perceptuális folyamatok által azonosított információk feldolgozása, vagyis megtörténik a vizuálisan észlelt ingerek azonosítása, amelynek eredménye a szavakat alkotó grafémák aktiválása. Mindeközben az egyes betűkhöz tartozó fonológiai egységek aktivációja is végbemegy, vagyis a szó azonosítása magában foglalja a grafémák és fonémák azonnali, együttes aktivációját. Eközben azonnal elindul a szemantikai aktiváció is, amely kétirányú módon közvetít a lexikon, valamint az ortografikus és fonológiai folyamatok között. A szóazonosítási folyamatok következtében aktiválódnak a szóhoz tartozó lexikai elemek, vagyis a szóazonosítási folyamat eredménye a szóalak és néhány, az aktiválódott szóhoz kapcsolódó, a mondat és szövegértéshez szükséges információ (grammatikai forma, tematikus szerep stb.) előhívása (Perfetti, 1999).

A lexikai hozzáférés egy olyan amodális folyamat, amely bármilyen input esetén hasonlóan működik (Perfetti, 1999; Perfetti & Stafura, 2014). Ennek segítségével történik a már létrejött szóazonosítási folyamat és a mentális lexikonban tárolt egységek közötti kapcsolat létrehozása. A mentális lexikon a szavak jelentésének tárhelye, amely a fogalmi rendszerrel átfedésben van. Ugyanakkor nem pusztán a szavak jelentése tárolódik itt, hanem a szóhoz tartozó egyéb információ is, mint például a szó hangalakja, szófaja és mondatokban betöltött szerepe. Az empirikus eredményekkel is alátámasztott szemantikai-háló model-

lek a mentális lexikon szerveződését hálózatos modellben definiálják, eszerint a szavak önmagukban álló egésként tárolódnak, amelyek a közöttük lévő kapcsolatok révén hálózatokba szerveződnek (összefoglaló a mentális lexikon felépítéséről bővebben lásd: Honbolygó, 2008). A szó kiolvasásával egyidejűleg a szemantikai információk is elérhetővé válnak az olvasó számára, ugyanakkor a szavaknak több jelentése is lehetséges (Perfetti, 1999).

A lexikai hozzáférést magyarázó modelleket tekintve alapvetően két főbb irányzat különíthető el, vannak a lexikon és vannak a tanulás-alapú modellek. Ezek alapvetően eltérő magyarázatot szolgáltatnak a lexikális reprezentációk és szabályok felépítését és működését illetően. A lexikon-alapú modellek feltételezik, hogy a szavak tő formájában tárolódnak, amelyekhez toldalékok csatolásával további szavak és jelentések hozhatók létre. A tanulás-alapú modellek a lexikon működését univerzálisnak tekintik, vagyis a hangsúly a már meglévő előzetes tudás hatásán van, a nyelvelsajátítás a kontrasztív tulajdonságok megkülönböztetése révén történik (Milin et al., 2018). A kontextus függvénye az, hogy melyik jelentés aktiválódik éppen. A megfelelő szójelentés kiválasztásával kapcsolatban több elmélet is napvilágot látott (Perfetti, 1999). A szelektív hozzáférési modell alapján a kontextushoz legjobban kapcsolódó jelentés aktiválódik csak, a többi jelentés nem mutat aktivációt (Glucksberg et al., 1986). A többszörös hozzáférési modell ennél valamelyest komplexebben írja le a szelekciós folyamatokat. E szerint a kontextusnak eleinte nincs hatása, a megfelelő jelentés kiválasztását egy nagyon rövid, általános aktiválási folyamat előzi meg, amelynek során egy automatikus és nagyon gyors folyamat következtében egy szónak egyszerre több jelentése is aktiválódik. Ezt követően történik a kontextushoz legjobban illeszkedő jelentés kiválasztása, amelyre nemcsak a kontextus felépítése van hatással, de a kétértelműséget előidéző szószerkezet is (Onifer & Swinney, 1981; Seidenberg et al., 1982). Az elrendelt keresési modell alapján a szószelekció korlátozott, többszörös hozzáféréseken alapszik, amelyet a szó jelentéseinek relatív gyakorisága szabályoz. Eszerint a szó leggyakrabban használt jelentése folyamatosan aktív, ugyanakkor kontextus hatására másik jelentés aktiválódhat (Duffy et al., 1988; Perfetti, 1999).

A szósztintű feldolgozási folyamatok befolyásolják a magasabb szintű megértési folyamatok hatékonyságát. Az eredményes, hatékony olvasás alapja a szóról való tudás komponenseinek megfelelő használata, ez magába foglalja a szóalakokkal (nyelvtani osztály), a szójelentéssel, valamint a szóhasználattal (ide tartozik a pragmatika is) kapcsolatos tudást. Ennek megfelelően a hatékony olvasás alapja nem annak gyorsasága, hanem a szó azonosításához szükséges komponensek megfelelő előhívása (Perfetti, 2007). Afázias személyekkel végzett vizsgálatok alapján a szöveg felépítése, azon belül is a szavak olyan tulajdonsága, mint a szófaj, a szóhosszúság és a szógyakoriság hatással van a szövegfeldolgozási képességekre. A szóhosszúság tekintetében szemmozgásos vizsgálatok eredményei szerint afáziasoknak a hosszabb szavak olvasásához több időre volt szükségük (DeDe, 2017). Nemcsak önmagában a szavak hosszúsága, de a szöveghosszúság is szignifikáns hatással van az olvasási időre, ugyanakkor Webster és munkatársai (2018) eredményei szerint mindez a szövegfeldolgozás pontosságát nem befolyásolja. Ellenben a szógyakoriság és a szófaj szövegfeldolgozási képességekre gyakorolt szignifikáns hatását sikerült kimutatni (Kaye & Cherney, 2016): a könnyebb szerkesztésű, nagy gyakoriságú szavakat tartalmazó szövegek feldolgozása szignifikánsan pontosabbnak mutatkozott, mint a kevésbé gyakori, periférián lévő szavakból felépülő szövegeké. A szószerkezetet tekintve nem minden szó feldolgozása működik egyformán, ezt igazolták Ciaccio és kollégáinak (2020) eredményei is, miszerint az általuk vizsgált afázias személyek több hibát vétettek az elő- és utótaggal rendelkező összetett szavak olvasásakor, mint az egyszerű szavak olvasásakor, továbbá a morfoszintaktikai feldolgozás olyan tényezőktől függött, mint a szintaktikai szerkezet, a szógyakoriság, a szófaj és a szószerkezet. A szófaj tekintetében hosszabb olvasási időt detektáltak a főnevek esetében a határozószókhoz képest (Sung et al., 2011).

Összegezve a fentieket látható, hogy a lexikon az olvasás rendszerében központi helyet foglal el, ugyanis két fontos folyamat között helyezkedik el. Az egyik ilyen a szóazonosítás, amely során a nyelvi és ortografikus tudás gyors és pontos használata szükséges. A másik a megértési, feldolgozási rendszer, amely a szóazonosítási és lexikai folyamatok alapján működik. Mindezek alapján lát-

ható, hogy az írott szavak formáinak és jelentéseinek ismerete központi szerepet játszik az olvasásban és hatással van a szövegértésre (Perfetti & Stafura, 2014).

2.3. Afázia

A fentiek alapján látható, hogy a kognícióval kapcsolatos kutatási eredmények hatására változott a nyelvről és a kognícióról alkotott gondolkodásmód. Mindezek hatására az afáziával kapcsolatos értelmezésekben is jelentős változás észlelhető. A következőkben Tóth (2018a) alapján az afáziák definiálásának főbb mérföldköveiről teszünk említést.

A kezdeti nézeteket a klinikai-funkcionális szemlélet uralta, amely alapján a nyelvi zavar definiálása a károsodott nyelvi komponensek feltárásán keresztül történt. Ennek megfelelően az afázia pontos meghatározását a nyelvi komponensek állapotának leírása adta. A neurolingvisztikai megközelítés viszont az afáziák feltárásakor már a sérült nyelvi szintek megismerésén keresztül közelít, ennek megfelelően az afáziák osztályozásának alapját az egyes nyelvi szintek adják. Tóth (2018a) az elmélet főbb képviselőjeként Whitakert (2007) említi, aki a következő nyelvi szintek alapján differenciál: fonológia, morfológia, szintaktika, szemantika, narratíva, alexia, agráfia. A neurolingvisztikai megközelítés már említést tesz a nyelvhez társuló kognitív folyamatok működésének fontosságáról (Tóth, 2018a,b; Molnár-Tóth, 2021). Napjaink nézeteit a kognitív funkciók egységes működésének értelmezése adja, ugyanis feltételezhetően a kognitív mechanizmusok mögött egy közös neurális architektúra áll, amelynek hatására a különböző folyamatok egymással interakcióban vannak (Hagoort, 2005; Baldo & Dronkers, 2007; Lindenberg et al., 2007; Amunts et al., 2010; Denes, 2011; Blumstein & Amso, 2013; Kunert et al., 2015; Tóth, 2018a,b; Molnár-Tóth, 2021).

Több magyar anyanyelvű afáziás személyekkel végzett kísérlet is ezt az álláspontot képviseli. Bánréti (2014) a munkamemória korlátozott kapacitásának nyelvre gyakorolt hatását tárta fel. Afáziás személyek mondatismétlési teszten nyújtott teljesítményének részletes elemzése alapján úgy tűnik, hogy az ismétlés nem elsődlegesen az auditív emlékezeti tároló, hanem mondatszerkezeti informá-

ciók feldolgozása és annak memóriában való tárolása alapján történik, továbbá a mondatismétlés során az afáziás személyek felhasználják a nyelvtani feldolgozáshoz szorosan kapcsolódó egyéb, még jól működő mentális rendszereket is. Mészáros (2007) 10 magyar anyanyelvű agrammatikus afáziás személy grammatikalitási döntési feladatokban produkált teljesítményét elemezte. Eredményei alapján összefüggés van a morfoszintaktikai feldolgozás és az ehhez szükséges munkamemória kapacitása között, ugyanis a grammatikailag helyes mondatoknál nincs szükség újraelemzésre, szemben az agrammatikus mondatokkal, ahol az újraelemzéshez viszont megfelelő munkamemória kapacitás szükséges. Afáziás személyek esetében ez a kapacitás korlátozott, így valószínűsíthető, hogy ez áll a morfoszintaktikai szerkezetek feldolgozásakor jelentkező zavarok hátterében. Szöllősi, Lukács & Zakariás (2015) a végrehajtó funkciók működését vizsgálták afáziás személyeknél. Eredményeik alapján afáziás személyeknél érintett a válaszalapú konfliktus gátlása és a munkamemória frissítése, továbbá összefüggést találtak a nyelvtani szerkezetek megértése és a munkamemória frissítése között.

Mindezek alapján az afáziát érdemes szindrómaként értelmezni, ennek következtében pedig a szerzett nyelvi zavar feltárásakor a nyelvben jelentkező tünetek leírásán túl érdemes figyelmet fordítani a többi, a nyelvi működést meghatározó kognitív folyamat vizsgálatára, beleértve többek között az olvasást is.

3. A kutatás célja

A kognitív funkciók elosztott hálózatos működése révén a szerzett nyelvi zavar következtében az olvasásban is jelentkeznek tünetek. Kutatásunk célja egy olyan olvasást vizsgáló eljárás kidolgozása, amely segítségével a felnőttkori szerzett nyelvi zavarok diagnosztikus eszköztárát bővíthetjük, ezzel segítve az afázia, mint szindróma univerzális feltárását. A pontos diagnosztika segítségével specifikus, egyénre szabott terápia tervezhető, ezzel növelve a hatékonyabb terápiás kimenetet, segítve a személy életminőségének javítását.

A fentiek alapján vizsgáltuk a szavak tulajdonsága, azon belül a szóhosszúság és a szófaj olvasási képességekre gyakorolt hatását. Jelenleg magyar anyanyelvű

afáziás személyekkel kapcsolatosan nem állnak rendelkezésre ilyen jellegű adatok. A nemzetközi szakirodalmi eredményekre alapozva feltételeztük, hogy a szavak hosszúsága és a szófaj szignifikáns hatással van az olvasási képességekre. Kutatásunk legfőbb kérdése, hogy a szavak hosszúsága és szófaja szerint milyen együttjárások tapasztalhatók az általunk vizsgált főbb nyelvi komponensek, valamint az olvasás különböző folyamatai között. Feltételeztük, hogy a különböző szótagszámú szavak olvasása különböző komponensekkel mutat összefüggést. Feltételeztük, hogy a szófaj szintén hatással van az olvasási képességekre, vagyis azt, hogy szófaj szerint eltérő együttjárásokat tapasztalunk majd. Amennyiben feltételezésünk beigazolódik, úgy tesztünk ezen része alkalmas a finomabb összefüggések kimutatására.

A jelen tanulmányban olvasásvizsgálatunk lexikai hozzáférést mérő feladatait elemezzük. A fentiek alapján látható, hogy az olvasás folyamatai során a lexikai hozzáférés működése meghatározó, emiatt vizsgálóeljárásunkban is kiemelt jelentőséget tulajdonítottunk a lexikai hozzáférést mérő feladatok kidolgozásának. Ebben a tanulmányban célunk a lexikai hozzáférést mérő feladatok és az olvasás többi komponense közötti összefüggések elemzése, továbbá vizsgáljuk a feladatok belső szerkezetének működését is.

4. Résztvevők és módszer

4.1. Résztvevők

Kutatásunkban összesen 19 stroke-on átesett afáziás személy (8 nő, 11 férfi) vett részt, akik átlagéletkora 62,26 (44–84) év volt. Az afázia típusát és súlyosságát a Western Afázia Teszt (Osmanné, 1991) magyar nyelvű változatával állapítottuk meg, amely alapján a vizsgálatban résztvevőknél az afázia súlyosságát tekintve 11 főnél enyhe, 8 személynél közepesen súlyos afázia igazolódott, az afázia típusát illetően 9 esetben anomikus, 5 főnél vezetésszerű, 3 főnél transzkortikális motoros és 2 főnél Wernicke típusú afáziát állapítottunk meg.

A résztvevők minden esetben jobbkezesek voltak, akiknél nem állt fent társútnetként beszédzavar, vagyis sem dysarthria, sem apraxia, továbbá a demenci-

ával vagy depresszióval diagnosztizált személyek kizárásra kerültek. A vizsgálatban való részvétel önkéntes alapon történt, amely a Szegedi Tudományegyetem Szent-Györgyi Albert Orvostudományi Kar Szent-Györgyi Albert Klinikai Központ Neurológiai Klinikán valósult meg. Vizsgálatunkat a 70/2020-SZTE számú etikai engedély birtokában végeztük.

4.2. Módszer

A kutatásban részt vevő afáziás személyek nyelvi képességprofiljának feltárása a Western Afázia Teszt (Osmanné, 1991) magyar nyelvű változatával történt. Ez alapján a nyelvi rendszer négy főbb komponenséről, a spontán beszéd fluenciájáról, a beszédmegértésről, az ismétlésről és a megnevezésről kaptunk információt, majd ennek a négy fő területnek a segítségével meghatároztuk az afázia együtthatót, amelyet az afáziakvóciens (AQ) demonstrál.

A pácienseknek a feladatokat egy laptopon keresztül mutattuk be. Ehhez egy informatikus segítségével elkészítettük a *Neurolinguistics Research 1.0* elnevezésű, online felületen elérhető programot, amely a reakcióidőt, valamint a lexikai hozzáférés és a morfoszintaktikai elemek feldolgozását mérő feladatok esetében a válaszok helyességét is rögzíti.

A kutatásunkhoz összeállított olvasásteszt összesen 10 feladatot tartalmaz, amely az alábbi főbb feladatcsoportokból áll: fonológia, lexéma, morfoszintaktika és szövegfeldolgozás (1. táblázat).

A fonológiát magánhangzók ($n = 26$) és mássalhangzók ($n = 28$) olvasásával vizsgáltuk. A vizsgálatban részt vevő személyeknek a feladata az volt, hogy a képernyőn látható ingereket olvassák fel hangosan. Az ingereket Meixner Ildikó 3. és 4. osztályos tanulóknak készített olvasást mérő lapjai közül választottuk ki (Meixner, 1995; Juhász, 2003).

A lexikát két feladatcsoporttal, a szóolvasással és a lexikai hozzáféréssel vizsgáltuk. A szóolvasás felmérésére négy feladatot állítottunk össze, kettő esetében szemantikai tartalommal rendelkező, kettő esetében szemantikai tartalommal nem rendelkező, álszavak felolvasását kértük a vizsgált személyektől. A szótágok és a szavak kiválasztásánál az ingereket Meixner Ildikó 3. és 4. osztályos

1. táblázat. Az olvasás teszt felépítése, külön kiemelve a jelen tanulmányunk elemzésének fókuszában álló feladatokat.

Feladatok		
<i>Fonológia</i>	magánhangzók olvasása ($n = 26$)	
	mássalhangzók olvasása ($n = 28$)	
<i>Lexika</i>	szóolvasás	szótagok ($n = 24$)
		két betűből álló szavak ($n=24$)
		szavak ($n = 50$)
		álszavak ($n = 36$)
	lexikon	kép-szókép párosítása ($n = 20$)
		szó-álszó lexikai döntés ($n = 50$)
		grammatikus/agrammatikus mondatok ($n = 20$)
<i>Szövegfeldolgozás</i>	szövegértés ($n = 21$)	

tanulóknak készített Olvasási felmérőlapjaiból (Meixner, 1995; Juhász, 2003), valamint Sipos (2013) feladatgyűjteményéből válogattuk ki. Az ingerek kiválasztásakor a szógyakoriságot is figyelembe vettük, amelyhez a Magyar Nemzeti Szövegtár által közzétett szógyakorisági listát használtuk. Az álszavak esetében Racsmány és munkatársainak (2006) verbális munkamemória mérőeljárásában az álszóismétlés felmérésére alkalmazott szavakat használtuk, megváltoztatva az ingerbemutató modalitását, vagyis az eredeti vizsgálathoz képest az ingerek prezentálása nem auditív, hanem vizuális úton történt. A résztvevők feladata minden esetben az volt, hogy olvassák fel hangosan a képernyőn látható ingereket. A szemantikai tartalommal nem rendelkező ingerek esetében szótagok ($n = 24$) és álszavak ($n = 36$) hangos olvasása volt a feladat. A szemantikai tartalommal rendelkező szavak felolvasását két betűből álló szavak ($n = 24$), valamint az egy szótagútól egészen a több szótagúig terjedő, összetett szavakat ($n = 50$) is tartalmazó ingerek bemutatásával mértük.

A lexikai hozzáférés mérésére, amelynek során a lexikai információkhoz való hozzáférést vizsgáltuk, három feladatot állítottunk össze: kép-szókép párosítása, szó-álszó döntési feladat és grammatikus/agrammatikus mondatokról ho-

zott döntés. A kép-szókép párosítása feladatnál a résztvevők feladata az volt, hogy három lehetőség közül válasszák ki a képernyő közepén bemutatott képhez tartozó megfelelő szót. A három alternatív válasz közül az egyik a helyes válasz, a másik a célszóhoz hasonló, de fonológiai hibát tartalmazó, a harmadik pedig a célszóval valamilyen szemantikai kapcsolatban lévő alternatív lehetőség volt. A feladat során főneveket ($n = 10$) és igéket ($n = 10$) mutattunk, az ingerek kiválasztásakor szintén a Magyar Nemzeti Szövegtár által közzétett szógyakorisági lista alapján választottunk gyakori, kevésbé gyakori, valamint ritka szavakat. Az ingerként felhasznált képeket egy nyílt hozzáférésű, ingyenesen használható adatbázisból válogattuk (<https://thenounproject.com/>). A szó-álszó lexikai döntési feladat esetében a fentiekben említett szógyakorisági lista alapján, minden szóosztályból (főnév, ige, melléknév, névmás, határozó) összesen 25, 1–5 szótagig terjedő célingert választottunk ki. Minden egyes célingerhez a Neurolinguistics Research 1.1 programmal, a magyar fonotaktikai szabályok, a szótagszám, valamint a CVC (magánhangzók-mássalhangzó kapcsolódások) figyelembevételével generáltunk álszópárt. Így összesen ez a feladat 50 ingert (25 szó, 25 álszó) tartalmazott. A betegek feladata az volt, hogy döntsek el a képernyőn látható szóról, hogy valódi, magyar, jelentéssel rendelkező szó-e. Amennyiben igen, akkor a feladat az volt, hogy kattintsanak a pipa ikonra, amennyiben nem, akkor az x-re. Mivel ennél a feladatnál célunk a szavakhoz tartozó lexikai hozzáférés mérése volt, emiatt a résztvevőknek a bemutatott ingereket nem kellett felolvasniuk. A grammatikus/agrammatikus mondatokról hozott döntési feladat esetében célunk a mondat szintű információk feldolgozásának vizsgálata volt, ennek során a résztvevőknek morfoszintaktikailag helyes, illetve hibás szerkesztésű mondatokat mutattunk. A feladatuk az volt, hogy döntsék el, a bemutatott inger helyes vagy helytelen. A mondatok szerkesztésénél a magyar nyelvre legjellemzőbb esetragokat használtuk fel (Kiefer, 2011), amelyek a következők voltak: accusativus (-t), dativus (-nak, -nek), instrumentális (-val, -vel), causalis-finalis (-ért), translativus-factitivus (-vá, -vé), inessivus (-ban, -ben), suppressivus (-on, -en, -ön), adessivus (-nál, -nél), allativus (-hoz, -hez, -höz), ablativus (-tól, -től), terminativus (-ig). A betegeket az előzőekhez

hasonlóan itt is arra kértük, hogy amennyiben helyes a bemutatott mondat, úgy kattintsanak a pipára, ha hibás, akkor az x-re.

Az olvasásteszt utolsó részében a szövegfeldolgozást vizsgáltuk, amelyhez egy, a *PIRLS* (Progress in International Reading Literacy Study) által közzétett, nyílt hozzáférésű szöveget választottunk. A szöveggel kapcsolatos kérdések egy részét átalakítottuk. A szövegértés értékelésénél az explicit ($n = 10$) és implicit ($n = 11$) információk feldolgozását külön értelmezzük.

5. Eredmények

Tanulmányunk célja a lexikai információkhoz való hozzáférés vizsgálatára létrehozott mérőeljárásunk tesztelése volt, amelyhez szerzett nyelvi zavarral diagnosztizált személyek nyelvi és olvasási folyamatait vizsgáltuk. Elemzéseink során célunk afázias személyeknél a lexikai információk feldolgozása és az olvasás többi folyamata közötti összefüggés feltárása volt, ehhez a lexikai hozzáférést mérő három feladatot, vagyis a kép-szókép párosítása, a szó-álszó lexikai döntés, valamint a grammatikus/agrammatikus mondatok döntési feladata során produkált eredmények és az olvasásteszt többi feladata közötti együttjárásokat elemeztük, mindehhez korrelációs elemzéseket végeztünk, amelyhez az *IBM SPSS Statistics 23* programot használtuk. Mivel adataink nem követték a normál eloszlást, így a korrelációs elemzések során Spearman-féle nem-paraméteres korrelációs próbát alkalmaztunk, a hibásáv minden esetben 95%-os konfidencia-intervallum volt.

5.1. A kép-szókép párosítása feladat eredményei

Elsőként a fentiekben ismertetett, a lexikai hozzáférés mérésére használt első feladat, a kép-szókép párosítása feladat összesített eredményeit elemeztük. Szignifikáns pozitív együttjárás mutatkozott a magánhangzók ($r(18) = 0,738$; $p < 0,001$), a mássalhangzók ($r(18) = 0,607$; $p = 0,006$), a szótagok ($r(18) = 0,658$; $p = 0,002$), a szavak ($r(18) = 0,655$; $p = 0,002$), az álszavak ($r(18) = 0,531$; $p = 0,019$), a szó-álszó lexikai döntés ($r(18) = 0,731$; $p = 0,001$), a grammatikus/agrammatikus mondatok megítélése ($r(18) = 0,481$; $p = 0,037$), valamint

a szövegértési feladatoknál az implicit információk feldolgozása ($r(18) = 0,512$; $p = 0,025$) között. A többi feladat esetében nem mutatkozott szignifikáns korreláció (bővebben lásd: 2. táblázat).

2. táblázat. A kép-szókép párosítása feladat és az olvasás teszt egyes feladatai közötti együttjárások (a statisztikailag szignifikáns együttjárásokat vastagon szedtük).

Feladatok	összesített		főnév		ige	
	r	p	r	p	r	p
spontán beszéd	0,349	0,143	0,080	0,744	0,385	0,104
auditoros verbális megértés	-0,019	0,940	-0,184	0,450	0,095	0,697
ismétlés	-0,308	0,200	-0,358	0,132	-0,173	0,478
megnevezés	0,149	0,543	0,278	0,250	0,051	0,834
magánhangzók	0,738	0,001	0,585	0,008	0,672	0,002
mássalhangzók	0,607	0,006	0,515	0,024	0,465	0,045
szótagok	0,658	0,002	0,558	0,013	0,522	0,022
két betűből álló szavak	0,445	0,056	0,491	0,033	0,455	0,050
szavak	0,655	0,002	0,517	0,024	0,651	0,003
álszavak	0,531	0,019	0,447	0,055	0,507	0,027
szó-álszó lexikai döntés	0,731	0,000	0,767	<0,001	0,705	<0,001
grammatikus/agrammatikus mondatok	0,481	0,037	0,386	0,102	0,441	0,059
szövegértés – összesített	0,538	0,017	0,536	0,018	0,506	0,027
szövegértés – explicit	0,442	0,058	0,251	0,299	0,542	0,017
szövegértés – implicit	0,512	0,025	0,669	0,002	0,397	0,093

A kép-szókép párosítása feladaton belül külön, szófaj szerint is elemeztük a résztvevők teljesítményét. Szignifikáns pozitív korreláció mutatkozott a kép-szókép párosítása feladatban használt főnevek feldolgozása és a magánhangzók ($r(18) = 0,585$; $p = 0,008$), a mássalhangzók ($r(18) = 0,515$; $p = 0,024$), a szótagok ($r(18) = 0,558$; $p = 0,013$), a két betűből álló szavak ($r(18) = 0,491$; $p = 0,033$) és a hosszabb szavak ($r(18) = 0,517$; $p = 0,024$) olvasása között, továbbá a szó-álszó lexikai döntés ($r(18) = 0,767$; $p < 0,001$), és a szövegértési feladatokon belül az implicit információk feldolgozása ($r(18) = 0,669$;

$p = 0,002$) között. A többi feladat esetében nem mutatkozott együttjárás. Vizsgáltuk a kép-szókép párosítása feladatban használt igék és a nyelvi, valamint az olvasás feladatok közötti esetleges együttjárásokat is. Szignifikáns pozitív korrelációt találtunk a mássalhangzók ($r(18) = 0,465$; $p = 0,045$), a szótagok ($r(18) = 0,522$; $p = 0,022$), a két betűből álló szavak ($r(18) = 0,455$; $p = 0,050$), a hosszabb szavak ($r(18) = 0,651$; $p = 0,003$), valamint az álszavak ($r(18) = 0,507$; $p = 0,027$) olvasása között, továbbá a szó-álszó lexikai döntés ($r(18) = 0,705$; $p < 0,001$) és az explicit szövegszintű információk feldolgozása között ($r(18) = 0,542$; $p = 0,017$). A többi feladat esetében nem mutatkozott együttjárás.

A fenti eredmények alapján látható a szófaj hatása az olvasásra, ugyanis a főnevek és az igék esetében csak részben hasonló együttjárásokat tapasztaltunk. Mind a két szófaj esetében szignifikáns pozitív korreláció mutatkozott a mássalhangzók, a szótagok, a két betűből álló szavak, a hosszabb szavak, valamint a szó-álszó lexikai döntés között. Néhány feladat esetében szófaj szerinti eltérések jelentek meg. A magánhangzók olvasása, valamint az implicit szövegszintű információk feldolgozása csak a főnevekkel, az álszavak olvasása és az explicit szövegszintű információk feldolgozása pedig csak az igékkel mutatott együttjárást.

5.2. Szó-álszó lexikai döntés feladat eredményei

A lexikai hozzáférés mérésére használt másik feladatunk a szó-álszó lexikai döntés volt, amely során a vizsgálati személyeknek növekvő szótagszámú, különböző szófajú ingerekről kellett döntést hozniuk. Az összesített eredményeket tekintve szignifikáns pozitív együttjárás mutatkozott a megnevezési képességek ($r(18) = 0,553$; $p = 0,014$), a magánhangzók ($r(18) = 0,508$; $p = 0,027$), a két betűből álló szavak ($r(18) = 0,598$; $p = 0,007$), a hosszabb szavak olvasása ($r(18) = 0,509$; $p = 0,026$), a kép-szókép párosítása ($r(18) = 0,731$; $p < 0,001$), a mondatok helyességének megítélése ($r(18) = 0,674$; $p = 0,002$), valamint az implicit szövegszintű információk feldolgozása ($r(18) = 0,631$; $p = 0,004$) kö-

zött. Az eredményeket külön elemeztük szótagszámonként, valamint szófajonként is.

Szótagszámonként elemezve az eredményeket az egyszótagú szavak feldolgozása szignifikáns pozitív korreláció mutatott a magánhangzók olvasása ($r(18) = 0,538$; $p = 0,018$), valamint a kép-szókép párosítása ($r(18) = 0,605$; $p = 0,006$) és az implicit szövegszintű információk feldolgozása ($r(18) = 0,615$; $p = 0,005$) között. A kétszótagú szavak esetében összefüggést találtunk a két betűből álló szavak ($r(18) = 0,538$; $p = 0,018$) és a hosszabb szavak olvasása ($r(18) = 0,657$; $p = 0,002$), valamint a mondatok szintaktikai szerkezetének megítélése között ($r(18) = 0,487$; $p = 0,035$). A három szótagból álló szavak feldolgozása egyedül a megnevezési képességekkel mutatott szignifikáns pozitív korrelációt ($r(18) = 0,459$; $p = 0,048$). A négy szótagú szavak feldolgozása együttjárást mutatott a megnevezési képességekkel ($r(18) = 0,729$; $p < 0,001$), a két betűből álló szavak ($r(18) = 0,637$; $p = 0,003$) olvasásával és a mondatok helyességének megítélésével ($r(18) = 0,668$; $p = 0,002$). Szignifikáns pozitív korreláció mutatkozott az öt szótagból álló szavak feldolgozása és a magánhangzók ($r(18) = 0,480$; $p = 0,037$), a szótagok ($r(18) = 0,497$; $p = 0,030$), valamint a két betűből álló szavak ($r(18) = 0,564$; $p = 0,012$) olvasása, továbbá a kép-szókép párosítása ($r(18) = 0,540$; $p = 0,017$) között (bővebben lásd: 3. táblázat).

Az eredményeket szófajonként elemezve a főnevek és a határozószók feldolgozása mutatta a legtöbb együttjárást a nyelvi és az olvasásteszt egyes komponenseivel. Szignifikáns pozitív együttjárás mutatkozott a főnevek és a spontán beszéd információtartalma és fluenciája ($r(18) = 0,492$; $p = 0,033$), a magánhangzók ($r(18) = 0,551$; $p = 0,014$), a két betűből álló szavak ($r(18) = 0,623$; $p = 0,004$) olvasása, továbbá a kép-szókép párosítása ($r(18) = 0,653$; $p = 0,002$), a mondatok szintaktikai szerkezetének megítélése ($r(18) = 0,641$; $p = 0,003$), valamint az implicit szövegszintű információk feldolgozása ($r(18) = 0,610$; $p = 0,006$) között. Szignifikáns pozitív korreláció mutatkozott az igék feldolgozása a két betűből álló szavak olvasása ($r(18) = 0,522$; $p = 0,022$), a kép-szókép párosítása ($r(18) = 0,544$; $p = 0,016$), valamint a szövegértés

3. táblázat. A szó-álszó lexikai döntés feladat és az olvasás teszt egyes feladatai közötti együttjárások elemzése a szótagszámok függvényében (a statisztikailag szignifikáns együttjárásokat vastagon szedtük).

Feladatok	szótagszám									
	1		2		3		4		5	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
spontán beszéd	0,249	0,303	0,449	0,054	0,141	0,564	0,234	0,335	0,039	0,873
auditoros verbális megértés	-0,055	0,824	0,098	0,691	-0,429	0,067	-0,097	0,692	-0,046	0,852
ismétlés	-0,036	0,883	0,262	0,279	-0,166	0,497	0,133	0,588	0,030	0,903
megnevezés	0,339	0,156	0,350	0,142	0,459	0,048	0,729	<0,001	0,352	0,139
magánhangzók	0,538	0,018	0,078	0,751	0,253	0,296	0,117	0,632	0,480	0,037
mássalhangzók	0,132	0,591	0,213	0,240	0,248	0,305	0,076	0,758	0,421	0,073
szótagok	0,339	0,156	-0,018	0,941	0,068	0,783	0,003	0,989	0,497	0,030
két betűből álló szavak	0,412	0,080	0,538	0,018	0,403	0,087	0,637	0,003	0,564	0,012
szavak	0,282	0,242	0,657	0,002	0,273	0,259	0,196	0,420	0,400	0,089
álszavak	0,152	0,535	0,310	0,196	0,209	0,392	-0,134	0,584	0,319	0,184
kép-szókép párosítása	0,605	0,006	0,378	0,111	0,443	0,058	0,325	0,175	0,540	0,017
grammatikus/agrammatikus mondatok	0,357	0,133	0,487	0,035	0,315	0,189	0,668	0,002	0,328	0,170
szövegértés – összesített	0,557	0,013	0,398	0,092	0,004	0,986	0,359	0,131	0,371	0,118
szövegértés – explicit	0,252	0,298	0,447	0,055	-0,140	0,567	0,281	0,244	0,354	0,137
szövegértés – implicit	0,615	0,005	0,232	0,339	0,212	0,384	0,322	0,179	0,384	0,105

összesített pontszáma között ($r(18) = 0,467$; $p = 0,044$). A melléknevek feldolgoása egyedül a két betűből álló ($r(18) = 0,694$; $p < 0,001$) és a hosszabb szavak olvasásával ($r(18) = 0,548$; $p = 0,015$) korrelált. Szignifikáns korreláció volt megfigyelhető a határozószók feldolgoása és a megnevezési képességek

($r(18) = 0,539$; $p = 0,017$), a magánhangzók ($r(18) = 0,486$; $p = 0,035$), a két betűből álló szavak ($r(18) = 0,572$; $p = 0,010$), valamint a hosszabb szavak ($r(18) = 0,456$; $p = 0,049$) olvasása, továbbá a kép-szókép párosítása ($r(18) = 0,681$; $p < 0,001$), a mondatok szintaktikai helyességének megítélése ($r(18) = 0,661$; $p = 0,002$) és az implicit szövegszintű információk feldolgozása ($r(18) = 0,601$; $p = 0,007$) között. A névmások esetében három komponenssel mutatkozott szignifikáns együttjárás: a megnevezési képességekkel ($r(18) = 0,562$; $p = 0,012$), a két betűből álló szavak olvasásával ($r(18) = 0,552$; $p = 0,014$) és a mondatok szintaktikai helyességének megítélése feladattal ($r(18) = 0,522$; $p = 0,022$) (bővebben lásd: 4. táblázat).

5.3. *Grammatikus/agrammatikus mondatok megítélése*

A résztvevők teljesítményét mondatszínten is vizsgáltuk, ennek során a betegeknek morfoszintaktikailag helyes, illetve helytelen mondatokról kellett eldönteniük, hogy a bemutatott inger helyes vagy helytelen. Az eredmények alapján szignifikáns korreláció mutatkozott a mondatok feldolgozása és a megnevezési képességek ($r(18) = 0,618$; $p = 0,005$), a két betűből álló szavak ($r(18) = 0,499$; $p = 0,030$) olvasása, továbbá a kép-szókép párosítása ($r(18) = 0,481$; $p = 0,037$) feladatok között.

Elemeztük a szó-álszó lexikai döntési feladatban használt ingerek és a mondatok helyességének megítélése közötti korrelációkat. Vizsgáltuk, hogy van-e összefüggés a szótagszám és a szófaj, valamint a mondatszintű információk feldolgozása között. A szótagszámok tekintetében szignifikáns együttjárás mutatkozott a mondatok helyességének megítélése és a két szótagú ($r(18) = 0,487$; $p = 0,035$), valamint a négy szótagú ($r(18) = 0,668$; $p = 0,002$) szavak esetében. A szófajok tekintetében szignifikáns korreláció mutatkozott a mondatok helyességének megítélése és a főnevek ($r(18) = 0,641$; $p = 0,003$), a határozószók ($r(18) = 0,661$; $p = 0,002$), valamint a névmások ($r(18) = 0,522$; $p = 0,022$) között.

4. táblázat. A szó-álszó lexikai döntés feladat és az olvasás teszt egyes feladatai közötti együttjárások elemzése a szófajok függvényében (a statisztikailag szignifikáns együttjárásokat vastagon szedtük).

Feladatok	szófaj									
	főnév		ige		melléknév		határozó		névmás	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
spontán beszéd	0,492	0,033	0,043	0,860	0,187	0,443	0,198	0,416	-0,019	0,937
auditoros verbális megértés	-0,127	0,605	-0,134	0,584	-0,163	0,504	0,051	0,834	0,054	0,827
ismétlés	-0,214	0,379	0,032	0,898	0,146	0,550	0,099	0,688	0,159	0,515
megnevezés	0,426	0,069	0,313	0,192	0,256	0,291	0,539	0,017	0,562	0,012
magánhangzók	0,551	0,014	0,197	0,418	0,437	0,061	0,486	0,035	0,275	0,255
mássalhangzók	0,273	0,259	0,354	0,137	0,339	0,156	0,412	0,080	0,084	0,732
szótagok	0,240	0,323	0,359	0,131	0,074	0,764	0,324	0,177	0,115	0,638
két betűből álló szavak	0,623	0,004	0,522	0,022	0,694	<0,001	0,572	0,010	0,552	0,014
szavak	0,513	0,025	0,315	0,190	0,548	0,015	0,456	0,049	0,404	0,086
álszavak	0,291	0,227	0,128	0,600	0,335	0,161	0,157	0,522	0,208	0,393
kép-szókép párosítása	0,653	0,002	0,544	0,016	0,419	0,074	0,681	<0,001	0,443	0,058
grammatikus/ agrammatikus mondatok	0,641	0,003	0,406	0,084	0,243	0,317	0,661	0,002	0,522	0,022
szövegértés – összesített	0,605	0,006	0,467	0,044	0,302	0,209	0,594	0,007	0,253	0,295
szövegértés – explicit	0,368	0,121	0,382	0,108	0,211	0,385	0,440	0,059	0,087	0,724
szövegértés – implicit	0,610	0,006	0,396	0,094	0,325	0,095	0,601	0,007	0,386	0,103

6. Összegzés

A nyelvi és kognitív folyamatok szorosan összefüggő, közös neurális architektúrával rendelkező mechanizmusok, ennek megfelelően az olvasás komplex

folyamatának megfelelő működéséhez számos kognitív folyamat szükséges. A kognitív működés hálózatos elrendeződésének köszönhetően bármely feldolgozási mechanizmusban jelentkező gyengébb, vagy nem megfelelő működés a többi folyamatra is hatással is. Emiatt az afáziával összefüggő olvasási zavarok jellegzetességeinek feltárása a nyelv, az olvasás és a megismerési folyamatok átfogó feltárása révén történhet. Az olvasás során a lexikai hozzáférés központi szerepet tölt be, ez a mechanizmus a szóazonosítási folyamatok és a magasabb megértési folyamatok között helyezkedik el. Tanulmányunkban vizsgáltuk a szóhosszúsági hatás, valamint a szófaj olvasási képességekre gyakorolt hatását. Célunk az volt, hogy elemezzük a szavak eltérő tulajdonságai mentén milyen összefüggések detektálhatók a különböző nyelvi és olvasási folyamatokban. Feltételeztük, hogy a szavak hosszúsága és a szófaj alapján eltérő mintázatok jelennek meg. Mindehhez a szerzett olvasászavarok felmérésére kialakított mérőeljárásunk lexikai hozzáférést vizsgáló feladatait használtuk. A fentiekén túl további célunk volt a mérőeljárásunk gyakorlatban való használhatóságának felmérése. Elemzésünk során a feladatok belső szerkezete és a többi olvasást mérő feladat közötti összefüggések feltárását helyeztük fókuszba.

A lexikai hozzáférést három, általunk kialakított feladattal mértük: képszókép párosítása, szó-álszó lexikai hozzáférés, grammatikus/agrammatikus mondatok megítélése. Eredményeink alapján a kép-szókép párosítása feladatok esetében szófajonként részben eltérő összefüggések mutatkoztak. Ennél a feladatrésznél a főnevek és az igék feldolgozása együttjárást mutatott a graféma-fonéma konverzió és a szóolvasási folyamatokkal, a szövegértési feladatoknál viszont eltérő együttjárásokat tapasztaltunk. A főnevek az implicit, az igék feldolgozása az explicit szövegszintű információk feldolgozásával mutattak szignifikáns összefüggést.

A szó-álszó lexikai döntés feladatai a megnevezési képességekkel, néhány szóolvasási feladattal, valamint a mondat és szövegszintű információk feldolgozásával mutattak szignifikáns együttjárásokat. Vizsgáltuk a lexikai döntési feladatok során a szótagszámok és a szófajok hatását. A szótagszámok alapján lényegi eltérések nem mutatkoztak, nagyrészt minden szótag esetében a két betűből álló

szavak olvasása és a mondat szintű információk feldolgozása között mutatkozott összefüggés. A szófajok tekintetében a határozók és a főnevek feldolgozásakor mutatkozott a legtöbb együttjárás, vagyis úgy tűnik, a határozók és a főnevek feldolgozása jelzi előre leginkább a mondat és szövegértési feladatokon nyújtott teljesítményt. Vizsgáltuk a mondatok helyességének megítélése és a többi olvasást mérő feladat közötti együttjárásokat is. Szignifikáns korrelációk mutatkoztak a megnevezési képességek, a szóolvasás, valamint a lexikai hozzáférést mérő feladatok esetében.

A fentiek alapján látható, hogy hipotézisünk részben beigazolódott. Adataink alapján úgy tűnik, az eltérő hosszúságú szavak feldolgozása között nincs lényegi különbség, ezzel szemben a szófaj hatása szignifikánsnak bizonyult, ugyanis szófaj szerint eltérő összefüggéseket tapasztaltunk. Első eredményeink azt mutatják, hogy már kis mintán is szignifikáns korrelációk mutatkoznak nemcsak feladatonként, de a feladatok belső szerkezete szerint is, ezek alapján pedig elmondható, hogy olvasásteresztünk lexikai hozzáférést vizsgáló próbái úgy tűnik, érzékenyen mérik az afáziához társuló szerzett olvasászavar jellegzetességeit. A résztvevők számának növelésével terveink szerint a jövőben lehetőség nyílik az olvasástereszt pszichometriai paramétereinek részletesebb vizsgálatára is, ezzel elősegítve egy megbízható, a gyakorlati munka során is jól használható olvasástereszt létrehozását.

A Felnőtt Olvasás Teszthez tartozó alkalmazás ingyenesen használható. A tesztrel kapcsolatos bővebb információ, továbbá a teszt használatához szükséges belépési kód az alábbi e-mail címen kérhető: felnottolvasasteszt@gmail.com

Támogatás

A tanulmány elkészítését a Magyar Tudományos Akadémia Közoktatás-fejlesztési Kutatási Programja támogatta.

This study was funded by the Scientific Foundations of Education Research Program of the Hungarian Academy of Sciences.

Metakogníció Kutatócsoport, SZTE, Metacognition Research Group, SZTE

Köszönetnyilvánítás

A szerzők köszönetet mondanak Dr. Molnár-Tóth Alinkának, aki konstruktív javaslataival hozzájárult a tanulmány elkészítéséhez.

Hivatkozások

- Alvermann, D. E., Unrau, N. J., & Ruddell, R. B. (2013). Models of reading and writing processes. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical Models and Processes of Reading (6th ed)* (pp. 691–698). Newark, DE: International Reading Association.
- Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., & Zilles, K. (2010). Broca’s region: Novel organizational principles and multiple receptor mapping. *PLOS Biol*, 8, e1000489. doi:<https://doi.org/10.1371/journal.pbio.1000489>.
- Baldo, J. V., & Dronkers, N. F. (2007). Neural correlates of arithmetic and language comprehension: a common substrate? *Neuropsychologia*, 45, 229–235.
- Blumstein, S. E., & Amso, D. (2013). Dynamic functional organization of language: Insights from functional neuroimaging. *Perspectives on Psychological Science*, 8, 44–48.
- Bánréti, Z. (2014). Az afázia. In C. Pléh, & A. Lukács (Eds.), *Pszicholingvisztika 1–2: Magyar pszicholingvisztikai kézikönyv* (pp. 1167–1241). Budapest: Akadémiai Kiadó.
- Ciaccio, L. A., Burchert, F., & Semenza, C. (2020). Derivational morphology in agrammatic aphasia: A comparison between prefixed and suffixes. *Frontiers in Psychology*, 11. doi:<https://doi.org/10.3389/fpsyg.2020.01070>.

- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589–608. doi:<https://doi.org/10.1037/0033-295X.100.4.589>.
- DeDe, G. (2017). Effects of lexical variables on silent reading comprehension in individuals with aphasia: Evidence from eye tracking. *Journal of Speech, Language, and Hearing Research*, 60, 2589–2602. doi:https://doi.org/10.1044/2017_JSLHR-L-16-0045.
- Denes, G. (2011). *Talking Heads: The Neuroscience of Language*. Hove: Psychology Press.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, 27, 429–446.
- Glucksberg, S., Kreuz, R. J., & Rho, S. H. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 323–335. doi:<https://doi.org/10.1037/0278-7393.12.3.323>.
- Hagoort, P. (2005). On broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9, 416–423.
- Hallowell, B., & Chapey, R. (2008). Introduction to language intervention strategies in adult aphasia. In R. Chapey (Ed.), *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders* (pp. 3–20). Philadelphia, PA: Lippincott Williams and Wilkins.
- Honbolygó, F. (2008). Beszédészlelés, mentális lexikon, beszédmegértés, beszédprodukción. In V. Csépe, M. Győri, & A. Ragó (Eds.), *Nyelv, tudat, gondolkodás* (pp. 59–116). Budapest: Osiris Kiadó.
- Juhász, A. (2003). *Logopédiai vizsgálatok kézikönyve*. Budapest: Logopédia Kiadó KKT.

- Kaye, R. C., & Cherney, L. R. (2016). Script templates: A practical approach to script training in aphasia. *Topics in Language Disorders*, 36, 136–153. doi:<https://doi.org/10.1097/TLD.0000000000000086>.
- Kiefer, F. (2011). A ragozás. In F. Kiefer (Ed.), *A magyar nyelv* (pp. 201–203). Budapest: Akadémiai Kiadó.
- Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in broca’s area: An fmri study. *PLOS One*, 10, 1–16.
- Lindenberg, R., Fangerau, H., & Seitz, R. J. (2007). Broca’s area. *as a collective term? Brain and Language*, 102, 22–29.
- Meixner, I. (1995). *A dyslexia prevenció, reedukáció módszere*. Bárczi Gusztáv Gyógypedagógiai Tanárképző Főiskola.
- Milin, P., Smolka, E., & Feldman, B. L. (2018). Models of lexical access and morphological processing. In E. M. Fernandez, & H. S. Cairns (Eds.), *The Handbook of Psycholinguistics* (pp. 204–268). John Wiley & Sons, Inc.
- Molnár-Tóth, A. (2021). Tudom, de nem tudom kimondani! a szerzett beszéd és nyelvi zavarok értelmezéséről: néhány elméleti megfontolás a klinikai gyakorlati munka tükrében. *Logopédia*, 5, 33–49.
- Mészáros, E. (2007). Mondatfeldolgozás magyar agrammatikus afáziásoknál. Doktori disszertáció.
- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition*, 15, 225–236.
- Osmanné, S. J. (1991). Az afázia klasszifikációja és diagnosztikája ii.: Az afázia egyes kategóriáinak jellemzői. *Ideggyógyászati Szemle*, 44, 351–362.

- Papathanasiou, I., & Coppens, P. (2013). Aphasia and related neurogenic communication disorders: Basic concepts and operational definitions. In I. Papathanasiou, P. Coppens, & C. Potagas (Eds.), *Aphasia and Related Neurogenic Communication Disorders*. Burlington, MA: Jones & Barlett Learning, LLC.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18, 22–37. doi:<https://doi.org/10.1080/10888438.2013.827687>.
- Perfetti, C. A. (1999). Comprehending written language: a blueprint of the reader. In M. Collin, & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 167–197). New York: Oxford University Press.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–383. doi:<https://doi.org/10.1080/10888430701530730>.
- Racsmány, M., Lukács, A., Németh, D., & Pléh, C. (2006). A verbális munkamemória magyar nyelvű vizsgálóeljárásai. *Magyar Pszichológiai Szemle*, 60, 479–506. doi:<https://doi.org/10.1556/MPSzle.60.2005.4.3>.
- Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18, 42–55.
- Riley, E. A., & Kendall, D. L. (2013). The acquired disorders of reading. In I. Papathanasiou, & P. Coppens (Eds.), *Aphasia and Related Neurogenic Communication Disorders: Basic Concepts and Operational Definitions* (pp. 157–172). Burlington, MA: Jones & Barlett Learning, LLC.
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. L., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14, 489–537.
- Sipos, Z. (2013). *Feladatgyűjtemény a hosszú szavak olvasásának gyakorlásához felső tagozatos tanulók részére*. Budapest: Meixner Alapítvány.

- Stafura, J., & Perfetti, C. (2017). Integrating word processing with text comprehension: Theoretical frameworks and empirical examples. In K. Cain, D. Compton, & R. K. Parrila (Eds.), *Theories of Reading Development. Vol. 15: Studies in written language and literacy* (pp. 9–32). Amsterdam: John Benjamins Publishing Company.
- Sung, J. E., McNeil, M. R., Pratt, S. R., Dickey, M. W., Fassbinder, W., Szuminsky, N. J., Kim, A., & Doyle, P. J. (2011). Real-time processing in reading sentence comprehension for normal adult individuals and persons with aphasia. *Aphasiology*, *25*, 57–70. doi:<https://doi.org/10.1080/02687031003714434>.
- Szöllősi, I., Lukács, A., & Zakariás, L. (2015). A végrehajtó funkciók zavara afáziában. *Magyar Pszichológiai Szemle*, *70*, 349–369.
- Tóth, A. (2018a). *Megnevezési feladathelyzetből nyert mintázatok értelmezése kognitív nyelvészeti perspektívából*. Szeged: Szegedi Tudományegyetem, Nyelvtudományi Doktori Iskola. Doktori disszertáció.
- Tóth, A. (2018b). Az afáziák definíciójának változása az idegtudományi eredmények tükrében. *Argumentum*, *14*, 97–109.
- Webster, J., Morris, J., Howard, D., & Garraffa, M. (2018). Reading for meaning: What influences paragraph understanding in aphasia? *American Journal of Speech-Language Pathology*, *27*, 423–437. doi:https://doi.org/10.1044/2017_AJSLP-16-0213.
- Whitaker, H. A. (2007). Language disorders, aphasia. In B. J. E. (Ed.), *Encyclopedia of Gerontology. Age, ageing and the aged. Second edition* (pp. 9–14). Oxford: Elsevier.

A megakadásjelenségek és a temporális paraméterek szerepe a borderline személyiségzavar felismerésében

Felletár Fanni^{1,2}, Gosztolya Gábor³, Hoffmann Ildikó^{4,5}, Babarczy Anna^{1,2}

¹Nyelvtudományi Kutatóközpont, Kísérleti Szemantika és Pragmatika Kutatócsoport

²Budapesti Műszaki és Gazdaságtudományi Egyetem, Kognitív Tudományi Tanszék

³ELKH-SZTE Mesterséges Intelligencia Kutatócsoport

⁴Nyelvtudományi Kutatóközpont, Nyelvi Folyamatok Agyi Reprezentációi Kutatócsoport

⁵Szegedi Tudományegyetem, Pszichiátriai Klinika

Abstract

Borderline personality disorder (BPD) is characterized by a pervasive pattern of instability of identity, emotions, and interpersonal relationships, and difficulty with emotional and impulse control. Due to the various co-occurrence patterns of symptoms and the interactions between them, the BPD population is relatively heterogeneous, making it difficult for clinicians to diagnose individuals. Since an individual's state of mind may to some extent be reflected in their speech behavior, an analysis of speech patterns may make a useful contribution to efforts of identifying the disorder. Our goal is to characterize spontaneous speech of BPD individuals based on the patterns of disfluencies and temporal parameters with the forced alignment method of automatic speech recognition, and to differentiate BPD individuals ($N = 27$) from healthy controls ($N = 27$) with a logistic regression statistical model. Our results have shown, that spontaneous speech of BPD individuals can be characterized by the frequency of silent pauses, filled pauses, and disturbances of grammatical encoding (grammatical errors and blendings), and the two groups can be differentiated by these features with 0.834 AUC.

Keywords: borderline personality disorder, spontaneous speech, speech disfluencies, temporal parameters, automatic speech recognition, classification

1. Bevezetés

1.1. A borderline személyiségzavar jellemzői és felismerésének nehézségei

A borderline személyiségzavarral (*borderline personality disorder*; BPD) élő személyeket az identitás, a társas kapcsolatok, valamint az érzelem- és indulatszabályozás zavarának tartós mintázata jellemzi. A DSM-5-TR (American

Email addresses: felletar.fanni@nytud.hu (Felletár Fanni), ggabor@inf.u-szeged.hu (Gosztolya Gábor), hoffmann.ildiko@nytud.hu (Hoffmann Ildikó), babarczy.anna@nytud.hu (Babarczy Anna)

Psychiatric Association [APA], 2022) dimenzionális modellje alapján a diagnózis felállításához az egyénnek a következő 7 személyiségvonásból legalább 4-gyel kell rendelkeznie, melyek közül legalább az egyik impulzivitás, kockázatvállalás vagy ellenségesség: érzelmi labilitás ($r = 0,61$; S. Nagy et al., 2022), depresszivitás ($r = 0,54$; S. Nagy et al., 2022), szorongásosság ($r = 0,50$; S. Nagy et al., 2022), szeparációs bizonytalanság ($r = 0,34$; S. Nagy et al., 2022), impulzivitás ($r = 0,62$; S. Nagy et al., 2022), kockázatvállalás ($r = 0,27$; S. Nagy et al., 2022) és ellenségesség ($r = 0,59$; S. Nagy et al., 2022).

Az érzelmi labilitás és az impulzivitás a BPD legmarkánsabb vonásai közé tartoznak (di Giacomo et al., 2017; S. Nagy et al., 2022). Az **érzelmi labilitás** a jellemzően interperszonális inger hatására könnyen kiváltódó, rövid ideig (órákig, napokig) fennálló, aránytalanul intenzív dühöt, szorongást vagy depressziót, valamint a „baseline”-hoz való lassú visszatérést jelenti (Linehan, 1993; Koenigsberg, 2010; Unoka et al., 2018; APA, 2022). Kialakulása összefüggésbe hozható a gyermekkorban tapasztalt érzelmi elhanyagolással, korai szeparációval, fizikai vagy szexuális bántalmazással (Gunderson & Sabo, 1993; Koenigsberg, 2010; Unoka et al., 2018), melynek eredményeként a BPD-ben érintett személyek a későbbi életszakaszaikban mások ártalmatlan – ugyanakkor számukra traumatikus emlékeket előhívó (Carter & Grenyer, 2012) – cselekedeteit kritikaként, visszaütésként vagy elhagyásként értelmezik (Millon et al., 2004; Koenigsberg, 2010). Az ennek hatására kiváltódó intenzív érzelmek – az indulati kontroll hiánya miatt (Kernberg, 1975) – könnyen vezethetnek impulzív, agresszív (pl. fenyegetőzés, verekedés) és öndestruktív (pl. szerhasználat, paraszucid viselkedés) cselekedetekhez (Linehan, 1993; Millon et al., 2004; Unoka et al., 2018). Az **impulzivitás** az intenzív érzelmek által kiváltott hirtelen viselkedéses válaszreakció, amelynek késleltetésére képtelen az egyén, és a következményeivel sem számol (APA, 2022). A BPD-ben érintett személyek meghatározó része azonban elfojtja ezeket az érzelmi reakciókat, és még akkor sem mutatják ki érzelmeiket, mikor indokolt lenne (Linehan, 1993).

A körkép központi komponense az **identitászavar**, amely ürességérzés, alacsony szorongástűrés, valamint változó énkép, célok és értékrend formájában je-

lenik meg (APA, 2022). Az érzelmek és az indulatok túl- vagy alulkontrollálása felnőttkorra instabil énképet eredményezhet (Linehan, 1993). Ugyanis tekintve, hogy az alulkontrollálás meggondolatlan cselekedetekhez, a túlkontrollálás, elfojtás pedig a túlzott szorongás hatására fellépő téveszmékhez és disszociációhoz vezethet, a BPD-vel élő személyek könnyen elveszthetik a saját magukba vetett hitüket (Linehan, 1993). Emellett a BPD-ben érintett személyekre jellemző, hogy a korábbi tapasztalataiktól izoláltan, az éppen aktuális társas érintkezés függvényében értékelik magukat, alakítják világlátásukat és hoznak döntéseket a jövőjükkel illetően (Linehan, 1993).

A BPD becsült prevalenciája 1–6% közé tehető a populációban. A mérések szerint a nemi arányban nem mutatható ki különbség a teljes populációban, a klinikai (diagnosztizált) populációban viszont a nők száma meghaladja a férfiakét. Bár az egyének csak a 18. életévük betöltését követően diagnosztizálhatók, az első tünetek már pubertás korban megjelennek, a 40-es évektől pedig enyhülni kezdenek. A DSM-5-TR (APA, 2022) statisztikái szerint a kórkép leggyakrabban szorongásos zavarokkal (pánikzavar, szociális szorongásos zavar), poszttraumás stressz szindrómával, major depresszióval, bipoláris zavarral, személyiségzavarokkal (dependens, hisztrionikus, nárcisztikus, antiszociális, paranoid, skizotípiás), evészavarokkal (bulimia, túlevéses zavar), szerhasználati zavarokkal és ADHD-val társul.

A pszichiátriai gyakorlatban továbbra is elterjedtebb a DSM kategóriális modellje, amelynek értelmében az egyén BPD diagnózist kap, amennyiben 9 diagnosztikai kritériumból legalább 5 igaz rá. A tünetek számos lehetséges permutációja és a köztük létesülő lehetséges kölcsönhatások **heterogén BPD populációt** eredményeznek (Millon et al., 2004; Smits et al., 2017; Unoka et al., 2018), így a kórkép felismerése nehéz feladatot ró a klinikusokra. A pszichiátriai vizsgálat elsősorban klinikus-vezette interjúk és tesztek segítségével történik. Ezek a módszerek főként a klinikus szakértelmén és a páciens saját állapotával kapcsolatos egyéni megélésén alapulnak, amelyek az eredményes diagnózis elengedhetetlen és pótolhatatlan komponensei. A számítógépes beszédelemzés fejlődésével megnyílt a lehetőség arra, hogy a klinikusok munkáját – a standard

módszereken túl – automatikus kvantitatív módszerek is megtámogassák (Low et al., 2020; Heinz et al., 2022). Wang és munkatársainak (2020; 2021) kutatási eredményeiből például arra következtethetünk, hogy a BPD-ben érintettek spontán beszédének egy lényegesen nagyobb mintán történő számítógépes nyelvészeti elemzése a jelenlegi diagnosztikai eljárások kiegészítő elemeként a jövőben hozzájárulhat a BPD sikereesebb felismeréséhez.

1.2. A beszéd mint lehetséges biomarker

Pszichiátriai állapotfelmérésnél a páciens magatartását vizsgálják. Tringer (2010) a magatartást három típusra osztotta: 1) külső magatartásra, amely azonos a megfigyelhető viselkedéssel; 2) belső magatartásra, amely a kognitív folyamatokat, az emóciókat, a motivációkat és az attitűdöket foglalja magában; és 3) kommunikatív magatartásra, amely által az egyén a belső magatartást külsővé teszi. A kommunikáció – és ezáltal a magatartás – egyik legfontosabb megnyilvánulási formája a **beszéd**. Egy adott magatartás akkor minősül pszichiátriai tünetnek, ha a helyzethez nem illik, illetve ha az elvárthoz képest túl gyakran vagy túl ritkán jelenik meg (Tringer, 2010). Ugyanez igaz a beszédre is. Beszélhet valaki túl sokat vagy túl keveset, túl gyorsan vagy túl lassan, illetve használhat olyan szófordulatokat, amelyek az adott beszédhelyzetben szokatlannak. Amennyiben ez a verbális viselkedés meghaladja a klinikusok által normálnak tekintett mértéket, a beszéd ilyen és ehhez hasonló megnyilvánulásai – többek között – valamilyen mentális zavar tüneteiként foghatók fel.

Mint ahogy azt az 1.1. alfejezetben tárgyaljuk, az érzelmi labilitás intenzív harag, szorongás vagy depresszió formájában jelenik meg. Az érzelmeket dimenzionális modellben értelmező tanulmányok (Johnstone & Scherer, 2000; Breitenstein et al., 2001; Scherer, 2003; Laukka et al., 2005) szerint a magas arousallel járó érzelmek – így például a harag és a szorongás – hatására nő a beszédtempó és rövidülnek a szünetek, míg az alacsony arousallel járó érzelmek – így például a depresszió – az előbbiekkal ellentétes irányú változást idéznek elő a felsorolt paraméterekben (Tisljár-Szabó, 2014). Mivel a szorongás szubjektív légszomjjal jár, beszéd közben az egyén gyakran tart néma szünetet levegővétellel

céljából (Tringer, 2010). A szorongó személy beszédében mindezeket túl gyakran előfordulnak kitöltött szünetek, töltelékszavak, agrammatikus mondatok, szótévesztések és parafáziák (Kasl & Mahl, 1965; Christenfeld & Creager, 1996; Tringer, 2010; Gósy et al., 2013).

1.2.1. Megakadásjelenségek és temporális paraméterek a spontán beszédben

A megakadásjelenségek a spontán beszéd pragmatikai funkcióval nem rendelkező elemei, amelyek a beszédtervezés és -kivitelezés között létrejött diszharmónia eredményeként megtörnek a beszéd artikulációs (néma szünet) vagy tartalmi (pl. ismétlések, töltelékszavak) folyamatosságát, és a beszédtervezési folyamat pillanatnyi zavarát jelzik (Gósy, 2002, 2005; Gyarmathy, 2015, 2017; Horváth, 2017). Természetes jelenségek ugyanakkor (Allwood et al., 1992; Gilquin & De Cock, 2011), a beszédproduktions mechanizmus univerzális volta miatt minden nyelvben előfordulnak, mennyiségük (szám, hossz) és minőségük (típus, együttjárás) azonban beszélőfüggőséget mutat (McDougall & Duckworth, 2017). Így például a beszélő személyisége és pszichés állapota is hatással van ezeknek a paramétereknek az alakulására.

A megakadásjelenségeket Gósy (2002) két csoportra osztotta: a beszélő bizonytalanságából adódó megakadásokra és téves kivitelezésből adódó hibajelenségekre. A beszélő bizonytalanságából adódó jelenségek a lexikális hozzáférés és az artikulációs tervezés közötti kapcsolat zavarából adódnak. Ide soroljuk a néma és kitöltött szüneteket, a nyújtásokat, az újraindításokat, az ismétléseket, valamint a töltelékszavakat. A néma és kitöltött szünetek gyakoriságából és hosszából kinyert jellemzőket (pl. artikulációs- és beszédtempó, szünetek száma és hossza) temporális paramétereknek nevezzük, amelyek például a neurodegeneratív zavarok (Hoffmann et al., 2010; Grósz et al., 2015; Sztahó et al., 2015; Hoffmann et al., 2017; Tündik et al., 2017; Tóth et al., 2018; Gosztolya et al., 2019; Svindt et al., 2020; Gosztolya et al., 2021; Egas-López; et al., 2021), a depresszió (Kiss et al., 2017; Low et al., 2020), a bipoláris zavar (Maxhuni et al., 2016; Guidi et al., 2017; Low et al., 2020), a skizofrénia (Kliper et al., 2010; Rapcan et al., 2010; Kliper et al., 2016; Bagi et al., 2019; Low et al., 2020) és a

BPD (Wang et al., 2020, 2021) felismerésében hatékonyak bizonyultak. Ezek a jellemzők a leggyorsabban és leghatékonyabban **automatikus beszédfelismeréssel** (*automatic speech recognition, ASR*) nyerhetők ki (ld. 2.3.2. alfejezet).

Hibának tekintjük a felszíni szerkezetben realizálódó, az adott nyelv szabályainak ellentmondó jelenségeket, amelyet nem szándékosan hangosított meg egy megfelelő nyelvi kompetenciával rendelkező anyanyelvi beszélő (Gósy, 2002; Gyarmathy, 2017). A hibajelenségeket Levelt (1989) beszédproduktív modellje alapján csoportosíthatjuk tovább. Így az elszólás és a malapropizmus a fogalmi előkészítés; a grammatikai hiba és a kontamináció a nyelvi enkódolás szintjéhez; a téves szótalálás, a téves kezdés és „a nyelvem hegyén van” jelenség a mentális lexikonhoz; az anticipáció, a perszeveráció és a metatézis az artikulációs tervezés szintjéhez; az egyszerű nyelvbottlás pedig az artikulációs kivitelezéshez sorolható (Levelt, 1989; Gósy, 2002; Gyarmathy, 2017). Véleményünk szerint ez az osztályozás jó eszközként szolgálhat egy adott beszélőcsoport zavart beszédtervezési szintjének vagy szintjeinek az azonosításához.

1.3. A BPD-ben érintett személyek beszédjellemzőinek vizsgálata

A BPD-vel élő személyek spontán beszédének vizsgálata viszonylag új kutatási iránynak számít, azonban néhány, jó kiindulópontként szolgáló tanulmány már megjelent a témában. A tanulmányok túlnyomó része a diskurzusszinttel foglalkozik, ennek is két típusával: a narratív monológok koherenciaelemzésével (Adler et al., 2012; Lind et al., 2019; Bendstrup et al., 2021) és a párbeszédok szerveződésének vizsgálatával (Wang et al., 2020, 2021). A szintaxis szintjén Carter & Grenyer (2012) a mondatok hosszát és komplexitását vizsgálta. A szavak szintjén a lexikai denzitást és diverzitást vizsgálták, valamint szógyakorisági elemzéseket végeztek például a következő kategóriákban: szófajok, beszédtema (pl. család, munka, halál), pszichés folyamatok (pl. kogníció, érzelmek), „abszolút” szavak (pl. *minden, semmi*), töltelékszavak (pl. *blah, I mean, you know*) és kitöltött szünetet jelölő szavak (pl. *mm, er, um*) (Carter & Grenyer, 2012; Rosenbach & Renneberg, 2015; Lyons et al., 2018; Wang et al., 2020; Spruit et al.,

2022). A fonetikai elemzések a prozódiai (pl. alaphang, energiakontúrok), az akusztikai (pl. jitter, shimmer) és a temporális paraméterekre (néma szünetek hossza) terjedtek ki (Villanueva-Valle et al., 2021; Wang et al., 2021).

Carter & Grenyer (2012) 20 BPD-ben érintett és 20 kontrollszemély beszédét hasonlította össze többek között a töltelékszavak és a kitöltött szünetet jelölő szavak gyakorisága alapján egy a gyermekkori emlékekre fókuszáló interjú keretében. Azt feltételezték, hogy a gyermekkori emlékek felidézése a BPD csoport esetében együtt jár a kifejező beszéd valamilyen zavarával. Számos jellemzőt figyelembe véve arra a következtetésre jutottak, hogy az emlékek felidézése szignifikánsan több zavart idéz elő a BPD csoport kifejező beszédében, mint a kontrollcsoportban. Azonban a töltelékszavak és a kitöltött szünetek gyakorisága tekintetében nem találtak szignifikáns különbséget a két csoport között.

Wang és munkatársai (2020; 2021) a töltelékszavak és a kitöltött szünetek gyakoriságát, a másodpercenkénti szószámot, valamint a néma szünetek átlagos hosszát vizsgálták egyéb paraméterek mellett 17 BPD-vel élő, 21 bipoláris zavarral élő és 12 kontrollszemélynél. Céljuk egy olyan klasszifikációs modell építése volt, amely automatikusan megkülönbözteti egymástól a BPD-vel élő, bipoláris zavarral élő és kontrollszemélyeket nem-klinikai interjúk szituációjában. A BPD-ben érintett és a kontrollszemélyek osztályozásához az első modelljükben – kutatásunk szempontjából nem releváns jellemzők mellett – a kitöltött szünetet jelző hangkapcsolatok gyakorisága, a másodpercenkénti szószám és a rövid szünetek (< 500 ms) átlagos hossza bizonyult jó prediktornak ($AUC = 0,733$; Wang et al., 2020). Második modelljükben pedig többek között a rövid és a hosszú (> 500 ms) szünetek átlagos hossza bírt megfelelő bejósolóerővel ($AUC = 0,841$; Wang et al., 2021), amely az egyedüli valódi temporális paraméter volt a felso-roltak közül.

1.4. A kutatás célja

Kutatásunk kiindulását a fentebb ismertetett nemzetközi és hazai kutatások eredményei adták, valamint olyan korábbi vizsgálatok, ahol a beszéd idői paramétereinek egy része biztonsággal szét tudta választani a klinikai és a kont-

rollcsoportot. A **beszédtempót** és a **néma szünetek hosszát** a harag, szorongás és depresszió során kimutatott normáltól eltérő értéke miatt (Johnstone & Scherer, 2000; Breitenstein et al., 2001; Scherer, 2003; Laukka et al., 2005; Tisljár-Szabó, 2014), a **néma szünetek gyakoriságát** a szorongásra jellemző gyakori lélegzetvétel miatt (Tringer, 2010), a **néma szünetek hosszát** és a **kitöltött szünetek gyakoriságát** Wang és munkatársai (2020; 2021) kutatásaiban bizonyított predikciós hatása miatt, valamint a **kitöltött szünetek**, a **verbális bizonytalansági megakadások**, a **nyelvi enkódolási hibák**, a **lexikális hozzáférés zavarából adódó hibák** és az **artikulációs kivitelezési hibák gyakoriságát** a szorongással való együttjárásuk miatt (Kasl & Mahl, 1965; Tringer, 2010; Gósy et al., 2013) választottuk vizsgálatunk középpontjává.

Jelen kutatásunk célja 1) a BPD csoport spontán beszédének leírása a megakadásjelenségek és a temporális paraméterek mérésén keresztül az automatikus beszédfelismerés kényszerített illesztés módszerének alkalmazásával; 2) a BPD és a kontrollcsoport automatikus elkülönítése a kinyert jellemzők alapján. Feltételezzük, hogy 1) a BPD csoport spontán beszéde jellemezhető a beszédtempó, a néma szünetek gyakorisága és hossza, a kitöltött szünetek gyakorisága, valamint a verbális bizonytalansági megakadások (ismétlések, újraindítások, töltelékszavak), a nyelvi enkódolási hibák (grammatikai hiba, kontamináció), a lexikális hozzáférés zavarából adódó hibák (téves szótalálás, téves kezdés, „a nyelvem hegyén van” jelenség) és az artikulációs kivitelezési hibák (egyszerű nyelvbotlások) gyakorisága alapján; 2) a BPD és a kontrollcsoport jól elkülöníthető ezen jellemzők mentén; 3) mivel a BPD-nek nem centrális tünete a beszédprodukciós nehézség, csak finom, műszeresen kimutatható eltérések lesznek megfigyelhetők a két csoport között.

A megakadásjelenségek típusait magasabb szintű kategóriákba soroltuk Gósy (2002) csoportosítása, valamint Levelt (1989) beszédprodukciós modelljének tervezési szintjei alapján (ld. 2. táblázat) . Kivételt képeztek ez alól a néma szünetek, a kitöltött szünetek és a nyújtások, amelyekből a temporális paramétereket számítottuk ki. Az elemzés első lépéseként egy beszédfelismerő rendszerre támaszkodva a megakadásjelenségek gyakoriságát és a temporális paramétereket

számítottuk ki, a másodikban pedig egy logisztikus regressziós modellt építettünk, amely a kinyert jellemzők alapján megkísérelte elkülöníteni a BPD csoportot a kontrollcsoporttól. A modell kiértékelésekor szenzitivitás és az AUC érték maximalizálására törekedtünk. Hosszú távú célunk egy valós klinikai hasznot nyújtó, nagyobb mintán betanított gépi osztályozó modell megépítése, amelyhez a jelenlegi és az ezt követő statisztikai elemzéseink eredményei szolgáltatnák a kiindulópontot.

2. Módszer

2.1. Résztvevők

Az adatgyűjtéshez BPD-vel élő és kontrollszemélyeket toboroztunk. Mivel a COVID-19 pandémia következtében kialakult helyzet miatt nem nyílt lehetőségünk közvetlenül felvenni a kapcsolatot pszichiátriai ellátásban részesülő BPD-ben érintett személyekkel, mindkét csoport esetében a **kényelmi és hólabda mintavételt** alkalmaztuk. Kényelmi mintavétel során a kutató az éppen elérhető intézményekből, csoportokból toboroz résztvevőket (Szokolszky, 2004). Mindkét csoport meghatározó részét a közösségi médiában jelen lévő BPD-vel élő személyeknek létrehozott és egyéb (kérdőívkitöltő, szakdolgozatíró, kísérleteket hirdető stb.) csoportokból toboroztuk. A hólabda mintavétel során a kísérleti személyeken keresztül léptünk kapcsolatba további potenciális résztvevőkkel (Szokolszky, 2004). A kutatásban már részt vett személyeket arra kértük, hogy osszák meg a részvételi lehetőséget ismerőseikkel.

Mindkét csoportnál alapkritérium volt, hogy a résztvevők magyar anyanyelvűek legyenek, illetve ép hallással és beszédszervekkel rendelkezzenek. A BPD-ben érintett személyek esetében a kutatásban való részvétel feltétele a BPD diagnózisát igazoló klinikai dokumentum bemutatása volt, amely tartalmazza az **F6030 – Érzelmileg labilis személyiségzavar** BNO-kódot (BNO-11, 2022). A komorbid zavarok nem számítottak kizáró oknak, mivel a populációban is gyakran előfordulnak. A kontrollszemélyként való részvétel feltétele az volt, hogy az egyének ne rendelkezzenek az alábbi zavarok egyikével sem: nyelvi

vagy beszédzavar (afázia, dadogás, hadarás, vokális tic, artikulációs zavar stb.), agyi vagy idegrendszeri betegség, sérülés (epilepszia, demencia, Parkinson-kór, agyi sérülés, agydaganat stb.), fejlődési zavar (ADHD, autizmus spektrum zavar stb.), pszichoaktív szerhasználat (gyógyszer-, alkohol-, drogfüggőség), mentális zavar (személyiségzavar, hangulatzavar, szorongásos zavar, skizofrénia stb.). A kontrollszemélyek önbevallás alapján adtak számot az egészségi állapotukról.

A toborzás eredményeként végül 27 BPD-vel élő és 27 kontrollszemély vett részt a kísérletben. A két csoport illesztésénél a résztvevők nemét, életkorát és sikeresen elvégzett tanéveinek számát vettük figyelembe, amelyek adatait az 1. táblázat foglalja össze.

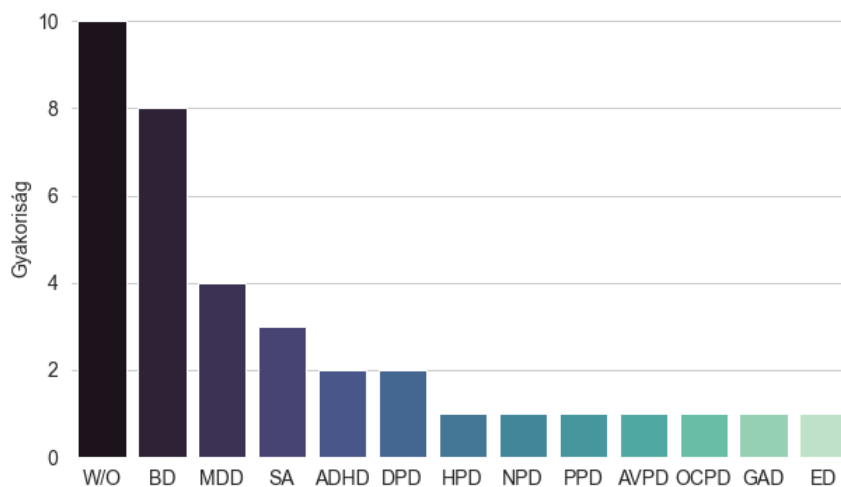
1. táblázat. A résztvevők demográfiai adatai.

N = elemszám; M = átlag; R = terjedelem.

Demográfiai jellemzők	BPD	Kontroll
N	27	27
nemi arány; nő:férfi	23:4	23:4
életkor; $M(R)$	27,26(20–50)	26,59(19–45)
elvégzett tanévek száma; $M(R)$	14,85(11–20)	15,52(11–21)

A táblázat alapján láthatjuk, hogy a két csoport adatai kiegyensúlyozottak. Az elemszám ($N = 27$) és nemi arány (23:4) teljesen megegyezik, az átlagéletkor és az elvégzett tanévek átlagos száma pedig közel azonos. A BPD csoportban – a klinikai populációhoz hasonlóan (APA, 2022) – a nők felülreprezentáltak. A BPD és a kontrollcsoport életkorának és elvégzett tanévei számának összevetésére Mann–Whitney-próbát alkalmaztunk, amely azt mutatta, hogy sem az életkor, sem az iskolázottság tekintetében nincs szignifikáns eltérés a két csoport között. A hasonló demográfiai adatok lehetővé teszik a valós különbségek megállapítását a beszédjellemzők vizsgálatakor. Az általános demográfiai adatok mellett számon tartottuk a BPD csoportban előforduló komorbid zavarok típusát és gyakoriságát, amelyet az 1. ábra szemléltet.

A mintán előforduló komorbid zavarok típusai jól tükrözik a BPD-vel leggyakrabban együttjáró komorbid zavarokat (ld. 1.1. alfejezet). A 27 BPD-ben



1. ábra. Komorbid zavarok típusa és gyakorisága.

(Az összlétszám nem egyezik meg a BPD csoport létszámával, mert egy személy több komorbid zavarral is rendelkezhet.)

W/O = nincs komorbid zavar; BD = bipoláris zavar; MDD = major depresszió; SA = szerabúzus; ADHD = figyelemhiányos hiperaktivitás zavar; DPD = dependens személyiségzavar; HPD = hisztrionikus személyiségzavar; NPD = nárcisztikus személyiségzavar; PPD = paranoid személyiségzavar; AVPD = elkerülő személyiségzavar; OCPD = kényszeres személyiségzavar; GAD = generalizált szorongásos zavar; ED = evészavar.

érintett személyből 17 (62,96%) rendelkezett legalább egy komorbid zavarral. Mivel célunk a BPD populáció természetes reprezentációja volt, nem tekintettük kizáró oknak a komorbid zavarok jelenlétét, és elemzéseinkben nem kíséreltük meg elválasztani egymástól a BPD-t egyéb pszichiátriai zavaroktól.

2.2. Adatgyűjtés

Az adatgyűjtéshez a **narratív interjú** módszerét választottuk. A narratív interjú egy személyes életeseményekre fókuszáló, kvalitatív, strukturálatlan interjúzási stratégia, melyben az adatrögzítő határozza meg a narratíva témáját, majd hagyja az adatközlőt minél szabadabban kibontakozni. Mivel fontos, hogy az adatközlő narratívája egy koherens egységet alkosson, az adatrögzítő az adatközlő beszédét nem szakítja félbe (Szokolszky, 2004). Azért esett a válasz-

tásunk erre a módszerre, mert 1) a pszichiátriai vizsgálatok során jellemzően strukturálatlan interjút alkalmaznak (Tringer, 2010; Bruchmüller et al., 2011; Cwik et al., 2016), 2) a BPD-vel élő személyek a kontrollszemélyekhez hasonlóan teljesítenek strukturált körülmények között (Kernberg, 1975; Edell, 1987; Millon et al., 2004), 3) így megfelelő közlésmennyiséget kapunk a megakadásjelenségek és a temporális paraméterek méréséhez.

Célunk az volt, hogy a vizsgálati személyeket **irányított spontán beszédre** készítsük. Kutatásunkban az adatközlőket arra kértük, hogy meséljék el a tegnapi napjukat: *Kérlek, meséld el részletesen a tegnapi napodat a felébredésedtől kezdve! Hogyan telt a napod? Mit csináltál? Mi történt veled? Hogy éreztet magad? A spontán beszéd kezdetétől fogva az adatrögzítő a hallgató szerepét vette fel; a narratívát nem szakította meg kérdésekkel, kommentárokkal akkor sem, ha az adatközlő elakadt. Az adatrögzítő csak nonverbális jelekkel (bólogatás, mosoly) támogatta a narratíva előrehaladását. Nem szabtuk meg minimum és maximum időkeretet, ezzel is elősegítve az adatközlő minél spontánabb megnyilvánulását. Az adatközlők beszédéről azonos körülmények között, hangcsillapított szobában, ugyanazon adatrögzítő közreműködésével készült hangfelvétel. A résztvevőket a kutatás menetéről tájékoztattuk, a beszéd felvételhez és annak tudományos célú felhasználásához írásos hozzájárulásukat adták.*

2.3. Az adatelemzés folyamata

2.3.1. Adatok előfeldolgozása

Az adatgyűjtés lezárultával a hangfelvételeket megvágtuk és felhangosítottuk az *Audacity* program segítségével (Audacity Team, 2021). A hangfelvételeket ezután lefuttattuk az Alrite automatikus beszédfelismerő programon (Régens, 2021), amelynek eredményeként szöveges átiratokat kaptunk. Az átiratok tartalmi és nyelvi helyességét manuálisan ellenőriztük, szükség esetén a hibákat kijavítottuk. Ezt követően annotáltuk a szövegben a 2. táblázatban felsorolt megakadásjelenségeket, amelyek meghatározását illetően Gósy (2002, 2005) és Gyarmathy (2017) munkájára támaszkodtunk.

2. táblázat. A vizsgált megakadásjelenségek csoportosítása.

Megakadás-kategóriák	Megakadásjelenségek
Temporális paraméterként kezelt jelenségek (ld. 3. táblázat)	néma szünetek
	kitöltött szünetek
	nyújtások
Verbális bizonytalansági megakadások	ismétlés
	újraindítás
	töltelékszavak
A nyelvi enkódolás hibái	grammatikai hibák
	kontamináció
A lexikális hozzáférés hibái	téves szótalálás
	téves szókezdés
	„a nyelvem hegyén van” jelenség
Artikulációs hibák	egyszerű nyelvbtlás

2.3.2. A jellemzők kinyerése

Az eljárás alapjául a hangfelvételek gépi, szószintű átirata szolgált, melyben manuálisan voltak annotálva a megakadásjelenségek. A folyamatot két szakaszra bontottuk: a megakadásjelenségek és a temporális paraméterek kinyerésére. Az első szakaszban a megakadásjelenségek gyakoriságát számítottuk ki gépi módszerrel, majd arányosítottuk az értékeket a szószám alapján. Ezt követően a megakadásjelenségeket összeadás útján magasabb szintű kategóriákba soroltuk. Kivételt képeztek ez alól a néma szünetek, a kitöltött szünetek és a nyújtások, amelyeket temporális paraméterek közé soroltunk be.

A második szakaszban a temporális beszédjellemzőket számítottuk ki egy automatikus beszédfelismerő rendszer segítségével. Az annotált megakadásjelenségek közül a néma és kitöltött szüneteket, valamint a nyújtásokat használtuk fel a jellemzők kinyeréséhez. A nyújtásokat – a nemzetközi irányelveket követve (Giannini, 2003; Peters, 2003; Gyarmathy, 2017) – kitöltött szünetként kezeltük, mivel a két kategória mind funkcióját, mind felszíni megjelenését tekintve hasonlóságokat mutat. Az írásjelek eltávolítása után a szavakat (az esetlegesen

jelölt nyújtásokat is figyelembe véve) egy automatikus szabályalapú eljárással beszédhang-sorozattá konvertáltuk. Ezután a beszédhang-sorozatba a szószintű átirat alapján néma és kitöltött szüneteket illesztettünk, így az egész felvételhez egy fonetikai címsorozatot kaptunk.

A beszédfelismerő rendszerünk segítségével ezután végrehajtottunk egy **kényszerített illesztés** eljárást. Ebben az eljárásban nem klasszikus beszédfelismerést végeztünk, ahol egyszerre próbáljuk meghatározni a hangfelvételhez legjobban illeszkedő fonetikaicímke-sorozatot és annak legvalószínűbb időbeli illesztését is. A kényszerített illesztés módszerrel a fonetikai címkéket (és azok sorrendjét) adottnak véve csupán azok legvalószínűbb pozícióját (kezdő és záró időpontjait) kerestük meg. Az eredmény így jóval kevesebb hibát tartalmaz, mintha – a szószintű átirat hiányában – a fonetikai címkéket is automatikusan határoznánk meg. Ennek a lépésnek az eredményeként a hangfelvételre a kézi átirat egy időbeli illesztését kaptuk, melyben minden szó minden beszédhangjának, valamint az összes (a szószintű átiratban jelölt) néma és kitöltött szünetnek adott a kezdő és záró időpontja. A kinyert temporális paramétereket a 3. táblázat foglalja össze.

3. táblázat. Vizsgált temporális paraméterek.

Tóth et al. (2015, 2018); Hoffmann et al. (2017) és Bagi et al. (2019) alapján.

Temporális paraméterek	Meghatározás
Artikulációs tempó	a kitöltött szünetek nélkül számított másodpercenkénti beszédhang-szám
Beszédtempó	az összes beszédhang-szám osztva a teljes beszédidővel
Néma szünetek számának aránya	a néma szünetek száma osztva az összes beszédhang-számmal
Néma szünetek gyakorisága	a néma szünetek előfordulásának száma osztva a teljes beszédidővel
Néma szünetek hosszának aránya	a néma szünetek összhossza osztva a teljes beszédidővel
Néma szünetek átlagos hossza	a néma szünetek összhossza osztva a néma szünetek számával
Kitöltött szünetek számának aránya	a kitöltött szünetek száma osztva az összes beszédhang-számmal
Kitöltött szünetek gyakorisága	a kitöltött szünetek előfordulásának száma osztva a teljes beszédidővel

Az egymáshoz nagyon hasonló jelenségeket mérő temporális paramétereket párokba rendeztük a statisztikai hatásuk későbbi összehasonlítása céljából. A beszédidőt és a szünetek hosszát másodpercben mértük. Az illesztés kimeneteiből a temporális paraméterek kiszámításához szükséges minden érték (pl. a beszédhangok vagy a szünetek száma, a beszédidő, a néma szünetek összhossza) könnyen és automatikusan meghatározható. (A beszédidő kiszámításakor a felvétel esetleges kezdő és záró néma szünetét figyelmen kívül hagytuk.)

Beszédfelismerő rendszerünk a *HTK* (*Hidden Markov Toolkit*; Young et al., 2006) volt, melyben egy mély neurális háló (*Deep Neural Network, DNN*; Hinton et al., 2012) akusztikus modellt használtunk. Az akusztikus modellt a BEA adatbázis (Gósy et al., 2012; Neuberger et al., 2014) hatvan órányi részhalmazán tanítottuk be; keretszintű jellemzőkként 40 mel szűrősor energiáit, illetve azok első- és másodrendű deriváltjait használtuk, kimenetként pedig 911 kapcsolt fonetikai állapotunk volt. (A felismerő rendszer további technikai részleteiről bővebben ld. Gosztolya et al., 2021.)

3. Eredmények

Az adatok előfeldolgozását követően leíróstatisztikai elemzést, majd hipotézisvizsgálatot végeztünk a *JASP* statisztikai szoftver segítségével (JASP Team, 2022).

3.1. Leíróstatisztikai elemzés

Első lépésként a hipotézisben foglalt változók átlagát, szórását és a két csoport közötti különbség mértékét (az r^2 determinációs együtthatót) vizsgáltuk meg. Az elemzés eredményét a 4. táblázat foglalja össze, ahol a temporális paraméter párokat vastag vonal választja el egymástól. Az utolsó négy sor a megakadásjelenségekre vonatkozik, melyeket nem volt szükséges párokba rendezni.

A táblázat alapján azt láthatjuk, hogy nem fedezhető fel kiemelkedő különbség a két csoport között. Az összes változó közül a **nyelvi enkódolási**

4. táblázat. A változók átlaga és szórása a BPD és a kontrollcsoportban, valamint a két csoport közötti különbség standardizált mértéke.

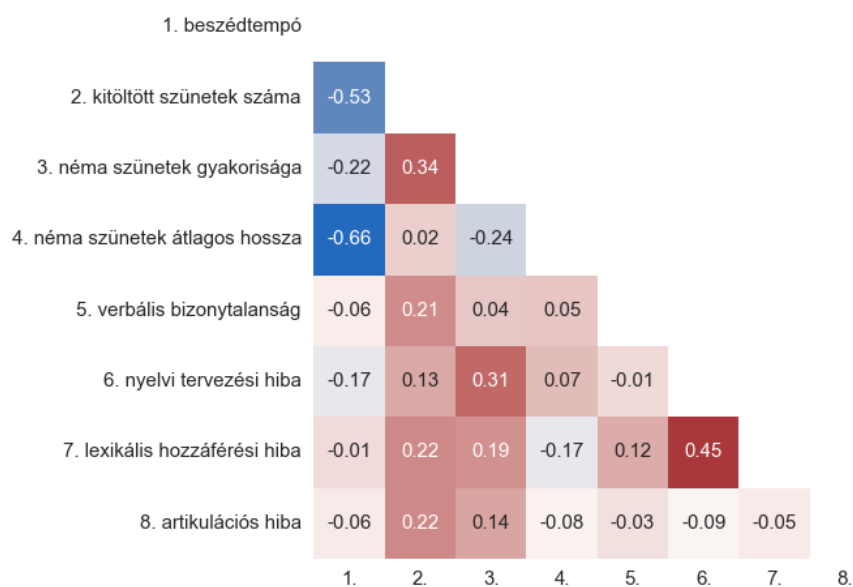
M = átlag; SD = szórás; r^2 = determinációs együttható; félkövér értékek = a magasabb determinációs együtthatóval rendelkező temporális paraméter-párrészek.

Változók	BPD M(SD)	Kontroll M(SD)	r^2
Artikulációs tempó (beszédhang/s)	9,994(2,145)	11,020(1,845)	0,064
Beszédtempó (beszédhang/s)	10,630(2,131)	11,647(1,805)	0,065
Néma szünetek számának aránya	0,041(0,011)	0,038(0,009)	0,011
Néma szünetek gyakorisága	0,413(0,070)	0,434(0,063)	0,025
Néma szünetek hosszának aránya	0,285(0,103)	0,259(0,081)	0,020
Néma szünetek átlagos hossza	0,698(0,243)	0,601(0,186)	0,049
Kitöltött szünetek számának aránya	0,022(0,008)	0,017(0,008)	0,079
Kitöltött szünetek gyakorisága	0,224(0,078)	0,193(0,071)	0,042
Verbális bizonytalanság aránya	0,085(0,053)	0,057(0,040)	0,085
Nyelvi enkódolási hibák aránya	0,008(0,012)	0,002(0,003)	0,087
Lexikális hozzáférési hiba aránya	0,007(0,008)	0,006(0,005)	0,002
Artikulációs hiba aránya	0,002(0,004)	0,002(0,005)	0,000

hibák aránya mutatta a legnagyobb különbséget a BPD és a kontrollcsoport között ($r^2 = 0,087$). A változók túlnyomó részében a BPD csoportnál magasabb átlagok és nagyobb szórások figyelhetők meg a kontrollcsoportéhoz képest. Az előbbieket alól kivételt képez a néma szünetek gyakorisága, valamint az artikulációs és beszédtempó, amelyek esetében azonban a kontrollcsoport mutatott átlagosan magasabb értékeket.

A párokba rendezett temporális paraméterek közül azokat tartottuk meg, amelyek markánsabb különbséget mutatnak a két csoport között az r^2 determinációs együttható szerint (ld. 4. táblázat; félkövérrel szedett determinációs együtthatók). A változó-párok összehasonlítása eredményeként tehát a következő változók maradtak meg: beszédtempó, néma szünetek gyakorisága, néma szünetek átlagos hossza, kitöltött szünetek számának aránya, verbális bizonytalansági megakadások aránya, a nyelvi enkódolás zavarából adódó hibák aránya,

a lexikális hozzáférés zavarából adódó hibák aránya és az artikulációs kivitelezési hibák aránya. A változók izolált elemzése azonban nem tárja fel az esetleges átfedéseket, együttes hatásokat, ezért egy predikciós modellben együtt elemeztük őket. A hipotézistesztesztelésre használt logisztikus regressziós modell előkészítéseként, a kollinearitás minimalizálása érdekében Spearman-korrelációelemzéssel megvizsgáltuk, mennyire korrelálnak a megmaradt változók egymással (ld. 2. ábra).



2. ábra. A változók közötti összefüggés Spearman-korrelációs együtthatói.

A változók kiszűréséhez a korrelációs erősség kritikus szintjét $|0,40|$ -nél húztuk meg. Az egymással jelentős átfedésben lévő változók redukálása után hat faktor maradt, amellyel a spontán beszéd általános tulajdonságait jellemezhetjük: a **néma szünetek gyakorisága**, a **néma szünetek átlagos hossza**, valamint a **kitöltött szünetek számának**, a **verbális bizonytalansági megakadások**, a **nyelvi enkódolási hibák** és az **artikulációs kivitelezési hibák aránya**. E hat változóval a logisztikus regressziós modell valamennyi előfeltétele teljesült. A multi-kollinearitás mértékét jellemző tolerancia és VIF (*Variance*

Inflation Factor) értékeket az 5. táblázat mutatja, az értékek mind a hat változó esetében elfogadhatók.

5. táblázat. A multi-kollinearitás vizsgálat eredménye.

Változók	Tolerancia	VIF
Néma szünetek gyakorisága	0,691	1,447
Néma szünetek átlagos hossza	0,844	1,185
Kitöltött szünetek számának aránya	0,660	1,516
Verbális bizonytalanság aránya	0,888	1,127
Nyelvi enkódolási hibák aránya	0,719	1,391
Artikulációs kivitelezési hiba aránya	0,890	1,124

3.2. Hipotézistesztelés

A predikciós modell felépítéséhez a logisztikus regresszió úgynevezett forward módszerét alkalmaztuk. A bináris eredményváltozó a csoport volt (BPD vagy kontroll), a bejósoló változók pedig az összefüggésvizsgálat eredményeként kapott hat faktor. A modell felépítése 4 körben történt: az első kör a referenciamodel volt, ami csak az interceptet tartalmazta. A második modellben a **nyelvi enkódolási hibák aránya** szerepelt, ami szignifikánsan növelte a modell bejósoló erejét. A harmadik modellbe a **kitöltött szünetek számának aránya**, majd az utolsó, negyedikbe a **néma szünetek gyakorisága** került be az előzőek mellé, amely szignifikánsan javította a modell bejósoló erejét. A néma szünetek átlagos hossza, a verbális bizonytalansági megakadások aránya és az artikulációs kivitelezési hibák aránya nem mutatkozott szignifikáns tényezőnek a két csoport megkülönböztetésében. A négy modell főbb mutatóit a 6. táblázat foglalja össze.

A táblázat alapján az látható, hogy a negyedik modell bejósoló ereje bizonyult a legpontosabbnak ($BIC_4 = 70,030$) az összes modell közül. Az R^2 szerint a nyelvi enkódolás zavarából adódó hibák aránya, a kitöltött szünetek számának aránya, valamint a néma szünetek gyakorisága a variancia közel 43%-át ma-

6. táblázat. A regressziós modellek főbb mutatói.

BIC = Bayesi információs kritérium; R^2 = Nagelkerke-féle pszeudo determinációs együttható; $\Delta\chi^2$ = a χ^2 -teszt eredménye; df = szabadságfok; p = szignifikanciaérték.

	Modell	BIC	R^2	$\Delta\chi^2$	df	p
1.	Csak intercept	78,849			53	
2.	Nyelvi enkódolási hibák aránya	74,971	0,181	7,867	52	0,005
3.	Nyelvi enkódolási hibák aránya Kitöltött szünetek számának aránya	74,635	0,269	4,326	51	0,038
4.	Nyelvi enkódolási hibák aránya Kitöltött szünetek számának aránya Néma szünetek gyakorisága	70,030	0,426	8,594	50	0,003

gyarazza. Az alábbi 7. táblázatban láthatók a végső modell változóinak főbb teljesítményértékei.

7. táblázat. A modell szignifikáns változóinak teljesítményértékei.

β = standardizált együttható; SE = a standardizált együttható standard hibája; z = a z -teszt eredménye; p = szignifikanciaérték.

	Változók	β	SE	z	p
1.	Nyelvi enkódolási hibák aránya	2,330	0,872	2,673	0,008
2.	Kitöltött szünetek számának aránya	1,073	0,399	2,692	0,007
3.	Néma szünetek gyakorisága	-1,060	0,400	-2,649	0,008

A végső modellben megmaradt változók együtthatói is erősen szignifikánsak ($p_1 = 0,008$; $p_2 = 0,007$; $p_3 = 0,008$). A nyelvi enkódolási hibák aránya a legerősebb magyarázó változó ($\beta_1 = 2,330$), ugyanis ha 1 szórással nő a nyelvi enkódolás zavarából adódó hibák aránya, akkor 2,33 szórással – azaz több mint a duplájával – nő annak a valószínűsége, hogy a személy a BPD csoportba tartozik. A másik két prediktor ($\beta_2 = 1,073$; $\beta_3 = -1,060$) valamivel kisebb, egymáshoz hasonló mértékben jósolja be a BPD csoporthoz való tartozás valószínűségét. A kitöltött szünetek arányának esetében egy szórásnyi növekedés 1,07 szórással növeli a BPD valószínűségét, a néma szünetek gyakoriságát

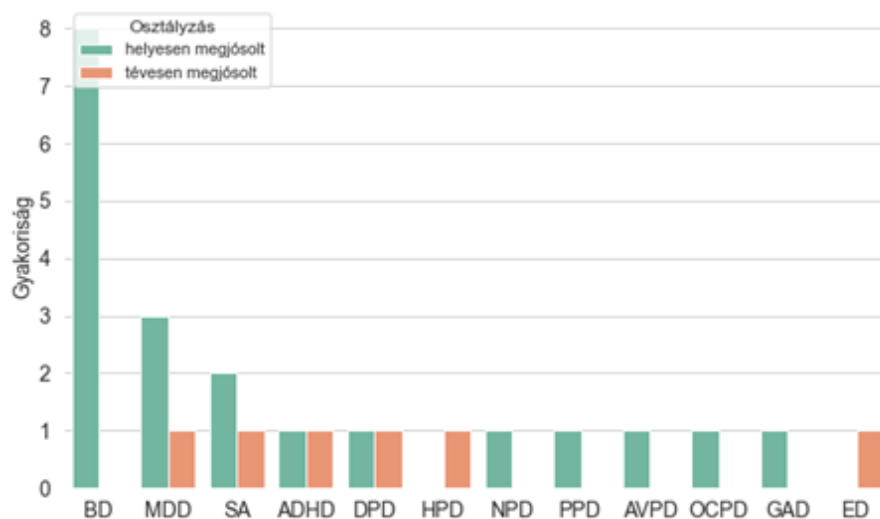
tekintve pedig hasonló mértékű, bár ellentétes előjelű összefüggést találtunk. Az adatokra illesztett modell validálatlan AUC-értéke 0,834 lett. Ha 50%-os valószínűsénél húzzuk meg a kísérleti csoporthoz tartozás határát, a modell szenzitivitása 70,37%, vagyis a 27 BPD-ben érintett személyből 19 személyt sorol be megfelelően, specifitása pedig 66,67%, tehát a 27 kontrollszemélyből 18 személyt kategorizál helyesen, míg 8 (29,63%) BPD-vel élő és 9 (33,33%) kontrollszemélyt diagnosztizál tévesen.

Mivel a komorbiditás befolyásolhatja az osztályozás kimenetelét, összehasonlítottuk a kategorizáció alakulását a komorbid zavarral rendelkező és nem rendelkező BPD-ben érintett személyeknél (8. táblázat).

8. táblázat. A helyes kategorizáció száma (százaléka) a komorbid zavarral rendelkező, illetve nem rendelkező BPD személyeknél.

	Van komorbid zavar	Nincs komorbid zavar
	N(%)	N(%)
Összesen	17(100%)	10(100%)
Helyesen kategorizált személyek	13(76%)	6(60%)
Tévesen kategorizált személyek	4(24%)	4(40%)

Bár jelentősnek tűnik a különbség a táblázat alapján, a helyes kategorizációk arányát tekintve nem mutatható ki szignifikáns különbség a komorbid zavarral rendelkező és a komorbid zavarral nem rendelkező személyek között ($\chi^2 = 0,819$; $p = 0,365$), így valószínűsíthető, hogy a komorbiditás mint átfogó kategória nem befolyásolta jelentősen az osztályozás eredményét. A különböző komorbid zavarokkal rendelkező BPD-vel élő személyek kategorizálásának helyességét a 3. ábra szemlélteti. A BPD-nek jószolt kontrollszemélyek (*false positive*; $N(\%) = 9(33,33\%)$) esetében egyik demográfiai változó sem volt hatással az osztályozásra, a kontrollnak jószolt BPD-ben érintett személyek esetében azonban az átlagéletkor (*false negative*; $M(SD) = 24,00(3,34)$) 3 évvel alacsonyabb volt, mint a teljes BPD csoport átlagéletkora ($M(SD) = 27,26(7,18)$).



3. ábra. A helyesen és a tévesen megjósolt komorbid zavarok gyakorisága.

BD = bipoláris zavar; MDD = major depresszió; SA = szerabúzus; ADHD = figyelemhiányos hiperaktivitás zavar; DPD = dependens személyiségzavar; HPD = hisztrionikus személyiségzavar; NPD = nárcisztikus személyiségzavar; PPD = paranoid személyiségzavar; AVPD = elkerülő személyiségzavar; OCPD = kényszeres személyiségzavar; GAD = generalizált szorongásos zavar; ED = evészavar.

4. Diszkusszió és összegzés

Tudomásunk szerint kutatásunk az első olyan vizsgálat, amely részletesen foglalkozik a BPD-ben érintettek spontán beszédében megjelenő megakadásjelenségekkel és a temporális paraméterekkel. Néhány korábbi kutatás is vizsgálta ezeket a paramétereket BPD-ben érintett személyeknél (Carter & Grenyer, 2012; Wang et al., 2020, 2021), ezek a vizsgálatok azonban csupán néhány jelenségre terjedtek ki (beszédtempó, néma szünetek, kitöltött szünetek, töltelékszavak). Első hipotézisünk szerint a BPD csoport spontán beszéde jellemezhető a beszédtempó, a néma szünetek gyakorisága és hossza, a kitöltött szünetek gyakorisága, valamint a verbális bizonytalansági megakadások, a nyelvi enkódolási hibák, a lexikális hozzáférés zavarából adódó hibák és az artikulációs kivitelezési hibák gyakorisága alapján. A leíróstatisztikai elemzés eredményei azt mutatták, hogy

a BPD csoport spontán beszédét átlagosan lassabb beszédtempó, hosszabb, de kevesebb néma szünet, több kitöltött szünet, valamint több verbális bizonytalansági megakadás, nyelvi enkódolási hiba, lexikális hozzáférési hiba és artikulációs kivitelezési hiba jellemzi, mint a kontrollcsoportét.

Feltételeztük, hogy a BPD és kontrollcsoport jól elkülöníthető a beszédtempó, a néma szünetek gyakorisága és hossza, a kitöltött szünetek gyakorisága, valamint a verbális bizonytalansági megakadások, a nyelvi enkódolási hibák, a lexikális hozzáférési hibák és az artikulációs kivitelezési hibák gyakorisága alapján. A logisztikus regressziós statisztikai modellünk a felsorolt jellemzők közül a nyelvi enkódolás zavarából adódó hibák arányát, a kitöltött szünetek számának arányát, valamint a néma szünetek gyakoriságát választotta ki prediktoroknak, amelyek együttesen 70,37%-os szenzitivitással jósolták meg a BPD csoportot. A szenzitivitás maximalizálására való törekvés azonban a *false positive* arány növekedését is magával vonta. A két csoportot 0,834 AUC pontossággal különítette el a modell. Eredményeink konzisztensek Wang és munkatársai (2020; 2021) modelljeivel. Első modelljük teljesítménye – amelyben többek között a beszédtempó, a kitöltött szünetek gyakorisága és a rövid szünetek átlagos hossza szerepelt – 0,733 AUC (Wang et al., 2020), második modelljüké – amelybe többek között a rövid és a hosszú szünetek átlagos hossza lett kiválasztva – pedig 0,841 AUC volt (Wang et al., 2021).

Feltételeztük továbbá, hogy csak finom, az ASR módszert felhasználva műszereken kimutatható eltérések lesznek megfigyelhetők a két csoport között. A leíróstatisztikai elemzés eredményei ugyan differenciát mutattak a BPD és a kontrollcsoport között a vizsgált változók tekintetében, ugyanakkor ezek a különbségek nem voltak jelentősek. Hogy méréseinket pontosíthassuk, a jövőben szükség lesz nagyobb beszélői minta kiértékelésére.

A statisztikai elemzés eredményeinek pusztá kiértékelésén túl érdemes ki térnünk az osztályozáshoz felhasznált prediktorok szignifikáns bejósoló erejének lehetséges magyarázataira a BPD populációban. A **nyelvi enkódolási hibák** BPD populációban becsült magasabb aránya magyarázható a BPD-ben érintett személyek **érzelmi labilitás**ával és **impulzivitás**ával összefüggő **elég-**

telen önmonitorozási képességgel. Azt feltételezzük, hogy a BPD-vel élő személyeknél ún. „áteresztő monitor” működik, amely révén a beszédtervezés folyamán keletkező hiba megjelenik a felszínen (Levelt, 1989; Gyarmathy, 2017), ugyanúgy ahogy egy-egy, az adott helyzethez képest inadekvát vagy aránytalan érzelem vagy indulat is a felszínre törhet. Mindkét típusú nehézséget tekinthetjük a gátlás funkcióhoz kapcsolódó problémának, vagyis az önmonitorozás egyik feltétele a végrehajtó funkciók hatékony működése a beszédtervezés során is. Unoka és Richman (2016) a végrehajtó funkciók erős, szignifikáns deficitjét mutatta ki BPD-ben érintett személyeknél kontrollszemélyekkel összehasonlítva. Hasonló eredményeket hozott Michopoulos és munkatársai (2021) kísérlete, amelyben a BPD-vel élő személyek a végrehajtó funkciókat mérő tesztek közül a *Stockings of Cambridge* és *Intradimensional/Extradimensional Attentional Set Shifting* (ID/ED) teszteken a kontrollszemélyekhez képest, az ID/ED teszten pedig a bipoláris személyekhez képest is szignifikánsan gyengébben teljesítettek.

A **kitöltött szünetek** jelenléte utalhat a fogalmi és a nyelvi megformálás háttérben zajló működésére (Beattie & Butterworth, 1979; Levelt, 1989; Shriberg, 2001; Watanabe et al., 2008; Gósy et al., 2013; Gyarmathy, 2017), illetve megjelenhetnek más megakadásjelenségek, például a hibajelenségek és javításuk kísérőjelenségeként (Gyarmathy, 2017) vagy a rejtett önmonitorozás felszíni jeleként (Christenfeld, 1996; Swerts, 1998; Clark & Fox Tree, 2002; Simpson, 2006; Finlayson & Corley, 2012; Gósy et al., 2013). A BPD-ben érintett személyeknél a nyelvi enkódolási hibák magas aránya miatt kevésbé valószínű, hogy a kitöltött szünetek a nyelvi megformálás és az önmonitorozás rejtett folyamatát jeleznék. Elképzelhető, hogy a kitöltött szünetek BPD populációban becsült magasabb aránya döntően a fogalmi szint, azon belül pedig az **önéletrajzi emlékek előhívásának nehézségével** magyarázható. Unoka és Richman (2016) az emlékezeti folyamatok erős, szignifikáns deficitjét mutatta ki BPD-vel élő személyeknél kontrollszemélyekkel összehasonlítva. Berán és Unoka (2016) szerint amikor a BPD-vel élő személyek epizodikus emlékeket hívnak elő, az epizodikus emlékeket alkotó élményközei (szenzoros-perceptuális-affektív) „epizodikus elemek” aktivációja gátolt, és csupán az epizodikus elemeket strukturáló és ér-

telmező „fogalmi keretekhez” (pl. „nem kedvel”, „kiközösítettek”) férnek hozzá (Conway, 2009). Számos kutatás eredményei konzisztensek ezzel a megállapítással, melyekben azt találták, hogy a BPD-ben érintett személyek szignifikánsan sematikusabb önéletrajzi emlékeket hívtak elő az egészséges kontrollszemélyekhez képest (Jones et al., 1999; Maurex et al., 2010; Reid & Startup, 2010; Rosenbach & Renneberg, 2015; Bendstrup et al., 2021). Mivel a szelf kontinuitását az önéletrajzi emlékek szekvenciája adja (Fivush, 2011), és az önéletrajzi emlékeknek lényeges vonása a specifikusság (Vanderveren et al., 2017), az önéletrajzi emlékek előhívásának nehézsége – véleményünk szerint – összefüggésbe hozható a BPD-vel élő személyekre jellemző **identitászavarral**.

A nyelvi enkódolási hibák és a kitöltött szünetek számának magas aránya utalhat **szorongásra** is (ld. 1.2. alfejezet), amely a BPD esetében az érzelmi labilitásból ered (ld. 1.1. alfejezet). Villarroel és munkatársai (2020) NEO személyiségtesztet vettek fel egy 43 BPD-ben érintett és 30 bipoláris személyt magában foglaló mintán, amelynek eredményeként azt kapták, hogy a BPD-vel élő személyek mind a kontrollcsoportnál, mind a BD csoportnál szignifikánsan magasabb értéket értek el a Big Five „neuroticizmus” dimenzióján, amely a DSM-5-TR negatív affektivitás tartományának feleltethető meg. Di Giacomo és munkatársai (2017) 248 BPD-ben érintett és 113 bipoláris személyen tesztelték a Hamilton Szorongás Skálát, amelyen a BPD-vel élő személyek szignifikánsan magasabb szorongást mutattak a bipoláris személyekhez viszonyítva. Az érzelmi labilitást – így a szorongást is – jellemzően interperszonális inger váltja ki, ez a BPD-ben érintett személyek esetében olyan traumatikus gyermekkori emlékeket idézhet fel, amelynek hatására az adott ingert kritikaként, visszautasításként vagy elhagyásként értelmezik (ld. 1.1. alfejezet), így elképzelhető, hogy az interjú – mint társas kontaktus – során a BPD-vel élő személyek kritikának érzik kitéve magukat.

A **néma szünetek** megjelenése történhet az artikulációhoz szükséges légáram biztosítása céljából, a közlés értelmi tagolása miatt, időt biztosíthat a mentális lexikonban való kereséshez, valamint a rejtett önmonitorozást is jelezheti (Gósy, 2000). Elképzelhető, hogy mivel a kontrollszemélyek összességében

kevesebb felszíni hibát vétettek, mint a BPD-ben érintett személyek (ld. 4. táblázat), a néma szünetek egészséges populációban becsült magasabb gyakorisága azzal magyarázható, hogy a BPD-vel élő személyekhez képest az egészséges személyek **jobb önmonitorozással** rendelkeznek, valamint nagyobb hangsúlyt fektetnek arra, hogy narratívájuk **értelmileg tagoltabb** legyen.

Az eredmények értelmezésekor szem előtt kell tartanunk a kutatás korlátait, amelyek jelen esetben a kis elemszámú mintát, a BPD csoport nagy szórásait és a kontrollszemélyek laza szűrési feltételeit foglalják magukban. A kis elemszám és a nagy szórás kevésbé reprezentatív mintát eredményez, melynek következményeként a statisztikai modell kevésbé lesz szenzitív a populációban fellelhető hatásokra. Mivel a kontrollszemélyek semmilyen klinikai szűrésen nem estek át, nem lehetünk biztosak abban, hogy egyikük sem rendelkezett semmilyen mentális zavarral. A felsorolt hiányosságok miatt a fenti magyarázatok jelenleg csupán spekulációnak tekinthetők, további vizsgálatok szükségesek érvényességük teszteléséhez. Az így kiegészített elméleti keret pedig újabb hipotézisekhez vezethet arra vonatkozóan, hogy mely beszédjellemzők statisztikai tulajdonságainak bevonásával pontosítható tovább a számítógépes modell.

Kutatásunk következő lépéseként ugyanezen a mintán tervezünk szógyakorisági (személyjelölés, szavak valenciája stb.) és koherenciaelemzést végezni az ezeken a területeken végzett kutatásokra alapozva (ld. 1.2.1. alfejezet). Ezt követően pedig – tekintettel a magas komorbiditásra és az ebből fakadó differenciál-diagnosztikai nehézségekre – kutatásunkat a BD-re is tervezzük kiterjeszteni. Távlati célunk, hogy a beszéd temporális jellemzőinek automatikus elemzése a jövőben egyrészt a nyelvelméleti kutatásokhoz, másrészt a klinikai differenciál-diagnózishoz is hozzájárulhasson. Ennek érdekében tervezzük az automatikus elemzést nagyobb és kiterjedtebb beszédmintán is elvégezni.

Köszönetnyilvánítás

Felletár Fanni kutatását a Magyar Tudományos Akadémia fiatal kutatói ösztöndíjprogramja támogatta. Gosztolya Gábor kutatásait a Mesterséges Intelli-

gencia Laboratórium (RRF-2.3.1-21-2022-00004) pályázata támogatta. Köszönjük Kneszl Beáta Carmennek, BPD aktivistának és írónak, hogy részt vett a kutatásban, és emellett népszerűsítette a részvételi lehetőséget a BPD közösségben, valamint Káldi Tamásnak, az ELKH Nyelvtudományi Kutatóközpont kutatójának, aki pedig az ELTE Bárczi Gusztáv Gyógypedagógiai Karon, egészséges populációban hirdette a lehetőséget. Köszönjük továbbá Bolya Mátyásnak, az ELKH Zenetudományi Intézet kutatójának, hogy közreműködött a hangfelvételek gépi átírásában. Végül köszönjük a kutatás összes résztvevőjének, hogy narratíváikkal hozzájárultak az adatgyűjtés sikerességéhez.

Hivatkozások

- Adler, J. M., Chin, E. D., Kolisetty, A. P., & Oltmanns, T. F. (2012). The distinguishing characteristics of narrative identity in adults with features of borderline personality disorder: An empirical investigation. *Journal of Personality Disorders*, 26, 498–512. doi:<https://doi.org/10.1521%2Fpedi.2012.26.4.498>.
- Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. *J. Semant.*, 9, 1–26.
- American Psychiatric Association (2022). *Diagnostic and statistical manual of mental disorders (5th ed., text rev.)*. Washington, DC: American Psychiatric Association.
- Audacity Team (2021). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 3.0.0. URL: <https://audacityteam.org/>.
- Bagi, A., Gosztolya, G., Szalóki, S., Szendi, I., & Hoffmann, I. (2019). Szkizofrénia azonosítása spontán beszéd temporális paraméterei alapján: egy pilot kutatás eredményei. In *MSZNY* (pp. 189–201). Szeged.
- Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, 22, 201–211.

- Bendstrup, G., Simonsen, E., Kongerslev, M. T., Jørgensen, M. S., Petersen, L. S., Thomsen, M. S., & Vestergaard, M. (2021). Narrative coherence of autobiographical memories in women with borderline personality disorder and associations with childhood adversity. *Borderline Personality Disorder and Emotion Dysregulation*, 8, 18. doi:<https://doi.org/10.1186/s40479-021-00159-5>.
- Berán, E., & Unoka, Z. (2016). *Élettörténetek a pszichoterápiában*. Budapest: Oriold és Társai.
- BNO-11 (2022). *BNO-11 Mentális zavarok*. Budapest: Animula Kiadó.
- Breitenstein, C., Lancker, D. V., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, 15, 57–79.
- Bruchmüller, K., Margraf, J., Suppiger, A., & Schneider, S. (2011). Popular or unpopular? therapists' use of structured interviews and their estimation of patient acceptance. *Behavior Therapy*, 42, 634–43. doi:<https://doi.org/10.1016/j.beth.2011.02.003>.
- Carter, P. E., & Grenyer, B. F. S. (2012). Expressive language disturbance in borderline personality disorder in response to emotional autobiographical stimuli. *Journal of Personality Disorders*, 26, 305–321. doi:<https://doi.org/10.1521/pedi.2012.26.3.305>.
- Christenfeld, N. (1996). Effects of a metronome on the filled pauses of fluent speakers. *Journal of Speech and Hearing Research*, 39, 1232–1238.
- Christenfeld, N., & Creager, B. (1996). Anxiety, alcohol, aphasia, and ums. *Journal of Personality and Social Psychology*, 70, 451–460. doi:<https://doi.org/10.1037/0022-3514.70.3.451>.
- Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84, 73–111.

- Conway, M. A. (2009). Episodic memories. *Neuropsychologia*, *47*, 2305–2313. doi:<https://doi.org/10.1016/j.neuropsychologia.2009.02.003>.
- Cwik, J. C., Papen, F., Lemke, J.-E., & Margraf, J. (2016). An investigation of diagnostic accuracy and confidence associated with diagnostic checklists as well as gender biases in relation to mental disorders. *Front. Psychol*, *7*. doi:<https://doi.org/10.3389/fpsyg.2016.01813>.
- Edell, W. S. (1987). Role of structure in disordered thinking in borderline and schizophrenic disorders. *Journal of Personality Assessment*, *51*, 23–41. doi:https://doi.org/10.1207/s15327752jpa5101_3.
- Egas-López, J. V., Balogh, R., Imre, N., Tóth, L., Vincze, V., Pákási, M., Kálmán, J., Hoffmann, I., & Gosztolya, G. (2021). Enyhe kognitív zavar detektálása beszédhangból x-vektor reprezentáció használatával. In *MSZNY* (pp. 147–156). Szeged.
- Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review*, *19*, 921–928.
- Fivush, R. (2011). The development of autobiographical memory. *Annu Rev Psychol*, *62*, 559–82. doi:<https://doi.org/10.1146/annurev.psych.121208.131702>.
- di Giacomo, E., Aspesi, F., Fotiadou, M., Arntz, A., Aguglia, E., Barone, L., Bellino, S., Carpiniello, B., Colmegna, F., Lazzari, M., Loretto, L., Pinna, F., Sicaro, A., Signorelli, M. S., Group, B., & Clerici, M. (2017). Unblending borderline personality and bipolar disorders. *Journal of Psychiatric Research*, *91*, 90–97. doi:<https://doi.org/10.1016/j.jpsychires.2017.03.006>.
- Giannini, A. (2003). Hesitation phenomena in spontaneous Italian. In M.-J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2653–2656). Barcelona: Universitat Autònoma de Barcelona.

- Gilquin, G., & De Cock, S. (2011). Errors and disfluencies in spoken corpora. *Int. J. Corpus Ling*, 16, 141–172. doi:<https://doi.org/10.1075/ijcl.16.2.01gil>.
- Gósy, M. (2000). A beszédszünetek kettős funkciója. *Beszéd kutatás 2000*, (pp. 1–14).
- Gósy, M. (2002). A megakadásjelenségek eredete a spontán beszéd tervezési folyamatában. *Magyar Nyelvőr*, 126, 192–204.
- Gósy, M. (2005). *Pszicholingvisztika*. Budapest: Osiris.
- Gósy, M., Bóna, J., Beke, A., & Horváth, V. (2013). A kitöltött szünetek fonetikai sajátosságai az életkor függvényében. *Beszéd kutatás 2013*, (pp. 121–143).
- Gósy, M., Gyarmathy, D., Horváth, V., Grácsi, T. E., Beke, A., Neuberger, T., & Nikléczy, P. (2012). Bea: Beszélt nyelvi adatbázis. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* (pp. 9–24). Budapest: Akadémiai Kiadó.
- Gosztolya, G., Balogh, R., Imre, N., Egas-López, J. V., Hoffmann, I., Vincze, V., Tóth, L., Devanand, D. P., Pákáski, M., & Kálmán, J. (2021). Cross-lingual detection of mild cognitive impairment based on temporal parameters of spontaneous speech. *Computer, Speech and Language*, 69. doi:<https://doi.org/10.1016/j.csl.2021.101215>. Article no. 101215.
- Gosztolya, G., Vincze, V., Tóth, L., Pákáski, M., Kálmán, J., & Hoffmann, I. (2019). Identifying mild cognitive impairment and mild alzheimer’s disease based on spontaneous speech using asr and linguistic features. *Computer Speech and Language*, 53, 181–197. doi:<https://doi.org/10.1016/j.csl.2018.07.007>.
- Grósz, T., Busa-Fekete, R., Gosztolya, G., & Tóth, L. (2015). Assessing the degree of nativeness and Parkinson’s condition using Gaussian processes and deep rectifier neural networks. In *Interspeech* (pp. 1339–1343). Dresden, Germany.

- Guidi, A., Schoentgen, J., Bertschy, G., Gentili, C., Scilingo, E. P., & Vanello, N. (2017). Features of vocal frequency contour and speech rhythm in bipolar disorder. *Biomed Signal Process Control*, 37, 23–31. doi:<https://doi.org/10.1016/j.bspc.2017.01.017>.
- Gunderson, J., & Sabo, A. (1993). The phenomenological and conceptual interface between borderline personality disorder and PTSD. *American Journal of Psychiatry*, 150, 19–27. doi:<https://doi.org/10.1176/ajp.150.1.19>.
- Gyarmathy, D. (2015). Diszharmóniás jelenségek, megakadások a beszédben. In M. Gósy (Ed.), *Diszharmóniás jelenségek a beszédben* (pp. 9–48). Budapest: MTA Nyelvtudományi Intézet.
- Gyarmathy, D. (2017). *Megakadásjelenségek a magyar spontán beszédben*. Budapest: MTA Nyelvtudományi Intézet.
- Heinz, M. V., Thomas, N. X., Nguyen, N. D., Griffin, T. Z., & Jacobson, N. C. (2022). Technological advances in clinical assessment. In G. J. G. Asmundson (Ed.), *Comprehensive Clinical Psychology, 2nd edition, Vol. 4* (pp. 301–320). Elsevier. doi:<https://dx.doi.org/10.1016/B978-0-12-818697-8.00171-0>.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Processing Magazine*, 29, 82–97.
- Hoffmann, I., Németh, D., Dye, C., Pákáski, M., Irinyi, T., & Kálmán, J. (2010). Temporal features of spontaneous speech in Alzheimer’s disease. *International Journal of Speech-Language Pathology*, 12, 29–34. doi:<https://doi.org/10.3109/17549500903137256>.
- Hoffmann, I., Tóth, L., Gosztolya, G., Szatlóczki, G., Vincze, V., Kárpáti, E., Pákáski, M., & Kálmán, J. (2017). Beszédfelismerés alapú eljárás az enyhe kognitív zavar automatikus felismerésére spontán beszéd alapján. In

- Z. Bánréti (Ed.), *Kísérletes nyelvészet* (pp. 385–405). Budapest: Akadémiai Kiadó.
- Horváth, V. (2017). Megakadásjelenségek és időzítési sajátosságai 6–9 éves gyermekek spontán narratíváiban. In J. Bóna (Ed.), *Új utak a gyermeknyelvi kutatásokban* (pp. 97–120). Budapest: ELTE Eötvös Kiadó.
- JASP Team (2022). JASP (Version 0.16.4)[Computer software]. URL: <https://jasp-stats.org/>.
- Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis, & J. Haviland-Jones (Eds.), *Handbook of emotions (Second edition)* (pp. 220–235). New York: Guilford Press.
- Jones, B., Heard, H., Startup, M., Swales, M., Williams, J. M., & Jones, R. S. P. (1999). Autobiographical memory and dissociation in borderline personality disorder. *Psychological Medicine*, 29, 1397–1404. doi:<https://doi.org/10.1017/s0033291799001208>.
- Kasl, S. V., & Mahl, G. F. (1965). Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, 1, 425–433. doi:<https://doi.org/10.1037/h0021918>.
- Kernberg, O. F. (1975). *Borderline Conditions and Pathological Narcissism*. New York: Jason Aronson.
- Kiss, G., Simon, L., & Vicsi, K. (2017). Depresszió súlyosságának becslése beszédjel alapján magyar nyelven. In *MSZNY* (pp. 125–135). Szeged.
- Klipper, R., Portuguese, S., & Weinshall, D. (2016). Prosodic analysis of speech and the underlying mental state. In S. Serino, A. Matic, D. Giakoumis, G. Lopez, & P. Cipresso (Eds.), *Pervasive Computing Paradigms for Mental Health* (pp. 52–62). NY, NY: Springer International Publishing.
- Klipper, R., Vaizman, Y., Weinshall, D., & Portuguese, S. (2010). Evidence for depression and schizophrenia in speech prosody. In *Third ISCA Workshop on Experimental Linguistics* (pp. 19–23). Saint-Malo, France.

- Koenigsberg, H. W. (2010). Affective instability: Toward an integration of neuroscience and psychological perspectives. *Journal of Personality Disorders*, *24*, 60–82.
- Laukka, P., Juslin, P. N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, *19*, 633–653.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Lind, M., Vanwoerden, S., Penner, F., & Sharp, C. (2019). Inpatient adolescents with borderline personality disorder features: Identity diffusion and narrative incoherence. *Personality Disorders*, *10*, 389–393. doi:<https://doi.org/10.1037/per0000338>.
- Linehan, M. M. (1993). *Cognitive-Behavioral Treatment of Borderline Personality Disorder*. New York, United States: Guilford Publications.
- Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*, *5*, 96–116. doi:<https://doi.org/10.1002/liv.2.354>.
- Lyons, M., Aksayli, N. D., & Brewer, G. (2018). Mental distress and language use: Linguistic analysis of discussion forum posts. *Computers in Human Behavior*, *87*, 207–211. doi:<https://doi.org/10.1016/j.chb.2018.05.035>.
- Maurex, L., Lekander, M., Nilsson, Å., Andersson, E. E., Åsberg, M., & Öhman, A. (2010). Social problem solving, autobiographical memory, trauma, and depression in women with borderline personality disorder and a history of suicide attempts. *British Journal of Clinical Psychology*, *49*, 327–342. doi:<https://doi.org/10.1348/014466509x454831>.
- Maxhuni, A., Muñoz-Meléndez, A., Osmani, V., Perez, H., Mayora, O., & Morales, E. F. (2016). Classification of bipolar disorder episodes based on analysis

- of voice and motor activity of patients. *Pervasive and Mobile Computing*, 31, 50–66. doi:<https://doi.org/10.1016/j.pmcj.2016.01.008>.
- McDougall, K., & Duckworth, M. (2017). Profiling fluency: An analysis of individual variation in disfluencies in adult males. *Speech Communication*, 95, 16–27. doi:<https://doi.org/10.1016/j.specom.2017.10.001>.
- Michopoulos, I., Tournikioti, K., Paraschakis, A., Karavia, A., Gournellis, R., Smyrnis, N., & Ferentinos, P. (2021). Similar or different neuropsychological profiles? only set shifting differentiates women with bipolar vs. borderline personality disorder. *Frontiers in Psychiatry*, 12. doi:<https://doi.org/10.3389/fpsy.2021.690808>.
- Millon, T., Grossman, S., Millon, C., Meagher, S., & Ramnath, R. (2004). *Personality Disorders in Modern Life (2nd ed.)*. Hoboken, NJ: Wiley.
- 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 *Proceedings of TSD* (pp. 424–431). Brno, Czech Republic. doi:http://dx.doi.org/10.1007/978-3-319-10816-2_51.
- Peters, B. (2003). Multiple cues for phonetic phrase boundaries in German spontaneous speech. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1795–1798).
- Rapcan, V., D’Arcy, S., Yeap, S., Afzal, N., Thakore, J., & Reilly, R. B. (2010). Acoustic and temporal analysis of speech: a potential biomarker for schizophrenia. *Medical Engineering & Physics*, 32, 1074–1079. doi:<https://doi.org/10.1016/j.medengphy.2010.07.013>.
- Reid, T., & Startup, M. (2010). Autobiographical memory specificity in borderline personality disorder: associations with co-morbid depression and intellectual ability. *British Journal of Clinical Psychology*, 49, 413–420. doi:<https://doi.org/10.1348/014466510x487059>.

- Rosenbach, C., & Renneberg, B. (2015). Remembering rejection: Specificity and linguistic styles of autobiographical memories in borderline personality disorder and depression. *Journal of Behavior Therapy and Experimental Psychiatry*, *46*, 85–92. doi:<https://doi.org/10.1016/j.jbtep.2014.09.002>.
- Régens (2021). Alrite. URL: <https://alrite.io/ai/>.
- S. Nagy, Z., Salgó, E., Bajzát, B., Hajduska-Dér, B., & Unoka, Z. S. (2022). Reliability and validity of the hungarian version of the personality inventory for DSM-5 (PID-5). *PLoS ONE*, *17*, e0266201. doi:<https://doi.org/10.1371/journal.pone.0266201>.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227–256. doi:[https://doi.org/10.1016/S0167-6393\(02\)00084-5](https://doi.org/10.1016/S0167-6393(02)00084-5).
- Shriberg, E. (2001). To „errrr” is human: ecology and acoustic of speech disfluencies. *Journal of the International Phonetic Association*, *31*, 153–169.
- Simpson, A. P. (2006). Phonetic processes in discourse. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (pp. 284–288). Amsterdam: Elsevier.
- Smits, M. L., Feenstra, D. J., Bales, D. L., de Vos, J., Lucas, Z., Verheul, R., & Luyten, P. (2017). Subtypes of borderline personality disorder patients: a cluster-analytic approach. *Borderline Personal Disord. Emot. Dysregul.*, *4*, 16. doi:<https://doi.org/10.1186/s40479-017-0066-4>.
- Spruit, M., Verkleij, S., de Schepper, K., & Scheepers, F. (2022). Exploring language markers of mental health in psychiatric stories. *Applied Sciences*, *12*. doi:<http://dx.doi.org/10.3390/app12042179>.
- Svindt, V., Bóna, J., & Hoffmann, I. (2020). Changes in temporal features of speech in secondary progressive multiple sclerosis (spms): case studies. *Clinical Linguistics & Phonetics*, *34*, 339–356. doi:<https://doi.org/10.1080/02699206.2019.1645885>.

- Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30, 485–496.
- Szokolszky, Á. (2004). *Kutatómunka a pszichológiában: Metodológia, módszerek, gyakorlat*. Budapest: Osiris.
- Sztahó, D., Kiss, G., & Vicsi, K. (2015). Estimating the severity of parkinson's disease from speech using linear regression and database partitioning. In *Interspeech* (pp. 498–502). Dresden, Germany.
- Tisljár-Szabó, E. (2014). Érzelmek és beszéd. In C. Pléh, & Á. Lukács (Eds.), *Pszicholingvisztika: Magyar pszicholingvisztikai kézikönyv 1* (pp. 951–987). Budapest: Akadémiai Kiadó.
- Tóth, L., Gosztolya, G., Vincze, V., Hoffmann, I., Szatlóczki, G., Biró, E., Zsura, F., Pákáski, M., & Kálmán, J. (2015). Automatic detection of mild cognitive impairment from spontaneous speech using ASR. In *Proceedings of Interspeech* (pp. 2694–2698). Dresden, Germany. doi:<http://dx.doi.org/10.21437/Interspeech.2015-568>.
- Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatlóczki, G., Bánréti, Z., Pákáski, M., & Kálmán, J. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research*, 15, 130–138. doi:<https://doi.org/10.2174/1567205014666171121114930>.
- Tringer, L. (2010). *A pszichiátria tankönyve*. Budapest: Semmelweis Kiadó.
- Tündik, M. Á., Kiss, G., Sztahó, D., & Szaszák, G. (2017). Automatikus frázisdetektáló módszereken alapuló patológiás beszédelemzés magyar nyelven. *MSZNY, Szeged*, (pp. 113–124).
- Unoka, Z., & Richman, M. J. (2016). Neuropsychological deficits in bpd patients and the moderator effects of co-occurring mental disorders: A meta-analysis. *Clinical Psychology Review*, 44, 1–12. doi:<https://doi.org/10.1016/j.cpr.2015.11.009>.

- Unoka, Z. S., Richman, M. J., & Czégel, D. (2018). How many causal pathways must symptoms form before we call them a borderline? a hierarchical network model of borderline personality disorder. *PsyArXiv*, . doi:<https://doi.org/10.31234/osf.io/93cnh>.
- Vanderveren, E., Bijttebier, P., & Hermans, D. (2017). The importance of memory specificity and memory coherence for the self: linking two characteristics of autobiographical memory. *Front Psychol*, 8. doi:<https://doi.org/10.3389/fpsyg.2017.02250>.
- Villanueva-Valle, J., Díaz, J. L., Jiménez, S., Rodríguez-Delgado, A., Arango de Montis, I., León-Bernal, A., Miranda-Terres, E., & Muñoz-Delgado, J. (2021). Facial and vocal expressions during clinical interviews suggest an emotional modulation paradox in borderline personality disorder: An explorative study. *Frontiers in Psychiatry*, 12. doi:<https://doi.org/10.3389/fpsyg.2021.628397>.
- Villarroel, J., Salinas, V., Silva, H., Herrera, L., Montes, C., Jerez, S., Vöhringer, P. A., & Bustamante, M. L. (2020). Beyond the categorical distinction between borderline personality disorder and bipolar II disorder through the identification of personality traits profiles. *Frontiers in Psychiatry*, 11. doi:<https://doi.org/10.3389/fpsyg.2020.00552>.
- Wang, B., Wu, Y., Taylor, N., Lyons, T., Liakata, M., Nevado-Holgado, A. J., & Saunders, K. E. A. (2020). Learning to detect bipolar disorder and borderline personality disorder with language and speech in non-clinical interviews. In *Interspeech 2020* (pp. 437–441). Shanghai, China. doi:<https://doi.org/10.21437/interspeech.2020-3040>.
- Wang, B., Wu, Y., Vaci, N., Liakata, M., Lyons, T., & Saunders, K. E. A. (2021). Modelling paralinguistic properties in conversational speech to detect bipolar disorder and borderline personality disorder. In *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7243–7247).

- Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication*, 50, 81–94.
- Young, S., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., & Woodland, P. (2006). *The HTK Book*. Cambridge, UK: Cambridge University Engineering Department.

Ybl a Google-ön. A szillabikus *l* és a magyar magánhangzó-harmónia esete

Blaskovics Ákos¹, Ittész Ambrus¹

¹ELTE BTK–Nyelvtudományi Kutatóközpont Kihelyezett Elméleti Nyelvészeti Központ

Abstract

In this paper, we examine a hitherto unexplored part of Hungarian loanword phonology, namely, the adaptation of words containing syllabic *l*, and the role that these words play in Hungarian vowel harmony. In Hungarian, there are no syllabic consonants, and word-final consonant–*l* clusters are almost completely absent. In a laboratory experiment, we studied how Hungarian speakers pronounce loanwords ending in a syllabic *l* and how they select suffix alternants. Data was collected from 28 speakers. They were asked to pronounce 6 target words in syntactic contexts that enforced the nominative, superessive (SUE; -on/-en/-øn) and allative (ALL; -hoz/-hez/-höz) forms. The target words included two surnames of German origin ([i:b], [ʃtro:b]) and four recent loans ([du:d], [gu:g], [lid], [paz]). After the experiment, the vowel quality of the suffixes attached to the target words and the number of realised syllables were examined along with the formant structure and length of *l* and the inserted vowel. As for results, firstly, speakers pronounced the target words in two syllables. The intonation peak in yes/no questions served as evidence. Secondly, both forms with inserted vowel and syllabic *l* ([gu:gV] ~ [gu:g]) are present. Insertion is more frequent in SUE forms than in ALL forms. Thirdly, neutral stems select mainly front alternants ([i:b]-höz, [lid]-øn), while in the case of back stems, both back and front suffixes occur ([paz]-on, [ʃtro:b]-höz). Therefore, syllabic *l* seems to be hardly transparent to labial harmony. Yet, it is so to palatal harmony to a larger extent. We claim also the emergence of a new harmonic class. However, the harmonic behaviour of the target words cannot be explained by the insertion of a reduced vowel. Only 60% of the forms with the front alternant contained one, and 78% of forms without any did select the front alternant. Thus, other factors may play a role, such as the acoustic similarity of [ø] and the alveolar consonants.

Keywords: loanword phonology, syllabic *l*, transparency, vowel harmony, vowel insertion

Email addresses: akosblaskovics@gmail.com (Blaskovics Ákos),
ittzes.ambrus@gmail.com (Ittész Ambrus)

1. Bevezetés

1.1. Kölcsönzésfonológia

A kölcsönyszavak fonológiája, tehát annak kutatása, hogy idegen szavak hangtanilag hogyan adaptálódnak egy nyelvbe, az utóbbi évtizedek szakirodalmában az érdeklődés középpontjába került. Ez többek között azzal magyarázható, hogy a kölcsönyszavak adaptálódásának vizsgálatával közvetlenül ragadhatók meg az adott nyelvre jellemző fonológiai folyamatok, így ez a teljes szinkrón rendszer megértéséhez is hozzásegíthet (Kenstowicz, 2003). A kölcsönzésfonológia elsősorban azt vizsgálja, hogy a fogadó nyelv milyen javítóstratégiákat alkalmaz, hogy az átadó nyelv szavait a saját hangtanába ágyazza.

A magyar kölcsönzésfonológiai szakirodalom elsősorban a mássalhangzónyúlás (Nádasdy, 1989) kérdéseivel foglalkozik. Kutatásunkban egy eddig nem vizsgált csoport, a szillabikus *l*-t tartalmazó névszók viselkedését és a magánhangzó-harmóniára való hatásukat vizsgáltuk.

A magyarban szótagmag-pozícióban nem állhat mássalhangzó, vagyis a magyar fonotaktikában ismeretlenek a szillabikus mássalhangzók. Több olyan nyelvből kerültek viszont jövevényszavak a magyarba, melyekben léteznek szillabikus mássalhangzók. A magyar szókincs részét képezi kb. 100 lexéma, melyeken megfigyelhető a kölcsönzésre alkalmazott egykori javítóstratégia, a szóvégi *i*-betoldás (pl. *Kipfel* → *kifli*, *Zettel* → *cetli*). Ezek nagyrészt a németből, azon belül is, ahogy a legtöbb német jövevényszó, a felnémet típushoz tartozó osztrák-bajor dialektusból kerültek a magyarba (Zsilinszky, 2003). Emellett előfordulnak jiddis (pl. *mázli*) és angol (pl. *szingli*) eredetű jövevényszavak is. A *-li* végződés mellett előfordul a *-ni* is (*fecni*, *zokni*), hiszen az *n* szintén szótagalkotó mássalhangzó lehet a németben és az angolban. A szillabikus *n*-nel a továbbiakban nem foglalkozunk.

A szóvégi *i*-betoldás mint javítóstratégia a mai magyarban már nem teljesen produktív (elég csak a [pazl], [gu:gl] szóalakokra gondolnunk). Anekdotikus ismereteink alapján a beszélők az újabb jövevényszavaknál egyaránt használják a három lehetséges stratégiát: a szóvégi *i*-betoldás mellett (pl. [kindli] ‘Kindle’)

használják ezeket az alakokat az utolsó két hang közé betoldott magánhangzóval, és kiejtve a magyar fonotaktikának nem részét képező szillabikus *l*-t is. Kutatásunk első kérdésköre az volt, hogy ha nem alkalmazzák az *i*-betoldást, akkor hogyan valósítják meg a beszélők a vizsgált szavakat.

1.2. A magyar magánhangzó-harmónia

Az alábbiakban áttekintjük a magyar magánhangzó-harmónia főbb jellegzetességeit, és azt, hogy miért jelent nehézséget a leírás számára a vizsgált jövevényszavak harmonikus viselkedése. Leírásunk Rebrus & Törkenczy (2019)-en alapul.

A magyar magánhangzó-harmónia tövezérelt, balról jobbra irányuló, harmonikus tartománya a szó. Megkülönböztethetünk előlségi és kerek ségi illeszkedést, ezek kiváltó és célosztályait az alábbi 1. táblázat foglalja össze. A kerek ségi illeszkedés parazitikus: csak olyan környezetekben fordul elő, ahol az előlségi illeszkedés is aktív. Egyetlen így kialakuló hármas harmonikus váltakozás figyelhető meg ($o \sim e \sim \ddot{o}$).

	Kiváltó osztály	Célosztály
Előlségi illeszkedés	{összes V}	{összes V \ {i, í}}
Kerek ségi illeszkedés	{elülső V-k}	{e, ö}

1. táblázat. Kiváltó és célosztályok

A magánhangzóképzőletet harmonikus szemponból három csoportra oszthatjuk, hátsó, elülső és semleges magánhangzókra. A csoportok áttekintését mutatja az alábbi 2 táblázat.

Megnevezés	Elemek	Jelölés
hátsó magánhangzók	{u, ú, o, ó, a, á}	B
elülső kerekített magánhangzók	{ü, ű, ö, ő}	F
semleges magánhangzók	{i, í, é, (e)}	N

2. táblázat. A magánhangzók csoportosítása a magánhangzó-harmónia szempontjából

A tő hangrendje váltakozó toldalékok esetén meghatározza a toldalékok hangrendjét is. Háromalakú toldalékoknál tehát léteznek a [B]B, [F]F és [N]N mintázatok, de nincsenek [B]F, [B]N, [F]B, [F]N, és [N]F szekvenciák (*naphoz*, *fülhöz*, *kézhez*, de **napön*, **napen*, **fülon*, **fülen*, **kézhöz*).¹ Többszótagú tövek esetén a toldalék harmóniáját legtöbbször a tő utolsó magánhangzója határozza meg. A semleges magánhangzók viszont többszótagú tövekben átlátszók lehetnek, ilyenkor a toldalék hangrendjét az első magánhangzó határozza meg (*radírhoz*, **radírhez*). A létező harmonikus mintázatok köre tehát erősen korlátozott. A továbbiakban a háromalakú toldalékok alternánsaira a következőképpen hivatkozunk: az *ö*-t tartalmazót elülső, az *e*-t semleges, az *o*-t tartalmazót hátsó alternánsnak fogjuk hívni.

Nem egyértelmű, hogy a beszélők milyen toldalékalternánst fognak használni egy szillabikus *l*-t tartalmazó szóalak esetén. A szillabikus mássalhangzók (minthogy nem részei a magyar hangkészletnek) nem tartoznak egyik harmonikus osztályba sem. Kérdéses, hogy ilyen helyzetben a tő „stabil” magánhangzója, vagy a szillabikus mássalhangzó számít-e, és ha a mássalhangzó, milyen toldalékot követel meg. Hogyan befolyásolja ezt, hogy megvalósul-e betoldott magánhangzó az *l* előtt? Kutatásunk második kérdésköre a szillabikus *l*-t tartalmazó szavak harmonikus viselkedése volt.

A korpuszkutatás mint lehetséges kutatási módszer nem tette volna lehetővé a fonetikai megvalósulás vizsgálatát, és a szóalakok ritkasága miatt a magánhangzó-harmóniára vonatkozóan is csak nagyon korlátozott következtetések levonását engedte volna meg. A felvetődő kérdések részletesebb vizsgálatára így akusztikai kísérletet terveztünk, amelynek eredményei alapján a fonetikaiakon túl a fentebb vázolt problémakört jobban megvilágító fonológiai állításokat tehetünk.

A kutatás során kizárólag háromalakú toldalékokat vizsgáltunk, az elülső és hátsó tövek közvetlen összehasonlíthatósága érdekében. A kétalakú toldalékok ugyanis semmit nem árulnak el az első magánhangzóként semleges magánhangzót

¹[N]B mintájú toldalékolt alakok léteznek, a semleges magánhangzót tartalmazó, de hátsó toldalékot kiváltó töveket antiharmonikusnak nevezzük. Jelen kutatás szempontjából nincs fontos szerepük.

tartalmazó tövek harmonikus viselkedéséről. Hiszen míg a *Strobl* szó megfelelő alakja lehet *Stroblnál* vagy *Stroblnél*, és ebből levonhatunk következtetéseket, a *Lidl* tő adessivusi esete *Lidlnél* lesz, akár szerepet játszik a harmóniában valamilyen második szótag, akár nem.

1.3. Hipotézisek

Kutatásunk tehát a szillabikus *l*-t tartalmazó szavak adaptálódását vizsgálta a magyar fonológiai rendszerbe. Korábbi kutatások híján mind a szavak fonetikai megvalósulásával, mind a harmonikus rendszerbe való illeszkedésével kapcsolatban fogalmaztunk meg hipotéziseket. Ezek a következők:

1. A beszélők két szótaggal ejtik a szillabikus *l*-t tartalmazó jövevényszavak töveit.
2. A második szótagok szótagmagja kétféle lehet: előfordul szillabikus *l* és a két szóvégi mássalhangzó közé betoldott redukált magánhangzó is (*gugVl* $\sim$ *gugl*).
3. Háromalakú toldalékok esetén az elülső kerekített alternáns jelenik meg (*lidlön*, *guglhöz*). Hátsó magánhangzót tartalmazó tövek után előfordul hátsó magánhangzó is, vagyis a második szótag részben átlátszó (*guglhöz* $\sim$ *guglhoz*).
4. Az elülső kerekített alternáns akkor fordul elő, ha a *Cl*-szekvencia két tagja közé redukált V-t told be a beszélő (*guglhoz*, de *gugVlhöz*).
5. Az *-On* toldalék előtt az *l* tartama nagyobb variabilitást mutat. Előfordulhat geminálódás (*lidlön* $\sim$ *lidllön*).

Tanulmányunk a következőképpen épül fel. A következő, 2. fejezetben a kísérlet megvalósítását, valamint a fonetikai és a statisztikai elemzés módszertanát részletezzük, míg a 3.-ban az elemzés során kapott eredményeket ismertetjük. A 4. fejezetben a jelen bevezetőben fölvetett fonológiai kérdések megválaszolására térünk vissza, végül az 5. fejezet az összefoglalást tartalmazza. A 2. (illetve a

3.) fejezetet egy online függelék egészíti ki, amely a következő linken érhető el: https://osf.io/mzgrd/?view_only=120e4fd17150472293a08b5e352b2cbe.

2. Módszertan

2.1. Kísérleti anyag

A hangfelvételeket a Nyelvtudományi Kutatóközpont fonetikai laboratóriumában készítettük, hangszigetelt stúdióban, omnidirekcionális fejmikrofonnal. A felvételekhez a SpeechRecorder szoftvert használtuk (Draxler & Jänsch, 2004).

Az adatközlők egy előttük lévő monitorról olvashatták le a feladatot: „Olvasd fel a mondatot úgy, hogy behelyettesíted az előtte látható szót a szerinted megfelelő toldalékkal ellátva!” (pl. *Strobl: A szobrok legfinomabb részletei sem tudtak kifogni _____; puzzle: Van még értelme este kilenckor hozzáfogni egy _____?*)

A kísérlethez hat célszót használtunk. Ezek között volt két német eredetű családnév (*Strobl, Ybl*), amelyek tehát egy védett lexikai réteghez tartoznak, a klasszikus javítóstratégia nem alkalmazódott rájuk. Emellett három márkanevet használtunk (*Lidl, Google, Doodle*),² amelyek a mai köznyelvben ingadozást mutatnak ([lidl:]~[lidli], [gu:gl:]~[gugli], [du:dl:]~[dudli]), és egy köznevet (*puzzle*), aminek kiejtése szintén ingadozik ([pazl:]~[pazli]).³

A vizsgált célszavakat három különböző szintaktikai környezetben kellett kimondaniuk a beszélőknek: egy zéró toldalékot, egy *-hoz/-hez/-höz-t* és egy *-on/-en/-ön-t* kikényszerítőben. (E toldalékokra a későbbiekben rendre allativusi és superessivusi toldalékokként is hivatkozunk.) Minden toldalékos alak kétszer szerepelt, egyszer kijelentő mondatban, egyszer eldöntendő kérdés végén. Az eldöntendő kérdésekre a szavak szótagszámának megállapításához volt szükség, ld. a 2.3. fejezetet. Összesen tehát minden tövet hatszor ejtett ki egy beszélő.

²Ahol semmi nem indokolja a fonetikai átírást, a szavak eredeti helyesírását fogjuk használni. Az ábrákon ettől eltérően egy nem fonetikai, de a kiejtést megközelítő (a magánhangzó-hosszúságot nem jelölő) átírást fogunk alkalmazni az ábrák könnyebb értelmezhetőségének kedvéért.

³A fenti felsorolásban az *i*-betoldásos alakok kiemelése végett nem vettük hozzá az egyes párokhoz a V-betoldást tartalmazóként tételezett [lidVI] típusú alakokat.

Az így kapott 36 célmondat mellé 72 disztraktor mondatot is elhelyeztünk. Ezekben a behelyettesítendő szavak szemantikailag hasonlítottak a célszavakra, és más harmonizáló toldalékokat várt el a szintaktikai környezet (pl. *Instagram: Te is percenként szoktad csekkolni az _____?*). A 108 mondatot randomizálva vetítettük az adatközlőknek, de úgy, hogy az egy lexémához tartozó alakok közül először mindig a toldalékolatlan alakú kijelentő mondatot lássák. Erre azért volt szükség, hogy az első mondat segítségével biztosítsuk, hogy a beszélők tisztában vannak a szó jelentésével (pl. *Doodle: A legtöbbek által használt időpont-egyeztető alkalmazás a _____.*)

2.2. Kísérleti személyek

A kísérletben 28 fő vett részt, ebből 14 férfi és 14 nő, életkoruk 19 és 57 év közötti (átlag: 29,1, szórás: 13,2). A többségük Budapesten vagy környékén lakik, származási helyük a településtípust és a Magyarországon belüli elhelyezkedést tekintve változó (a 28-ból egy beszélő partiumi származású volt). Egy vajdasági származású beszélőt kizártunk a kísérletből, mert a célszavaknak érdemben eltérő alakját használta ([ibɛl], [pa:zɛl]), ezzel nem érintve az általunk vizsgálni kívánt harmonikus mintákat.

A kísérlet korlátainak tekinthetjük azt a tényezőt, hogy az adatközlők túlnyomó többsége felsőoktatási végzettséggel rendelkezik, vagy jelenleg egyetemen tanul, a kutatás tehát semmiképp nem ad átfogó szociolingvisztikai képet. A kísérlet során laboratóriumi környezetben kellett produkálni a szavakat, amik eredeti helyesírásukkal voltak a mondatok elé írva. (Az íráskép szerepéről a kölcsönszavak adaptációjában lásd pl. Vendelin & Peperkamp 2006.)

A szóvégi *i*-betoldás és egyéb javítóstratégiák alkalmazása nem képezte a kutatás tárgyát. Az ilyen előfordulásokat később kizártuk az adatok elemzéséből. (Ugyanakkor egyértelművé vált a jelenség megléte. Az egyik adatközlő például a kísérlet után elmondta, hogy szándékosan alkalmazta egy helyen az *i*-betoldást, hogy jelezze a kísérlet összeállítói számára: élőbeszédben azt az alakot is használná.)

2.3. Módszer

A kísérlet során keletkezett hangfájlokon a fonetikai elemzést *Praatt*-al végeztük (Boersma & Weenink, 2021). A hangfájlokban a célmondatokban szereplő célszavakat, illetve az ezekben szereplő esetleges betoldott magánhangzót (a továbbiakban V-vel rövidítve) és az *l*-t annotáltuk, illetve a toldalékolt célszavak esetén a toldalék magánhangzóját is.

Az annotálást ketten végeztük. Egy beszélőt mindketten végigannotáltunk, és egy tapasztalt konzulens is felkértünk a beszélő annotálására. A megjelölt határokon fennálló különbségeket a 3. táblázat foglalja össze.

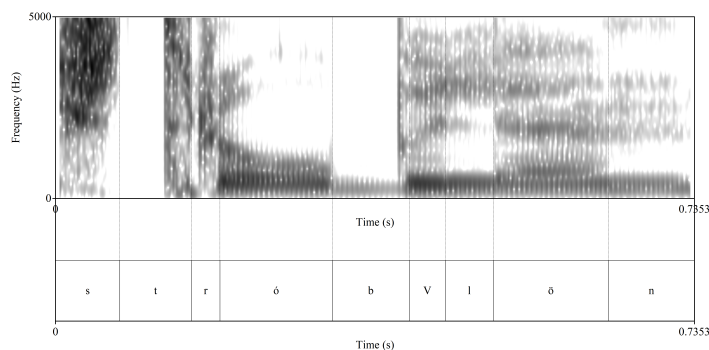
	1. annotátor	2. annotátor
(V) <i>l</i> szakasz kezdete	−2,4	0
V és <i>l</i> határa (ha van)	−1,2	+7,1
<i>l</i> vége	+5,4	+3,8

3. táblázat. A két annotátor átlagos eltérései a konzulenshez képest a próbaannotáció során (ms)

A V és *l* határának jelölése közti viszonylag nagyobb eltérést az okozta, hogy a 2. annotátor a formánsmenetek egészét a V-hez vette, a továbbiakban az 1. annotátorhoz hasonlóan ő is a formánsmenetek közepét jelölte.

A közösen kialakított stratégia alapján beszélőnként egyikünk annotálta a célszavakat, másikunk pedig a V-t, az *l*-t és – toldalékos alak esetén – a toldalék magánhangzóját. Az annotálás a spektrogram vizuális inspekciójával, a formánsmenetek elemzésével, illetve percepciók ítéletek segítségével történt. Az *l* és a V+*l* szekvenciák elkülönítésében szintén ezek játszottak szerepet (a formánsszerkezet tekintetében a formánsmenetek növekedését, valamint csökkenését, illetve az antiformánsokat vettük figyelembe). Az 1. ábrán látható egy példa egy olyan célszóra, aminél az antiformánsok éles jelenléte jól elkülöníthetővé teszi a betoldott V-t és az *l*-t.

Kérdéses esetben konzulenseinkhez fordultunk. Bár az *l*-ek szóvégen és mássalhangzók előtt vokalizálódhatnak, ezzel a jelenséggel nem foglalkoztunk mélyrehatóan, pusztán azt vizsgáltuk, hogy van-e magánhangzó-betoldás az *l*-et megjelenítő egység előtt.



1. ábra. Egy könnyen szegmentálható célszó

Az annotálás során a lehető legkonzisztensebb eljárásra törekedtünk, de az elemzéseket (és ebből kifolyólag az eredményeinket is) kellő óvatossággal kell kezelni. Egy magánhangzó és egy koronális likvida között nem mindig húzható éles határ, különösen, mivel a magánhangzó (mint látni fogjuk) nem teljesértékű, az *l* pedig gyakran vokalizálódik.

Az annotálás végeztével egy *Praat*-szkript segítségével minden fájl esetén kinyertük a célszó, a *V* és az *l* tartamát, továbbá a *V*, az *l* és a toldalék-magánhangzó esetén az annotált szakaszok középső 40%-án (30. és 70. percentilise között) vett első és második formánsok (F_1 és F_2) mediánját. A toldalék-magánhangzókat percpiciósan jegyeztük le.

Az adatokból töröltük azokat, amelyek tévesztés vagy hezitálás miatt egyáltalán nem értékelhetők (27 db az 1008-ból, kb. 2,7%). Ugyancsak kizártuk a további elemzésből az *i*-betoldással ejtett alakokat, ahogy azokat a superessivusi alakokat is, amelyekben a toldalék nem tartalmazott magánhangzót (pl. *lidln*). A harmóniát vizsgáló hipotéziseknél ez a döntés triviálisan következik, a többi esetében a megfelelő szakasznál fogjuk motiválni ezt a döntést. (23 ilyen alak fordult elő összesen, így az összes felvett adat kb. 95%-ával, 958 megfigyeléssel dolgoztunk tovább.) Csupán az 3.5.3 alszakasz fonológiai elemzése során térünk vissza erre az alaktípusra. Ahogyan azokra is, amelyeknek tövét egy szótagban ejtették a beszélők (részletesebben ld. 2.3.1.-et), ez további 24 kizárt megfigyelést eredményezett, vagyis összesen 934 megfigyeléssel dolgoztunk a következőkben

(ami az összes felvett adat kb. 92,7%-a). Ezek után a kapott adatokon statisztikai elemzést végeztünk az R statisztikai elemző szoftver segítségével, melynek módszertanát a következőkben foglaljuk össze. Részletes módszertani ismertetés a Függelékben olvasható.

2.3.1. *H1 – A tövek szótagszáma*

Az 1. hipotézis kapcsán az alakok szótagszámának megállapítására az „el-döntendő kérdés” tesztet (ld. Siptár & Törkenczy 2000) szintén mint percepció-s ítéletet alkalmaztuk. Fontos ugyanakkor leszögeznünk, hogy szótagszámot csak az (el-döntendő) kérdő mondatokban szereplő célszóalakok esetén állapítottunk meg, ugyanis a szótagszámot ezen mondatok intonációja alapján tudtuk meghatározni. Éppen ezért az ezzel kapcsolatos állításaink kizárólag az ezekben a mondatokban szereplő célszavak halmazára állhatnak. A célszavak a mondatokban valamennyi alkalommal nem-hangsúlyos pozícióban szerepeltek, így a frázis utolsó előtti szótagján kell lennie a dallam csúcsának: tehát várakozásaink szerint amennyiben a töveket két szótaggal (vagyis a toldalékolt alakokat összesen hárommal) ejtik, a toldalékolatlan alakok esetében a dallamcsúcsnak a tő első magánhangzójára kell esnie, míg a toldalékoltak esetében a toldalékot megelőző szótagra, amely az *l*-et is tartalmazza.

A *i*-betoldással javított alakok kizárása nem volt véletlen: az intonációelhe-lyezésük teljességgel megjósolható: tőalakban a tő első szótagjára, egyszótagos toldalék esetén a tő második szótagjára fog kerülni a magas tónus. Hasonlóan kevésbé meglepő azon alakok kizárása, ahol az „el-döntendő kérdés” teszt alap-ján egyszótagú tövet detektáltunk (tehát *strob.lon* és *strobl* típusú alakokat–*stro.bV.lön* és *stro.bl* típusúak helyett). Ugyanis ezekben az alakokban triviálisan nem jelenik meg betoldás (mert amennyiben megjelenne, éppen hogy létrejön egy újabb szótag), így ezek az adatok torzítanák a további hipotézisekre alkotott modellek kiértékelését, ahol a *V* megléte akár függő, akár független változóként szerepel. Sőt, az ilyen típusú allativusi alakok (pl. *strobl.hoz*) némelyikében magát az *l* szegmentumot sem tudtuk fölfedezni a spektrogramon. Bár a tesztet csak a kérdő mondati alakokra tudjuk alkalmazni, ennek ellenére a kérdő mon-

datban felfedett egyszótagos tövű alakok mellett kijelentő mondati párjukat is kizártuk, egyfelől feltételezve, hogy kijelentő mondatban nem eltérő szótagszámúak a tárgyalt alakok, másfelől azért, hogy minél nagyobb kiegyensúlyozottsággal bírjon a későbbiekben elemzett adatsor (tekintve, hogy a továbbiakban némelyik modellben a kérdés–kijelentés változó is szerepel).

2.3.2. H2 – A magánhangzó-betoldás megjelenése és időtartama

A 2. hipotézis kapcsán két modellt állítottunk föl. Egyrészt egy binomiális általánosított lineáris kevert modellt (*binomiális GLMM*) futtattunk a V-betoldás meglétére (ez a logisztikus regresszió alapszik), másrészt egy lineáris kevert modellt (*LMM*) arra, hogy milyen más változókkal áll összefüggésben a betoldás tartama. (Előbbi modelltípushoz a `GLMMadaptive` R-package `mixed_model`, utóbbihoz az `lme4` R-package `lmer` függvényét használtuk. Mivel különböző, a függelékben tárgyalt okokból az LMM-ek paraméterei torzultak (*biased*) lehetnek, ezek robusztus változatát is használtuk a `robustlmm` package-ből, ld. az `rlmer` függvényt.) A független változók közé mindkét alkalommal a toldaléktípus, a célszó, ezen kettő interakciója, illetve a mondattípus (kérdés–kijelentés) került; a random hatások az egyes beszélőknek feleltek meg, továbbá ezen belül random meredekséget számoltunk a toldaléktípusra is.

2.3.3. H3 – Az alternáns megválasztása háromalakú toldalékok esetén

A 3. hipotézisnél az egyes célszavak harmonikus tendenciáit írjuk pusztán le, statisztikai modellalkotás nélkül.

2.3.4. H4 – Összefüggés a magánhangzó-betoldás és az elülső alternáns megjelenése között

A 4. hipotézis vizsgálatához ugyancsak egy binomiális GLMM-et alkalmaztunk, azt vizsgálva, hogy mitől függ az elülső *ő* alternáns megjelenése a toldalékokban. Jelen esetben az adatok közül csak a célszavak toldalékolt alakjait vizsgáltuk, ezek közül is kizárva azokat, amelynek toldalékában nem szerepelt magánhangzó, vagy éppen szerepelt, de a szót két szótagban ejtették az „eldöntendő kérdés” teszt alapján (ld. a *lidln* említésekor fentebb az 1. hipotézisnél,

2.3.1), vagy pedig a tő első szótagjával antiharmonikus magánhangzó szerepelt (elenyésző számú ilyen eset fordult elő, az eddig használt 605 toldalékolt megfigyelésből 3 db, mindhárom a *-hOz* esetén). A modell független változói a célszó, a betoldás megléte és ezen kettő interakciója lettek.

2.3.5. *H5 – Variabilitás az l időtartamában*

Az 5. hipotézis (amely az l tartamára vonatkozó feltételezéseinket tartalmazta) kapcsán két LMM-et állítottunk fel, egyet az l , egy másikat pedig a Vl -szekvencia standardizált (z -score-osított) tartamára. A független változók mindkét esetben a betoldás-, a toldaléktípus-, a célszó- és a kérdés-változók lettek, különböző interakciókkal (részletesebben ld. a Függelék 5.1. és 5.2. szakaszát).

3. Eredmények

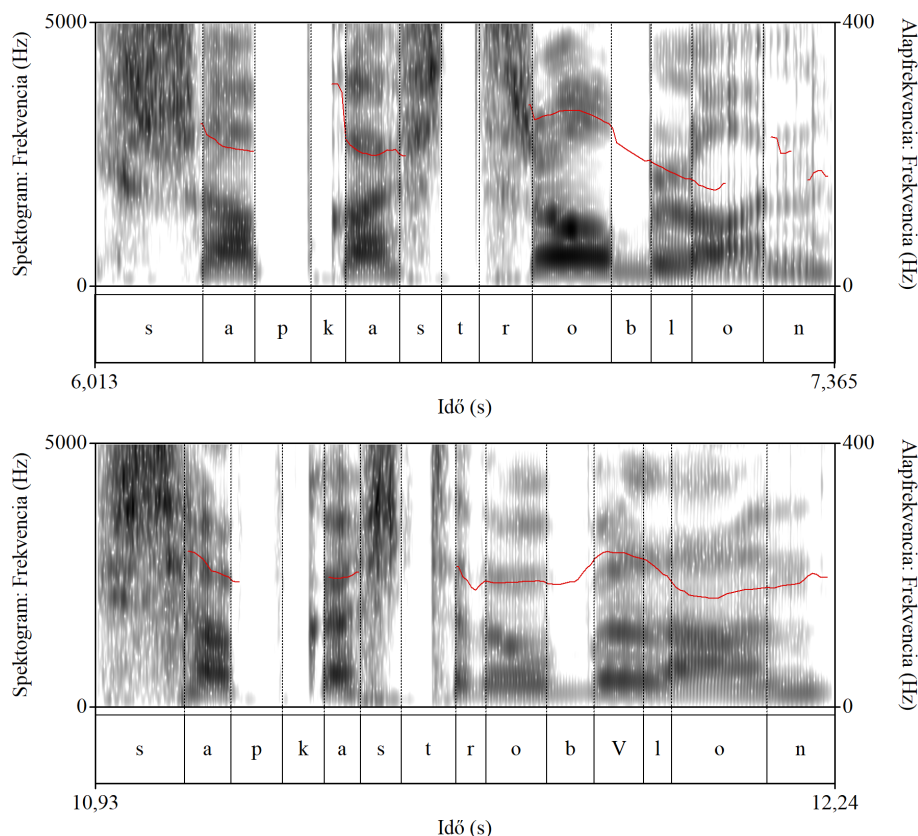
A következőkben a statisztikai elemzés során kapott eredményeinket mutatjuk be a korábban felsorolt hipotézisekre vonatkoztatva.

3.1. *H1 – A tövek szótagszáma*

Az 1. hipotézisünk arra vonatkozott, hogy a beszélők két szótaggal ejtik a szillabikus l -t tartalmazó szavakat. A toldalékolatlan alakok esetében valóban azt találtuk, hogy eldöntendő kérdés végén a dallamcsúcs három beszélő egyetlen szavától (*Strobl*) eltekintve mindahányszor a tő első magánhangzójára került, így 98,21%-os arányt kaptunk. Mivel a tő első magánhangzója a frázis utolsó előtti magánhangzója, ebből következik, hogy utána még egy szótagnak állnia kell, ez pedig nem lehet más, mint az l -t tartalmazó szótag. Így pedig arra jutunk, hogy a beszélők két szótagban ejtették a célszavakat. (Arról, hogy az l -t tartalmazó szótagban az l szótagmag-e, a következőkben még szót fogunk ejteni.) A 2. ábrán a különböző szótagszámú ejtéstípusokat mutatjuk be.

A *-hOz* vagy *-On* toldalékot viselő, kérdő mondat végén elhelyezkedő célszavak között ez az arány 97,41%-os, a 309 célszóból 301-et ejtettek három szótaggal, tehát kétszótagos tövel a beszélők. A nem ekképp ejtett nyolc megfigyelés mind a *Strobl*-nak vagy az *Ybl*-nek valamely alakja. Érdekes látni, hogy míg éppen

a *Strobl* és az *Ybl* azok, amelyeknek nem létezik *i*-betoldással javított alakjuk (védett lexikai réteghez tartozván), addig éppen ezt a kettőt ejtik potenciálisan egyszótagú tövel a beszélők.



2. ábra. A *Vajon a kopaszsága miatt volt mindig sapka Stroblon?* mondat *Stroblon* szavában felfedezhető intonációs különbség a szó két-, illetve háromszótagos ejtése esetén.

Az alsó spektrogramon jól látható egy V jelenléte a *b* és az *l* között (toldalékkal együtt háromszótagos alak!), szemben a föntivel. Másrészt az is kitűnik, hogy az alsón a dallamcsúcs a megjelenő V-n van, míg a felső ábrán az *o*-n – éppen ebből következtettünk arra, hogy előbbinél a célszó maga (*Strobl*) kétszótagos. (A felső ábra *n*-jénél található frekvenciaértékek az irreguláris zöngé miatt félremért értékek, nem valósak.)

3.2. H2 – A magánhangzó-betoldás megjelenése és időtartama

A 2. hipotézis szerint a tövek második szótagjainak szótagmagja variálódik, előfordul szillabikus *l* és betoldott redukált magánhangzó is. A hipotézis

igazolásához definiáltuk a *betoldási arányt*:

$$v = \frac{\text{V-s szóalakok}}{\text{összes szóalak}} \cdot 100 \quad (1)$$

Ez a szám kérdéseknél nagyobb, mint kijelentéseknél ($v_{\text{kérdés}} > v_{\text{kijelentés}}$), illetve a superessivusi toldalék (*-On*) előtt nagyobb, mint az allativus toldaléka (*-hOz*) és a tőalak esetében ($v_{\text{On}} > v_{\text{hOz}} > v_{\text{tőalak}}$) – a pontos arányokért ld. az alábbi 4. táblázatot.

	eldöntendő kérdésben	kijelentésben	összesen
toldalékolatlan alakban	37,80%	27,88%	32,83%
<i>-hOz</i> előtt	47,06%	45,28%	46,15%
<i>-On</i> előtt	77,93%	68,92%	73,38%
az összes alakban együtt	53,46%	46,61%	50%

4. táblázat. Betoldott V-k aránya toldalék- és mondatfípusonként

Vagyis egyrészt megállapíthatjuk, hogy eldöntendő kérdés végén gyakrabban jelenik meg betoldott magánhangzó, mint kijelentésekben. Mielőtt még azonban ezzel összefüggésben megvizsgálánk a betoldás meglétének kondicionáltságát elemző binomiális GLMM átlagos marginális hatásait, bemutatjuk, mi motiválja, hogy redukálnak tekintsük a betoldott magánhangzót.

Két lehetséges szempontból tudjuk vizsgálni, hogy miért tekintjük redukálnak a betoldást. Az egyik a szegmentum hosszúsága, ugyanakkor nem rendelkezünk megfelelő teljes stabil magánhangzóval, mivel a toldalék-magánhangzóknak csak a formánsait mértük, tartamukat nem. Elképzelhető volna az is, hogy a V-k tartamát az *l*-ekével hasonlítsuk össze, hiszen ha egy *li.dV.lön* típusú szótag-szerkezet feltételezése esetén (ld. az 3.5.3 alszakaszt) a szótagkezdet pozícióját elfoglaló *l* hosszabb lenne, mint az öt megelőző szótagmag V, az erős indok lehetne arra nézvést, hogy a V-t redukáltként írjuk le. Ugyanakkor amint a függelékbeli 5.1.1. ábráról kitűnik, a V és az öt követő *l* tartama közel azonos, toldalékolt alakokban még valamivel hosszabb is a V az öt követő *l*-nél.

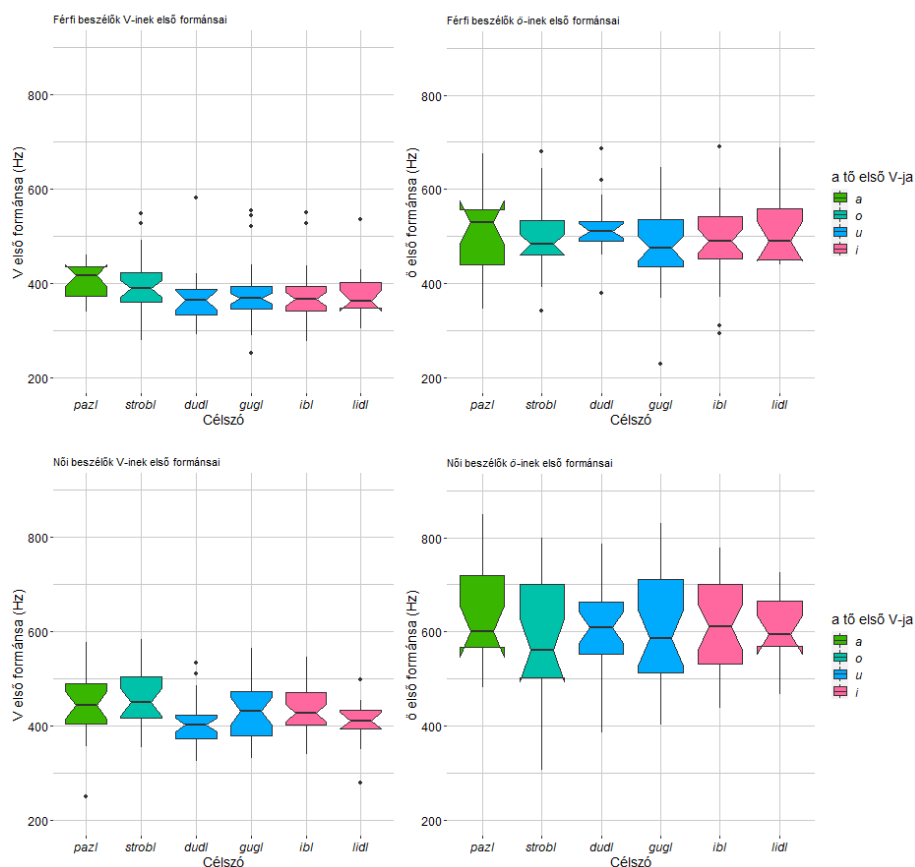
A másik lehetséges szempont a V formánsszerkezetének összevetése a toldalék-magánhangzókéval (utóbbiak esetén az annotált szakaszuk középső 40%-án mért formánsok mediánjai állnak rendelkezésünkre). A 3. és a 4. ábra az elülső toldalékalternánst tartalmazó szavakban szereplő betoldásoknak és a betoldást tartalmazó, elülsőként harmonizáló szavak toldalék-magánhangzójának rendre első és második formánsát ábrázolja a beszélők neme szerint szétválasztva. Az, hogy egy boxplotban vagy csak az elülső, vagy csak a hátsó toldalék-magánhangzó szerepelhet, triviális, viszont azért döntöttünk amellett, hogy ezzel párhuzamosan csak az adott alternánst tartalmazó szavak betoldásait ábrázoljuk, mert potenciálisan elképzelhető, hogy a betoldás minősége befolyásolja a toldalékalternáns megválasztását (hogy a betoldás jelenlétének van-e befolyása, arról ld. 3.4.2-et; a betoldás minőségének harmóniára való befolyását ugyanakkor jelen tanulmány nem vizsgálja). Mindazonáltal csupán néhány tendencia leírására szorítkozunk a kis számú megfigyelés miatt.⁴ Ennek érdekében ún. bemetszett (*notched*) boxplotokat alkalmaztunk, mivel ezekről leolvasható, ha két csoport mediánja szignifikánsan különbözik (ezt az jelzi, ha két boxploton a bemetszések egyáltalán nem fednek át).

Az első formánsokat illetően (3. ábra) a V variabilitása jóval kisebb az egyes célszavakon belül, mint az elülső magánhangzó-alternánsé, vagyis az *ő*-é, a formánsátlagok célszavak között viszont *mindkét* esetben mérsékelt variációt mutatnak. A második formánsokat illetően (4. ábra) viszont épp ellenkezőleg azt látjuk, hogy az egy-egy célszón belüli variabilitás a két csoporton belül egészen hasonló (az első formánsokhoz képest mindenképp), ám a formáns(átlag)ok célszavak közti variációja jóval feltűnőbb a V esetében, mint az *ő*-ében.

Ráadásul a V mediánkülönbségei épp a harmónia szempontjából jelentékenyebb második formáns tekintetében öltenek nagyobb mértéket. Jelentékenyebb, hiszen ez indikálja egy magánhangzó előlségét és kerektségét, ellentétben a magassággal kapcsolatban álló első formánssal. Vagyis az *ő* második formánsának

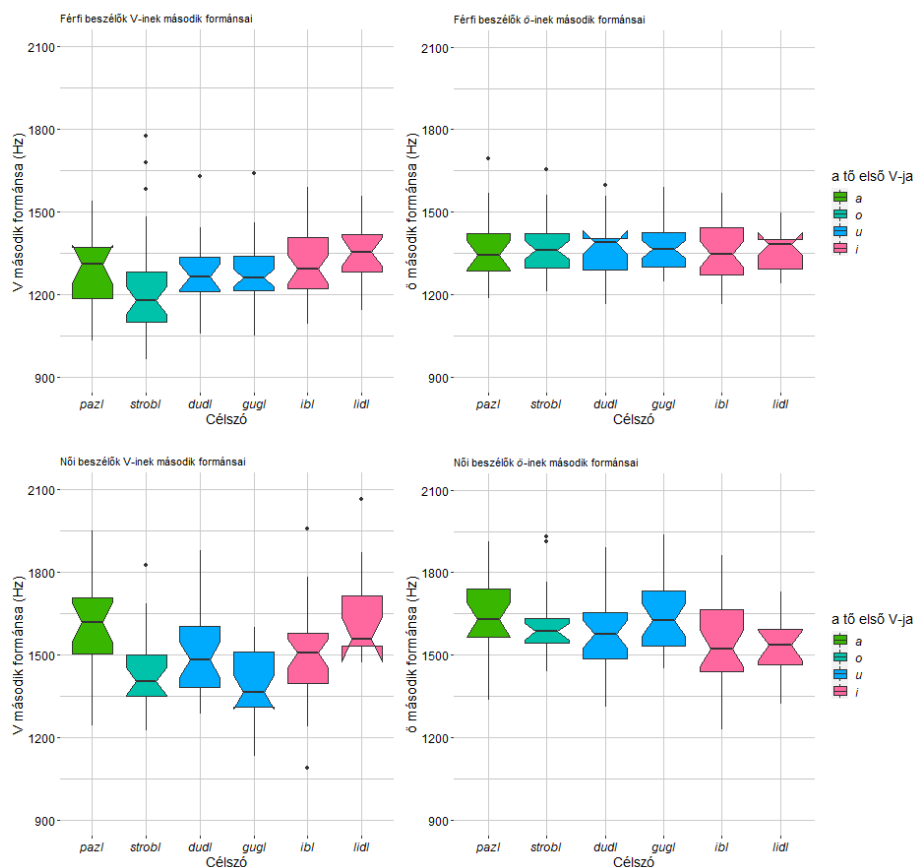
⁴Hátsó toldalék-magánhangzót tartalmazó alakból jelentősen kevesebb fordult elő – ld. a 7. táblázatot –, a kapcsolódó boxplotok ennél fogva csak a Függelékben kaptak helyet, ld. az ottani 2.1.1. és 2.1.2. ábrát.

és ezáltal harmonikus viselkedésének stabil voltához képest a V második formán-
sa igen változatos képet mutat. Bár az itt ábrázolt V-k mind *ö*-t tartalmazó
toldalékot váltottak ki, ez a változatosság a magánhangzó instabilitására utal, és
afelé mutathat, hogy – talán meglepő módon, de – az *l*-nek is szerepe lehet a
harmónia meghatározásában (ld. még 3.4.2 konklúzióját). E miatt az említett
instabilitás miatt hívjuk a V-t redukált magánhangzónak.



3. ábra. A toldalék-*ö* esetén szereplő betoldás (V) és a betoldás esetén szereplő toldalék-*ö* első formánsai célszavanként és nemenként. Fent: férfi beszélők formánsértékei, lent: női beszélők formánsértékei; bal: V-formánsok, jobb: *ö*-formánsok

A felső kettő, férfi beszélőket tükröző ábrákon célszavanként rendre 18, 32, 18, 41, 34, 17 megfigyelést, az alsó kettő, női beszélőket tükröző ábrákon célszavanként rendre 21, 22, 30, 27, 32, 14 megfigyelést tartalmaznak a boxplotok.



4. ábra. A toldalék-ö esetén szereplő betoldás és a betoldás esetén szereplő toldalék-ö második formánsai célszavanként és nemenként. Fent: férfi beszélők formánsértékei, lent: női beszélők formánsértékei; bal: V-formánsok, jobb: ö-formánsok

A felső kettő, férfi beszélőket tükröző ábrákon célszavanként rendre 18, 32, 18, 41, 34, 17 megfigyelést, az alsó kettő, női beszélőket tükröző ábrákon célszavanként rendre 21, 22, 30, 27, 32, 14 megfigyelést tartalmaznak a boxplotok.

3.2.1. A magánhangzó-betoldás megjelenése

Az átlagos marginális hatások vizsgálata előtt röviden kitérnénk a modell illeszkedési jóságára. A marginális pszeudo- R^2 -e 0,2065, a kondicionális pedig kb. kétharmadával nagyobb (0,3409), mint a marginális (vagyis a csak a fix

hatásokat figyelembe vevő) változat. Tehát a random hatások a függő változó varianciájának további kb. 13%-át képesek magyarázni.⁵

A modell átlagos marginális hatásait és az ezekhez tartozó p-értékeket a Függelék 6.1. táblázata tartalmazza. A független változók közül először is szignifikánsnak bizonyult a kérdés-változó átlagos marginális hatása ($p = 0,0205$). A modellben a változó referenciakategóriája a kérdés érték volt, így az együtt-hatója a kijelentésre vonatkozik, ez pedig $-0,0848$, vagyis egy kijelentés esetén átlagosan 8,48%-ponttal kisebb a valószínűsége a betoldásnak (ez egybevág a 4. táblázatban látott tendenciával, amely szerint kérdés esetén gyakrabban fordul elő betoldás). Ezt a tendenciát a következőképpen magyarázhatjuk.

Ahogy már korábban is megjegyeztük, az eldöntendő kérdés intonációja az utolsó előtti szótagig emelkedik, ott a legmagasabb (feltéve, hogy az utolsó hangsúlyos prozódiai frázis legalább három szótag hosszú, de ez a célmondatok mindegyikére teljesül). Továbbá tudjuk, hogy az intonáció magassága (jelen esetben egy kvázi magas tónus) a szótagmagon érzékelhető leginkább, illetve azt is, hogy egy magánhangzó előnyösebb szótagmagként, mint egy mássalhangzó (jelen esetben az *l*). Tehát az intonációjelölés kényszere kérdésekben jobban motiválja a magánhangzó-betoldást, mint kijelentő mondatokban, ahol nem szükséges valamely specifikus szótag (illetve annak szótagmagja) intonációjának kiemeltebb érzékeltetése.

Szignifikánsnak bizonyult a toldaléktípus-változó is. Mivel ez része volt az interakciós hatásnak is, ezért közvetlenül az egyes értékpárok közti szignifikanciát tekintjük a becsült határátlagok (*estimated marginal means*, a továbbiakban *EMM-ek*) módszerének **emmeans** package-beli megvalósítása segítségével. Ezek a post-hoc tesztek azt mutatják, hogy mindkét toldalékolt alakban átlagosan szignifikánsan (ld. a Függelék 6.2. táblázatát) különbözik a V-betoldás valószínűsége a tőalakitól, a *-hOz*-zal ellátott alakokban 18,72%-ponttal, az *-On*-nal ellátottakban 49,2%-ponttal valószínűbb a betoldás. A két toldalékolt alak típus

⁵A pszeudo- R^2 -ek meghatározásához a **MuMIn** package **r.squaredGLMM** függvényét (illetve segédletként az **afex** R-package *mixed* függvényét) használtuk.

V-betoldási valószínűsége is szignifikánsan különbözik: az előzőleg megadott két érték különbségével, vagyis 30,48%-ponttal valószínűbb a betoldás egy *-On*-nal ellátott alakban. (Az egyes relatív gyakoriságokért ld. a 4. táblázatot.)

Az előzőek magyarázataként az *-On* és a *-hOz* közötti különbséget a hiátus-kerüléssel tudjuk indokolni. Mint az 1. hipotézis kapcsán kapott eredményeink mutatták, az *l* a tövek második szótagjában valósul meg (legalábbis a kérdés mondatokban bizonyosan). Épp ezért ha a vizsgált szavakban az *l* szótagalkotó lenne (amennyiben nincs betoldás, ez szükségszerűen így van), és az őt tartalmazó szótagot egy magánhangzó-kezdetű szótag követi, akkor két szótagmag (*N*) kerül egymás mellé, hiszen előbbiben nincsen kóda (*Co*), utóbbiban nincsen szótagkezdet (*O*), vagyis hiátus alakul ki, ami a szótagszerkezetek szempontjából hátrányos, a nyelvek törekednek ennek elkerülésére, ld. (2)-t. (Kérdéses lehet ugyanakkor, hogy vajon egy mássalhangzó-kapcsolat feloldása nem előnyösebb legalább annyira, mint egy hiátus feloldása, és így a *-hOz* esetén való betoldás legalább annyira, mint az *-On* esetén történő.) Ellenben ha a beszélő beszúr egy – akár csak – redukált magánhangzót is, a hiátus megszűnik, hiszen ebben az esetben a betoldott magánhangzó lesz a szótagmag, és mivel a következő szótagnak nincs szótagkezdete, az *l* átszótagolódik ennek a helyébe (mivel előnyösebb itt, mint kódában),⁶ így pedig a két szótagmag között már találunk egy harmadik hangot, tehát a hiátus megszűnt, ld. (3).⁷

$$\sigma(O,N) + \sigma(N,Co) = CL + On \rightarrow \text{hiátus} \quad (2)$$

$$\sigma(O,N) + \sigma(O,N,Co) = CV + lOn \rightarrow \text{nincs hiátus} \quad (3)$$

⁶Már csak ennek megerősítésére is érdemes volna megvizsgálni az *l*-vokalizációt: ugyanis ez csak kódában megy végbe. Így ha azt találjuk, hogy a betoldott alakokban *-On* esetén nincs egyetlen vokalizált *l* sem, míg *-hOz* esetén igen, az erős indikátora lenne annak, melyik esetben tölti be a szótagalkotó szerepét az *l*.

⁷Itt nem vetettük föl annak a lehetőségét, hogy az *l* előtti mássalhangzó a tő első szótagjának kódájába kerüljön, míg az *l* mint szótagkezdet és az *-On* együtt alkotja a következő szótagot. Ezt az eshetőséget másokkal egyetemben az 3.5.3. alszakaszban fejtjük ki.

Arra is ki kell térnünk, hogy a tőalak esetén való betoldás gyakorisága mindkét toldalék betoldási arányánál kisebb. Ez megerősíteni látszik Jurgec & Bjorkman (2018) azon megállapítását, amely szerint a toldalékolatlan szavak gyakrabban kivételesek (kevésbé adaptálják őket a beszélők), vagyis a kölcsönszavakat nagyobb mértékben adaptáljuk toldalékolás során. Mindenesetre ez csak tendenciaként fogalmazható meg, továbbá az is kérdéses, hogy vajon kivételesebb-e annyival egy szillabikus mássalhangzó, mint egy betoldott redukált magánhangzó, hogy az előbbi megállapítás megálljon.

A célszó-változó kategorikus lévén a modell kiértékelése során dummyváltozókra bomlik (miképp az előzőek is), így meg tudjuk figyelni, hogy mely értéke esetén tapasztalunk szignifikáns eltérést a referenciakategória (*Doodle*) esetén vett betoldási valószínűségtől, és hogy az milyen mértékű. Viszont mivel a többi közti szignifikáns eltéréseket így nem nyerjük ki, és mivel a modellben interakciók is szerepelnek, a post-hoc tesztek eredményeit vizsgáljuk. Azt látjuk, hogy a *Google* és a *Lidl* minden mástól és egymástól is szignifikánsan különbözik. A *Google*-ban minden más vizsgált szónál gyakoribb a betoldás, a legkevésbé a *Strobl*-tól tér el (átlagosan 17,66%-ponttal), a leginkább a *Lidl*-tól (átlagosan 51,56%-ponttal), a többi eltérésért ld. a függelékbeli 6.2., illetve a pontos gyakoriságokért a 5. táblázatot. Ez arra utalhat, hogy a szó ebben a betoldással ellátott alakban lexikalizálódott.⁸

A *Lidl*-ben viszont minden más vizsgált szónál alacsonyabb a betoldási valószínűség. A legkevésbé a *puzzle*-tól tér el (átlagosan 17,59%-ponttal kevésbé valószínű a betoldás), a leginkább a *Google*-tól. Hogy mi az oka annak, hogy míg a *Google* inkább betoldással, addig a *Lidl* inkább betoldás nélkül lexikalizálódhatott, kérdéses. Előfordulhat, hogy a *Lidl* inkább az *i*-betoldott [lidli] vagy *l*-csonkolt [lid] alakban jelenik meg a hétköznapi beszédben, és ezért a produkált alakok jórészt a betűejtésnek köszönhetőek. Így egy olyan használt alakot, ahol a

⁸Az is a jelenség mögött állhat – de ennek elemzését jelen tanulmány nem tartalmazza –, hogy a *Doodle*-ban szereplő *Cl*-kapcsolat homorgán, míg *Google*-ban nem, és ennek okán maguk az artikulációs gesztusok közti átmenet eredményezhetné a betoldást. Köszönjük az egyik anonim lektornak ezt az észrevételét!

C és az l között nem szerepel új szegmentum, jóval könnyebben fölülír az írásképp, mint egy olyat, ahol már szerepel egy új szegmentum (ld. a *Google* betoldással ellátott alakjának erőteljes használatát). Szerepet játszhat ugyanakkor az átadó nyelv különbsége is.

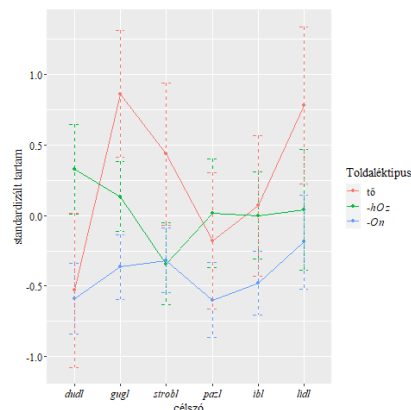
3.2.2. A magánhangzó-betoldás időtartama

Térjünk át a 2. hipotézis kapcsán fölállított másik modellünkre! A V-betoldás standardizált tartamára futtatott robusztus LMM szignifikáns együtthatóinak vizsgálata előtt röviden kitérünk a modell illeszkedési jóságára: a robusztus modell marginális pszeudo- R^2 -e ugyan nagyobb, mint a nem robusztusé ($0,2200 > 0,2083$), a kondicionális pszeudo- R^2 viszont valamivel kisebb a robusztus verzió esetén ($0,4236 < 0,5098$). Az eltéréstől eltekintve mindenesetre az látszik, hogy a random hatások a független változóbeli variáció további 20-30%-át (a csak a fix hatások által magyarázott 1–1,5-szeresét) magyarázzák, valamint a teljes modell képes ezen variancia (közel) felét magyarázni.

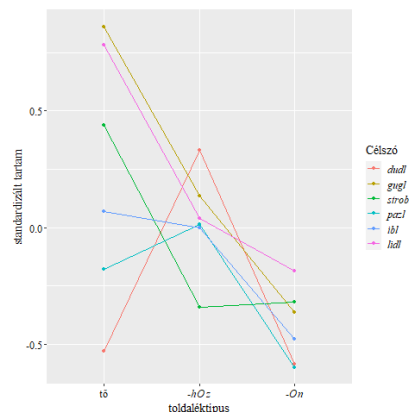
Azt találtuk, hogy mind a három dummyváltozó-csoport (toldaléktípus, célszó, mondat típus) külön-külön szignifikáns. Ezeken belül a következőket találjuk (a részletes eredményeket ld. a Függelék 6.3. táblázatában). A mondat típus-változó referenciakategóriája újfent a kérdés-érték volt, egyúttal azt is látjuk, hogy a kijelentés-dummy szignifikáns ($p < 0,001$). Ennek együtthatója $-0,2275$, vagyis amennyiben a célszó kijelentő mondatban szerepelt, átlagosan a V-tartam szórásának (28,83 ms) kb. negyedével (6,56 ms-mal) volt rövidebb a V-betoldás. Ezt hasonlóképp a V kérdés esetén vett nagyobb gyakoriságához (ld. az előző modellt) az eldöntendő kérdésbeli intonációjelölés szükségességével magyarázzuk.

A toldaléktípusok közül mind a *-hOz*, mind a *tő* szignifikánsan ($p < 0,001$) különbözőnek mutatkozik az *-On* esetétől a post-hoc tesztek alapján. Előbbi 0,4518 egységnyi, utóbbi 0,6622 egységnyi szórással hosszabb. A *tő* és a *-hOz* betoldáshossza nem különbözik szignifikánsan (ld. a Függelék 6.4. táblázatát). A 5. és az 6. ábrán is jól látszik, hogy az *-On*-nal ellátott alakok betoldása rendre kisebb a *-hOz* és a *tő* esetén megvalósulóknál. Mindazonáltal hozzá kell tennünk, hogy míg az összes vizsgált *-hOz*-os alak 46,15%-a betoldott, addig az

-*On*-os alakok közt ez az arány 73,38% (312, ill. 293 db-ból rendre 144, illetve 215 db). Ellenben ez nem befolyásolja a betoldás meglétét vizsgáló és a jelen modell közti látszólagos paradoxon magyarázhatóságát, amelyre a következő kapcsolódó megfigyelés után ki is térünk.



5. ábra. A betoldott V standardizált tartama toldaléktípusonként



6. ábra. A betoldott V standardizált tartama célszavanként

Egy olyan tendencia is megfigyelhető, hogy a *-hOz* és az *-On* esetén való betoldás között erős negatív korreláció áll fenn: jól látszik a 7. ábrán is, hogy azon beszélők, akik a beszélők átlagához képest átlagosan hosszabb betoldást produkáltak *-hOz* esetén, a beszélők átlagához képest átlagosan rövidebb betoldást produkáltak *-On* esetén (a Pearson-féle korrelációs együttható $-0,6469$). Ezt az motiválhatja, hogy amennyiben a beszélő a betoldást választja javítóstratégiául (pl. az *l* megnyújtása helyett, ld. az 3.5.3. alszakasz eredményeinek kifejtését), egy hármas mássalhangzó-kapcsolat (*VClhV*) feloldásához, így a klaszter tagjai szeparálásának érzékeltetéséhez inkább van szükség egy hosszabb betoldásra, mint egy kettes klaszter esetén (*VClV*). (Mindettől természetesen még nem kerülünk ellentmondásba az előző szakasz megállapításával, amely szerint *-On* esetén szisztematikusan valószínűbb a betoldás, hiszen mint ki is emeltük, ez a jelenség ezen a javítóstratégián (a V-betoldáson) belül értelmezhető.)

A 7. ábráról az is kitűnik, hogy a tövek esetén jóval nagyobb a variáció a beszélőközi átlagtól való eltérésben (illetve alátámasztja azt, hogy pusztán az

-*On*-tól különbözik szignifikánsan: míg a -*hOz* eltéréseivel vett korrelációja erőteljesen pozitív: 0,7210, az -*On* eltéréseivel vett jócskán negatív: -0,5115). Az eltérések varianciája tövek esetén 0,3727, míg a két toldalékos alak esetén 0,0775 (-*hOz*), illetve 0,0503 (-*On*). A beszélőnkénti eltérések szerint vett konfidenciaintervallumok hosszának varianciája is jóval nagyobb a tövek esetén (0,4259 a 0,0136-del és a 0,0084-del szemben). Ez legalább részint azzal magyarázható, hogy a *t*őalakok nemcsak hogy kivétel nélkül a célmondatok utolsó szavaként szerepeltek (ez néhány kijelentő toldalékolt alak esetében nem teljesült), de maga az *l* vagy az azt tartalmalmazó szekvencia is közvetlenül a frázis végén szerepelt (toldalék nem lévén). A frázisvégi nyújtás pedig általános fonetikai tendencia.

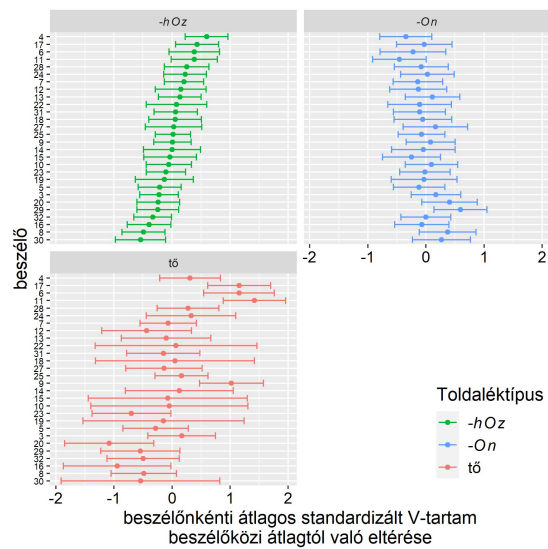
A célszó-dummyk közül, úgy tűnhet, a referenciakategóriától (*Doodle*) egyedül a *Strobl* különbözik szignifikánsan ($p < 0,001$). Viszont mivel több interakciós hatás is szignifikáns, közvetlenül az EMM-eket vizsgáljuk meg (ld. szintén a függelékbeli 6.4. táblázatot). Ezek azt mutatják, hogy a *Google* csak a *Lidl*-től és a *Strobl*-tól nem különbözik szignifikánsan (vö. 6. ábra), míg a másik öt között csupán a *Lidl* a *puzzle*-től és a *Doodle*-tól. (A nem robusztus modellben a *Google* a *Strobl*-tól is szignifikánsan különbözik, viszont a *Lidl* a *Doodle*-tól nem.) Ez, mint a betoldás meglétének vizsgálata során is, arra utalhat, hogy – azáltal, hogy több célszótól különbözik, mint a többi célszó egyenként – a *Google* lexikalizálódottsági foka jóval magasabb a többiénél. Azzal kapcsolatban pedig, hogy a *Lidl*-ben megjelenő betoldás tartama megközelíti a *Google*-ban megjelenőét,⁹ azt mondhatjuk, hogy mindazoknál, akiknél a *Lidl* V-betoldásos (nem pedig *i*-betoldásos vagy éppen csonkolt) alakban lexikalizálódott, ez a lexikalizálódottsági fok magas a többi szóéhoz képest.

Ez nem is különösebben implauzibilis: igen valószínű, hogy a *Lidl* szó általánosabban nagyobb használati gyakoriságú a *Strobl*, az *Ybl*, a *puzzle* és a *Doodle* szavaknál.

⁹Persze a betoldásos *Lidl*-alakok száma és aránya is jóval alacsonyabb, mint a *Google* alakjaié, ld. az 5. táblázatot.

	betoldás (467)	standardizált tartam
Google	71,97% (113)	0,2555
Strobl	59,06% (88)	-0,0855
Ybl	54,09% (86)	-0,0715
Doodle	44,38% (71)	-0,2100
puzzle	42,86% (66)	-0,2199
Lidl	27,74% (43)	0,3308

5. táblázat. Betoldás gyakorisága és a standardizált tartam célszavanként (összes alak)

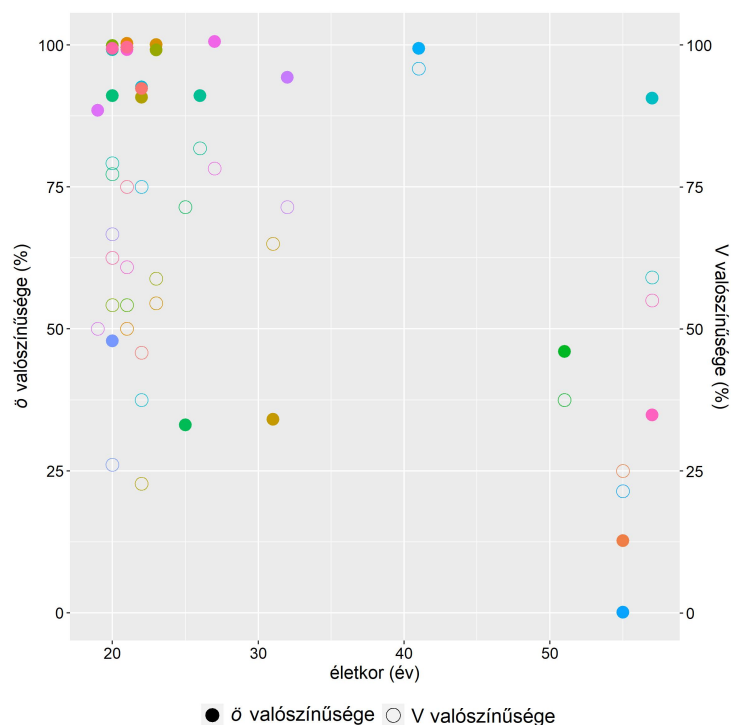


7. ábra. A betoldás standardizált tartamának beszélőnkénti eltérései toldaléktípusonként

3.3. H3 – Az alternáns megválasztása háromalakú toldalékok esetén

A 3. hipotézisünk azt tartalmazta, hogy háromalakú toldalékok esetén az elülső alternáns (-ön, -höz) jelenik meg a vizsgált szótövek után, azzal a megszorítással, hogy első szótagjukban hátsó magánhangzót tartalmazó töveket (a továbbiakban *hátsó* tövek) követhet hátsó magánhangzójú toldalék is, tehát a második szótag az ilyen esetekben átlátszó lehet. Ezek a tövek a kísérletben a *Strobl*, *Doodle*, *Google* és a *puzzle* voltak.

A beszélők konzisztenciájára itt nem térünk ki, viszont annyit röviden elmondhatunk, hogy elkülöníthető egy olyan csoport, amely minden toldalékkolt



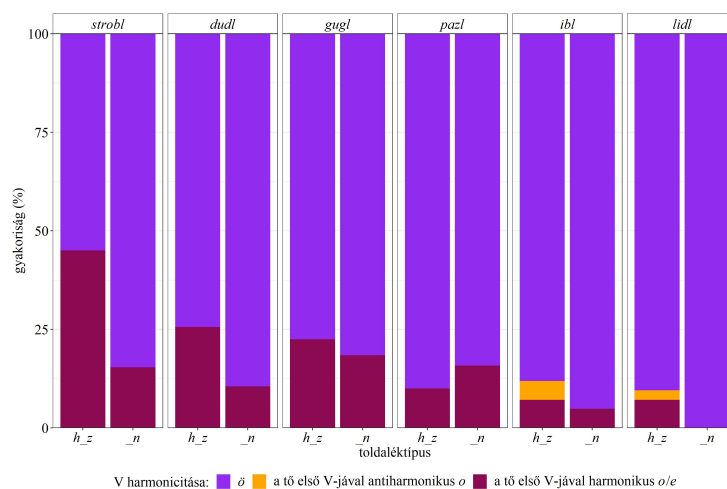
8. ábra. Az elülső alternáns és a betoldás valószínűsége beszélőnként az életkor függvényében. Egy szín egy beszélőt jelöl. Az egymásra eső pontokra jitterelést végeztünk, hogy minimálisan szétszóródjanak a jobb láthatóság érdekében.

alakban az elülső alternánst választotta. Emellett volt egy olyan, amelyikre ez szintén igaz, de annyi megkötéssel, hogy a *Strobl* bizonyos alakjaiban hátsó toldalékot használt, viszont csak ezekben használt ilyet. Illetve egy olyan is, amelyik viszonylag következetesen semleges magánhangzót tartalmazó tövek (a továbbiakban *semleges* tövek) után elülső (kerekítetlen) toldalékot, hátsó tövek után hátsó toldalékot használt.¹⁰

Azt sem hagyhatjuk figyelmen kívül – bár ez nem képezte vizsgálatunk tárgyát, hiszen ahhoz a minta nem volt kellően rétegzett –, hogy az életkor és a

¹⁰A kísérletben nem vizsgáltunk olyan töveket, melyek első magánhangzóként elülső (nem semleges) magánhangzót tartalmaznak (ilyenek lehetnek volna a *Löbl* vagy a *Deep Purple* tulajdonnevek). Ezek viszont a harmóniával kapcsolatban nem szolgáltak volna információval: irreálisnak tartjuk, hogy ezeket a szavakat ne az *ő*-s alternánssal toldalékolják a beszélők, ami viszont nem tesz különbséget a stabil magánhangzó és az *l* között.

beszélő gyerek- vagy fiatalkori lakhelye alapján is látunk eltéréseket. Különösen jellemző a fiatal budapesti beszélőkre a hipotézisünk első fele szerinti viselkedés, tehát hogy a vizsgált szavak minden toldalékos alakjában az elülső alternánst választották. Tendenciaszerűen igaz az is, hogy inkább az idősebb beszélők használtak hátsó toldalékokat hátsó tövek után. Ez megmutatkozik a 8. ábrán is, amin minden beszélő esetén látjuk, hogy milyen valószínűséggel használja az elülső toldalékalternánst, illetve alkalmaz V-betoldást a beszélő. Az ábra alapján úgy tűnik, hogy a hat 40 év fölötti beszélő közül négy esetében a fiatalabb beszélőkhöz képest jóval kevésbé valószínű, hogy az elülső alternánst választja, a maradék kettő viszont hasonló gyakorisággal alkalmazza azt, mint a fiatalabbak. A fiatalabbakra pedig, a kísérlet folyamán kialakult benyomásunkat megerősítve, sokkal inkább jellemző, hogy az elülső toldalékalternánst választják. A betoldás valószínűségét tekintve viszont már hasonló jól leolvasható tendenciák nem tűnnek elő. (Az életkor és a V-betoldás gyakoriságára kiszámolt Kendall-féle τ rendkívül alacsony, csaknem 0 ($-0,0083$). Az $\bar{o}$ és az életkor közti összefüggésre kiszámolt viszont bár nem különösebben magas, de egyértelműen negatív: $-0,3455$.)



9. ábra. Szillabikus *l*-t tartalmazó tövekhez kapcsolt toldalékok harmóniája

A 9. ábráról leolvasható az eredmények gradualitása. A beszélők minden tőnél az esetek nagy többségében az elülső toldalékokat használták. Ezt a

harmonikus viselkedést nem a tövek magánhangzója váltja ki, hanem a szillabikus *l*-t tartalmazó szótag (hogy a szótagon belül pontosan mi, azt az 5. hipotézisnél tárgyaljuk). A 3. hipotézis első fele tehát igazolódott.

A hipotézisben feltételezett opcionális átlátszóság szintén megvalósul, de az eredmények itt tövenként és toldalékonként is eltérőek. Valóban a hátsó töveknél fordulhat elő inkább a *tő* első magánhangzójával harmonizáló toldalék, de némi variáció semleges töveknél is megfigyelhető. Vagyis néhányszor előfordultak a *-hez*, *-en* alternánsok is. Jelentős variáció három hátsó *tő*nél fordul elő (*Strobl*, *Doodle*, *Google*), de a *puzzle* esetén nem ritkább az elülső alternáns, mint az *Ybl*-nél. Az, hogy a *tő* hátsó legyen, szükséges, de nem elégséges feltétele az opcionális átlátszóságnak. Tendenciaszerűen megfigyelhető különbség van a toldalékok között is: a *-hOz*-os alakokban gyakrabban fordul elő hátsó toldalék, mint az *-On*-osokban, vagyis a *Stroblhoz* típusú alakok gyakoribbak, mint a *Stroblon* típusúak. Két *tő* összehasonlításakor előfordulhat, hogy csak az egyik toldalékkal van különbség: a *doodle-hoz* alak arányaiban csaknem háromszor gyakoribb a *puzzle-hoz* alaknál, a *doodle-on* és a *puzzle-on* előfordulása viszont közel azonos.

Megfordítva a fentebbi állítást: *-On* esetén gyakrabban jelenik meg az első szótagok alapján motiválatlannak tűnő elülső toldalék, mint a *-hOz* esetén. Mivel magyarázható ez a különbség? Kézenfekvő magyarázatként szolgálhatna, hogy *-On* előtt (ahogy a 2. hipotézis kapcsán láttuk), gyakoribb a redukált magánhangzó betoldása az *l* elé, mint *-hOz* előtt, tehát a betoldott (és szótagmaggá vált) magánhangzó váltja ki az *-ön*-nel való toldalékolást. Ezt a magyarázatot vizsgáltuk a 4. hipotézis tesztelésekor.

3.4. H_4 – Összefüggés a magánhangzó-betoldás és az elülső alternáns megjelenése között

3.4.1. Betoldási és előlségi arány

Azt láttuk, hogy mindazokat az alakokat, ahol elülső (kerekített) magánhangzót tartalmaz a toldalék, nem magyarázza a *tő* első magánhangzója (a semleges töveknél kerekiségben, a hátsó töveknél előlségben). Tehát felmerül a kérdés,

hogy mi motiválja az *ö*-s toldalékok megjelenését. Erre keresi a választ a 4. hipotézisünk, amely szerint ennek kiváltója az *l* szillabicitását feloldó betoldott redukált magánhangzó. A korábban már definiált betoldási arányt ezúttal csak a toldalékolt alakokra értelmezzük, hiszen a *tő*alakok esetén triviálisan nem beszélhetünk arról, hogy harmonikusan mit vált ki egy esetleges betoldás:

$$v = \frac{\text{V-s toldalékolt szóalakok száma}}{\text{összes toldalékolt szóalak}} \cdot 100 \quad (4)$$

Definiáljuk továbbá a labialitási arányt, mely az elülső toldalékot tartalmazó szóalakok és az összes toldalékolt szóalak százalékos arányát takarja:

$$\phi = \frac{\text{elülső toldalékot tartalmazó szóalakok}}{\text{összes toldalékolt szóalak}} \cdot 100 \quad (5)$$

A 6. táblázatban látható a két arányszám szavankénti bontásban. Összevetve a két arányszámot az egyes szavak esetében, a következő tendenciát találjuk: többnyire jóval nagyobb arányban fordul elő elülső alternáns, mint amekkora arányban betoldás tapasztalható, vagyis általában $v < \phi$. Különösen látványos ez az arány a *Lidl* esetében: szinte háromszor akkora arányban találunk *ö*-s toldalékot, mint betoldást ($v_{\text{Lidl}} = 31,63\% < \phi_{\text{Lidl}} = 96,94\%$), de a *puzzle* esetében is csaknem kétszeres az arány, ahogyan a 6. táblázat mutatja. Mivel az elülső toldalék olyankor is megjelenik, amikor nincs betoldott magánhangzó, az utóbbi megléte nem magyarázhatja az előbbiét.

Külön is felhívjuk a figyelmet a *Google* esetére: jól láthatóan ennél a legnagyobb a betoldási arány, csak nagyjából minden hatodik alkalommal nem toldottak be a beszélők magánhangzót az *l* elé. (A 28-ból 23 beszélő legalább fele alkalommal (6-ból 3-szor) ejtett betoldott magánhangzót is az *l* előtt, tizenkettőjük ezen belül is legalább 5 alkalommal.) Mint ahogy már fentebb is többször utaltunk erre, ez arra enged következtetni, hogy a *Google* betoldást tartalmazó alakja jóval lexikalizáltabb, mint a többié. Viszont újra felmerül a kérdés, hogy

	v (betoldás) (358)		ϕ (harmónia) (507)
Doodle	55,24% (58)	<	80,00% (84)
Google	84,16% (85)	>	76,24% (77)
Strobl	71,72% (71)	$\approx$	68,69% (68)
puzzle	46,94% (46)	$\ll$	88,78% (87)
Ybl	66,34% (67)	<	95,05% (96)
Lidl	31,63% (31)	$\ll$	96,94% (95)

6. táblázat. Betoldott V-k és elülső kerekített toldalékok aránya szótövenként, zárójelben az abszolút gyakoriságokkal

fonetikailag-fonológiaiag hogyan jellemezhetők a betoldott magánhangzók, ha legalább kb. 8%-nyi esetben láthatóan átlátszóak voltak a *Google* esetében is, illetve ha feltételezzük, hogy egyébként elülső alternánsok kiváltói. (Hiszen ennyivel (8 százalékpont) többször látunk betoldásos alakot, mint amekkora arányban az elvileg általa kiváltott elülső alternánst.)

3.4.2. A betoldás és az elülső alternáns függetlensége

A kérdést formálisabban megvizsgálva, a korábban már vázolt, az $\bar{o}$ megjelenésének, valamint a betoldásnak és a célszónak az összefüggésére specifikált binomiális GLMM eredményeül a következőt kaptuk. Egyetlen változó átlagos marginális hatása sem szignifikáns (ld. a Függelék 6.5. táblázatát), kivéve a *Lidl* és a betoldás meglétének interakcióját ($p < 0,001$; ez nem véletlen, hiszen a 6. táblázatból is jól látható, hogy a *Lidl* toldalékolt alakjai közt jelentősen kevesebb a betoldás nélküli alak, mint a többi célszónál). Az egész modell egybevéve azonban szignifikáns (közös hipotézissel megvizsgálva a változók β -értékeinek (amely a független változó logit-értékeiben való változást adja meg) egy-egy párját, a legtöbb párra el tudjuk utasítani, hogy ne lenne szignifikáns). A post-hoc tesztek is ugyanezt mutatják (ld. a Függelék 6.6. táblázatát). Mindezt akképpen interpretálhatjuk, hogy egyrészt valamennyi célszónál olyan mértékű az elülső toldalékos alakok aránya, hogy ebben nincs közöttük számottevő különbség, más-

részt nincs különösebb jelentősége az elülső toldalék megjelenésének valószínűsége szempontjából annak sem, hogy történt-e az adott alakban betoldás.¹¹

A modell illeszkedési jóságát tekintve azt látjuk, hogy a marginális pszeudo- R^2 alacsony (0,0874), míg a kondicionális pszeudo- R^2 relatíve magas (0,6178), ami azt mutatja, hogy a random hatások jóval több variációt magyaráznak a függő változóból, mint amennyit a fix hatások, sőt: abszolút értékben tekintve is a variáció több mint felét képesek magyarázni utóbbiak. A marginális érték alacsony volta egyfelől jól leképezi azt, hogy a modellben szereplő változók egyesével szinte kivétel nélkül nem szignifikánsak, a kondicionális érték (ehhez képesti) nagysága pedig indikátora annak, hogy az $\bar{o}$ megjelenése nagyban, sőt nagyjából beszélőfüggő! (Ennek köze lehet az előző hipotézis eredményeinek bemutatásánál megemlített, az adatok kiegyensúlyozatlansága miatt csupán gyenge tendenciához is, amely a beszélő életkorára és az $\bar{o}$ megjelenési valószínűségére vonatkozott.)

Egy némileg egyszerűbb mérőszámot tekintve: annyit ugyan megállapíthatunk az elülső toldalék megjelenésének feltételes valószínűségei alapján, hogy betoldás esetén (éppen hogy) valamivel nagyobb arányban (85,47%) fordul elő ilyen toldalék, mint a betoldás nélküli alakokban (82,38%), de az $\bar{o}$ -t tartalmazó alakoknak is csupán 60,36%-ában található betoldás. Vagyis függni nem függ az $\bar{o}$ választása a betoldás jelenlététől. A betoldás jelenlétének és a harmóniának a függetlenségét (H_0) az összes toldalékolt alakra (ld. 7. táblázat) χ^2 -próbával tesztelve 5%-os szignifikanciaszinten nem tudjuk elutasítani (p -érték: 0,363), ezenfelül a korrigált kontingenciakoefficiens is alacsony ($C^* = 0,0524$). A Cohen-féle w hatásnagyság ugyancsak alacsony hatást mutat ($w = 0,0371$) – ami

¹¹Ha a binomiális GLMM-et célszavanként futtatjuk le, akkor a *Strobl* esetében a betoldás jelenlététől mégiscsak szignifikánsan ($p = 0,0151$) függ az $\bar{o}$ megjelenése. Ez feltehetőleg azzal magyarázható, hogy egyedül a *Strobl* esetében szerepel jelentős számú nem $\bar{o}$ -t tartalmazó toldalék az adatok közt. Az pedig, hogy a bemutatott modellben miért nem szignifikáns mégsem, múlhat azon is, hogy a célszó-változó önálló dummyja mellett egy adott célszó a betoldás meglétének dummyjával is interakcióban van, ami képes befolyásolni az egyes becsült értékek szignifikanciáját.

még jobban megerősíti az eredményt, hiszen ez utóbbi nem érzékeny a minta nagyságára, míg a χ^2 -próba igen.

	ő a toldalék	nem ő a toldalék
van betoldás	306	52
nincs betoldás	201	43

7. táblázat. Betoldás és ő-s toldalék kapcsolata

Tehát a 4. hipotézist nem erősítették meg az adatok: gyakran előfordult, hogy betoldás nélkül is az ő-s toldalékot választották a beszélők, holott ezt a tő első szótagjának magánhangzója sem indokolja. Ennek magyarázatában viszont egy másik felvetődő tényező lehet az, hogy az l formánsszerkezete hasonlít az ő-éhez. Ez a hasonlóság megerősítést nyert saját adataink alapján is. Látható, hogy a Markó (2017) szerinti adatokhoz (ld. a 8. táblázatban) közel esnek az általunk mért formánsértékek is (ld. a 10a–10d. ábrákat). Ugyanakkor azt csak további, percepciós kísérletek erősíthetik meg, hogy valóban az l -e a kiváltója az elülső toldalékoknak (vagy pedig az l (egy részének) esetleges elülsőként való percepciója).

	ő	l
F ₁ (Hz)	470 (ffi) / 520 (nő)	300-400
F ₂ (Hz)	1300 (ffi) / 1550 (nő)	1300-1600

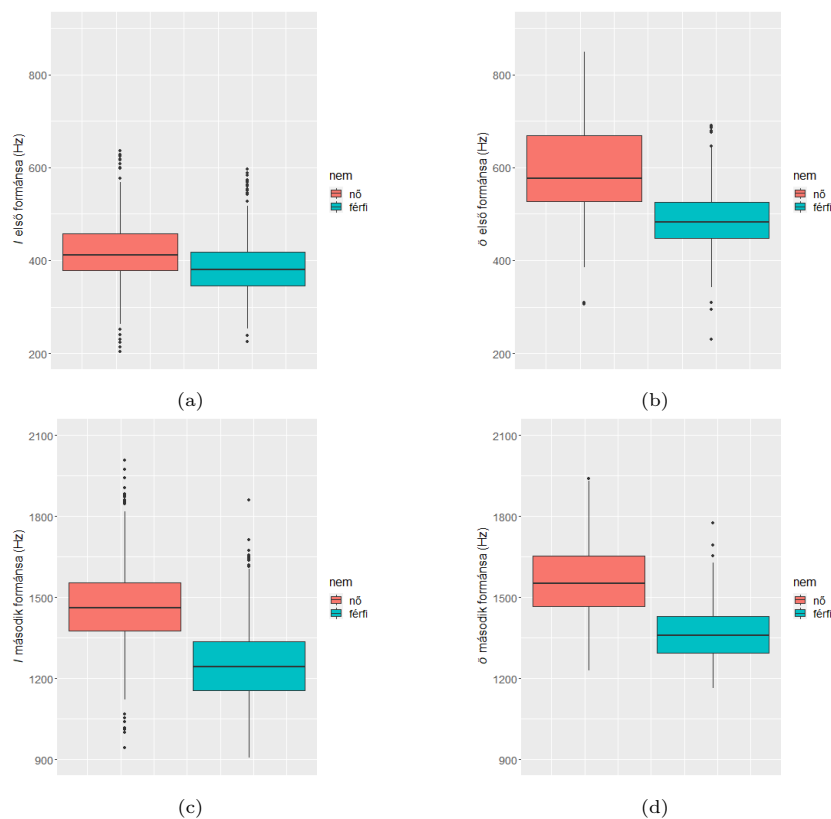
8. táblázat. A toldalék ő és az l formánsai Markó (2017) alapján

3.5. H5 – Variabilitás az l időtartamában

Az 5. hipotézis kapcsán rátérünk az l tartamának és szórásának, illetve a VI szekvencia tartamának az elemzésére.

3.5.1. Az l -tartam

Az első modellben az l tartamát vizsgáltuk. Mivel a négy változóból három interakciókban is részt vesz, ezért egyből a post-hoc tesztek (EMM-ek) szignifikanciáját érdemes tekinteni (az összes együtthatóért ld. a Függelék 6.7., a



10. ábra. Az *ö* és az *l* formánsai

post-hoc tesztek eredményéért a Függelék 6.8. táblázatát). Itt azt látjuk először is, hogy betoldás-változó két értékét összehasonlítva az *l*-tartam szignifikánsan különbözik ($p < 0,001$). Betoldás esetén az *l* tartama átlagosan 1,2172 szórással rövidebb volt, mint a betoldás nélküli alakokban. Ez arra mutathat, hogy a betoldás mintha kiegészítené az *l*-et olyan mértékig, mint amikor nincs betoldás.

A toldaléktípus-dummyk közül a *tő* mind a robusztus, mind a nem robusztus verzióban igen szignifikánsan ($p < 0,001$) különbözik a *-hOz*-tól és az *-On*-tól is, előbbinél átlagosan 0,6816 egységnyi szórással, utóbbinál átlagosan 0,8531 egységnyi szórással hosszabb. Ez a tény már csak az általános frázisvégi nyúlásból is következhet. A robusztus modellben a két toldalékolt alak *l*-tartama is különbözik egymástól, igaz jóval kevésbé szignifikánsan ($p = 0,0439$). Eszerint

egy allativusi alakban az l átlagosan 0,1714 szórásnival hosszabb. Ez főképp arra lehet visszavezethető, hogy *-On* esetén jóval gyakoribb a betoldás, és mivel az előbb láttuk, hogy betoldás esetén jóval rövidebb az l , így ez az *-On* tekintetében is igaz lehet. (Ezzel áll összefüggésben az is, hogy az l egy része mindenképp szótagkezdetbe kerül (ld. erről a 3.5.3. alszakaszt), ahol tipikusan rövidebb, mint egyebütt a szótagban.)

A célszó-dummyk EMM-jei alapján azt látjuk, hogy a *Doodle* minden más célszó l -tartamától szignifikánsan különbözik, de a többi célszóé egymás közt nem. A *Doodle*-beli l -tartam a többi szóénál 0,1347 és 0,2539 egység közötti szórással hosszabb. A *Doodle* többbitől való erőteljes elkülönülését motiválni is tudjuk. Ugyanis több beszélő is jelezte utólag, hogy ők ezt a szót (vagyis az alkalmazást, amit jelöl) nem ismerték, így elképzelhető, hogy az ismeretlenség okozta bizonytalanság egy hosszabb l -tartamban nyilvánul meg.

A kérdés-változó is szignifikánsnak bizonyult ($p < 0,001$), az együtthatója $-0,1043$, tehát a két kategória (kérdés és kijelentés) együttes szórásának kb. egytizedével rövidebbek a kijelentésbeli l -ek. Ezt az 3.2.2. modell eredménye-
ihez (V-betoldás tartama) hasonlóan tudjuk értékelni: annak a szótagnak az intonációs jelentősége, amely az l -t tartalmazza, magával vonja azt is, hogy a szegmentum is, amin megjelenik, kiemeltebb legyen. Bár az intonációs kiemelés (magas tónus) az adott szótag szótagmagjára vonatkozik, tudjuk, hogy amennyiben nincs V-betoldás, az intonáció elhelyezéséből következően az l -nek kell képeznie a (ha van – rag előtti) utolsó szótag magját. Tehát akár ő, akár a V a szótagmag, a beszélők a magas tónus könnyebb érzékeltetésére valamelyest megnyújtják az adott szegmentumot. Sőt, ha van V-betoldás, az l akkor is hosszabb kérdésben, mint kijelentésben, vagyis annak ellenére, hogy elméletben be sem tölti a szótagmag pozícióját.

Azt, hogy ez a nyújtás szignifikáns-e a teljes Vl -szekvencia esetén is, illetve van-e különbség annak hosszában aszerint, hogy van-e V-betoldás vagy nincs, azt a következő alszakaszban mutatjuk be.

A modell illeszkedési jósága igen nagy: a marginális pseudo- R^2 0,7519, a kondicionális pedig 0,8457, ami azt jelenti, hogy a független változók a modell

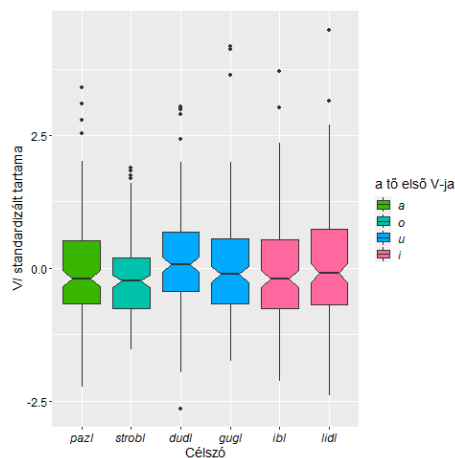
egy tekintélyes részét képesek magyarázni, ebből kb. 9%-pontnyit a random hatások, vagyis a beszélőközi variáció. (A nem robusztus modellben is igen hasonlóra jutunk: 0,6929, ill. 0,8220.) Itt emlékeztetnénk arra, hogy a V-tartam vizsgálatára specifikált modellben (és a V meglétére vonatkozóban is) a random hatások jóval nagyobb részét magyarázták a V-tartam (-jelenlét) variabilitásának, mind abszolút, mind relatív értelemben. A V-tartam esetében a random hatásoknak a magyarázott variabilitás kb. 50%-a tulajdonítható, ami az összesnek kb. 20%-a (marginális pszeudo- R^2 : 0,2200; kondicionális pszeudo- R^2 : 0,4236). A V megléte esetében pedig a magyarázott variabilitás szűk felét (az összes kb. 13%-át) magyarázták a random hatások (0,2065, ill. 0,3409). Tehát a V-ben megjelenő variáció sokkal inkább függ az egyes beszélőktől az l -ben megjelenőhöz képest.

Más specifikációkat is megvizsgáltunk, és azt találtuk, hogy az a modell, amelyben csak a betoldás- és a toldaléktípus-változó szerepel interakciójukkal együtt, önmagában szinte akkora részét magyarázza az l -tartam variabilitásának, mint a teljes modell: a marginális pszeudo- R^2 0,7416, a kondicionális pedig 0,8370.

Ellenőrzésképpen készítettünk egy olyan modellt is, amelybe csak a betoldás-változót vettük föl, és így számoltuk ki a pszeudo- R^2 -et. Azt látjuk, hogy az az egy változó önmagában képes fix hatásként az l -tartam variabilitásának több mint felét (61,33%) magyarázni, a random hatásokkal együtt pedig 66,09%-át. A marginális és a kondicionális változat közti kis különbség fényében térjünk rá arra, hogy vajon szignifikáns-e a betoldás megléte a teljes Vl -szekvencia esetén is. (A név ellenére nem kell feltétlenül V-t, vagyis betoldást tartalmaznia, elképzelhető, hogy pusztán egy (hosszabb) l szerepel az így hívott szekvenciában.)

3.5.2. A Vl -tartam

A teljes Vl -szekvencia tartamát vizsgáló modellben is viszonylag magas a pszeudo- R^2 : a marginális verzió 0,4925, tehát a fix hatások a modell közel felét magyarázzák, a kevert hatásokkal együtt (kondicionális verzió) a Vl hosszában

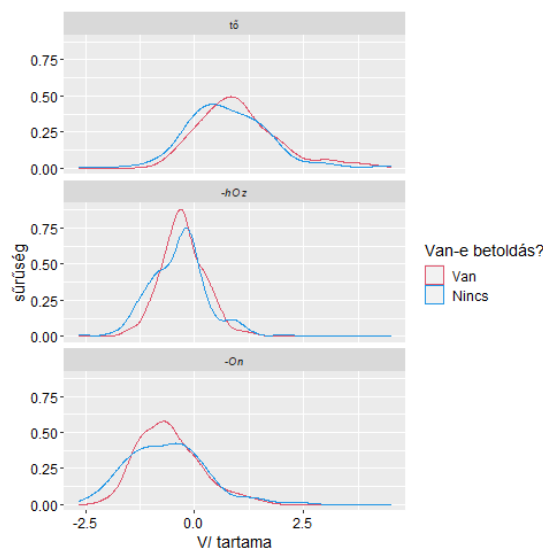


11. ábra. *Vl*-tartam célszavanként

rejlő variabilitás 74,47%-át magyarázza a modell. (A nem robusztus modell hasonló értékeket ad: 0,4276, ill. 0,7332.)

A változókat sorra véve egyrészt megerősíthetjük a célszavakat illetően a *Doodle* minimális elkülönülését (ld. a 11. ábrát) a teljes szekvencia szintjén is azzal, hogy itt minden mástól igen, de a *Lidl*től nem különbözik szignifikánsan ($p = 0,4104$; ezt és a további eredményeket ld. a Függelék 6.10. táblázatában). Ez alapján a *Doodle*-beli *Vl*-tartam 0,1934 és 0,2916 egység közti szórással hosszabb a többiben található nál. Ahogy előzőleg is, itt is a szó részleges ismeretlenségével tudjuk indokolni a többihez képest hosszabb *Vl*-szekvenciát. (Mindezen kívül egyedül a *Lidl* különbözik még szignifikánsan ($p = 0,0237$) a *Strobl*tól.)

Másrészt a toldaléktípusok közti EMM-ek mindhárom toldaléktípust egymástól szignifikánsan különböztetnek mutatják ($p < 0,001$). Ezek alapján átlagosan 1,1854 egységnyi szórással hosszabb egy tőalak *Vl*-tartama egy *-hOz*-zal ellátott alakhoz képest, míg 1,5776 egységnyi szórással egy *-On*-nal ellátotthoz képest. Ennek az okát a már említett frázisvégi nyúlásban kereshetjük. A két toldalékolt alak az előbbi szórástávolságok különbségével tér el egymástól, vagyis az allativusi alakok 0,3922 egységnyi szórással hosszabbak. Ez a különbség az *l*-tartam esetén kb. feleekkora (0,1714) volt, de a *V*-tartambeli különbséggel összevetve indokolt lehet, és utalhat is a különböző szótagszerkezetre az egyes toldaléktípusoknál.



12. ábra. A *VI*-tartam eloszlása toldaléktípusonként

Ugyanis míg az *-On* esetében az *l* (legalább részben) szótagkezdet, addig a másik két végződés esetében semmiképp sem az. Kóda (és szótagmag) esetén pedig esélyes, hogy hosszabb az *l*.

Harmadrészt a kérdés tekintetében valóban megáll szignifikánsan ($p < 0,001$) a teljes szekvenciára, hogy az kérdés esetén hosszabb az intonációjelölés érzékelte céljából, függetlenül attól, hogy hogyan épül fel a kiemelendő szótag szerkezete (mi a szótagmag). (A változó együttthatója $-0,1953$, vagyis $0,1953$ egységnyi szórással rövidebbek a kijelentésbeli *VI*-tartamok.)

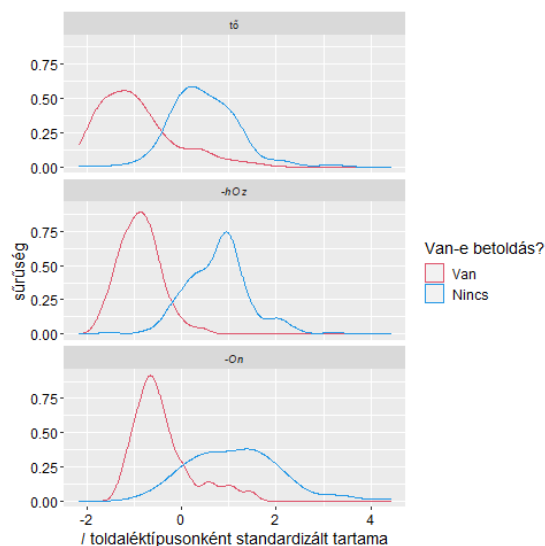
Ez utóbbihoz jön még hozzá a következő megfigyelés, és ez az, ami igazán jelentős (valamint konzisztensen is magyarázható különbség az előző modellt figyelembe véve). Bár a betolás-változóra végzett post-hoc teszt azt mutatja, hogy a két érték szignifikánsan különbözik ($p = 0,0024$, ld. a Függelék 6.10. táblázatát), de ha a toldaléktípussal vett interakciójának az EMM-jeit vizsgáljuk, máris azt találjuk, hogy egy-egy toldaléktípuson belül rendre nem különböznek szignifikánsan a betoldásos és a betoldás nélküli alakok a *VI*-tartam tekintetében (ld. a Függelék 6.11. táblázatát). Ugyanezt tapasztaljuk vizuálisan is: a 12. ábrát elnézve igencsak úgy tűnik, mintha nem volna jelentősebb különbség aszerint a

Vl egység tartamában, hogy a *V* megjelenik-e vagy sem. (Ugyanerre mutat a Függelék 5.2.1. (a) ábrája, ahol a modell illesztett *y*-értékei és a reziduálisok vannak ábrázolva a betoldás-változó szerint felbontva.) Ezt érdemes továbbá összehasonlítani a 13. vagy a függelékbeli 5.1.1. és 5.1.3. (a) ábrával, amelyeken jól látszik, hogy az *l*-tartam esetében viszont mennyire számít a betoldás jelenléte. Összességében tehát attól függetlenül, hogy a többeli *l* előtti utolsó mássalhangzó és a rag (vagy a szóvég) közötti elméleti *l*-szegmentumnak egy hosszabb *l* a megvalósulása vagy pedig egy *V*-betoldás és egy rövidebb *l* kapcsolata, a kettő tartamban érdemi különbséget nem mutat (vö. 12. ábra). Ez pedig így pusztán a *Cl*-szekvenciára vonatkozó javítóstratégiák közti individuális preferenciák eltérését mutatja, illetve azt, hogy azok alkalmazása nem eredményez különböző tartamot. A preferenciák szerepével összefüggésben állhat az is, hogy míg az *l*-tartam magyarázott variabilitásának mindössze kb. 11%-ért felelt a beszélőközi variáció (a random hatások), addig ez a *V*-tartam esetében kb. 48%-ot tett ki. Sőt, mindez a 4. hipotézis elemzése során kapott eredménnyel együtt (amely szerint az *ő* harmonikus megjelenése nem függ a betoldás jelenlététől) tovább erősíti azt a feltételezést, hogy nincs erőteljes fonológiai jelentősége a betoldásnak! Az pusztán egyes javítóstratégia-preferenciák artikulációs megnyilvánulása, de az elméleti szillabikus *l* harmonikus besorolását nem befolyásolja.

3.5.3. Szótagszerkezeti elemzés az *l*-tartam alapján

A hipotézis kapcsán kapott eredmények során többször hivatkoztunk az *l*-t tartalmazó szótagok és szavak szótagszerkezetére, így ezt röviden össze is foglaljuk egy külön fonológiai elemzés keretében a *Lidl* superessivusi alakját példaként használva.

A „szillabikus” *l*-t tartalmazó superessivusi alakok tekintetében két eltérő fonológiai mintázat ütközik: (1) szótagmag után az *-n* allomorf következik (pl. *ajtó-n*), kóda után *-Vn* (pl. *rajt-on*), (2) magánhangzó után ugyancsak az *-n* járul a tőhöz, míg mássalhangzók után *-Vn*. Ez a két mintázat a vizsgálatunk tárgyát képező szótípustól eltekintve a magyarban egybeesik, hiszen szótagmag csak magánhangzó, kóda pedig csak mássalhangzó lehetne. Ám ez az összhang a



13. ábra. Az *l*-tartam eloszlása toldaléktípusonként

szillabikus mássalhangzók (jelen esetben az *l*) megjelenésével megbomlik, hiszen az egyfelől szótagmag, másfelől pedig magánhangzó, így kérdéses, hogyan jár el a beszélő a kötőhangzót illetően, milyen javítóstratégiát alkalmaz.

Hogyan tudjuk ezt az ellentmondást a superessivusi alakok esetében feloldani? (Az allativusi toldaléknál nem merül föl a kérdés, hiszen ott a toldalék mindenképp mássalhangzó-kezdetű, kötőhang nélküli.)

Az első lehetséges megoldás, hogy a beszélők az *l* szillabikus voltát súlyozzák nagyobb mértékben, mint a mássalhangzó voltát. Ebből az adódik, hogy a ragnak kötőhangzó nélkül kell megjelennie: *li.dLn* (1).¹² Ilyen és hasonló alakok egy-egy beszélőnél előfordultak a felvétel során. Kérdéses ugyanakkor, hogy ezeket spontán, mindennapi beszédben is használnák-e a beszélők, vagy csupán a kísérlet alatt oldották így föl a két vázolt minta előbukkanó ellentétét, reakcióként egy addig ritkán látott esetre. (Meg kell jegyeznünk itt is, hogy előfordultak olyan, ebbe a típusba tartozó alakok, ahol volt V-betoldás. Akármennyire is

¹²Ebben a szakaszban a szillabikus, vagyis a szótagmag pozícióját betöltő *l*-t kiskapitális *l* betűvel jelöltük.

kevesebb mint 20%-át tették ki az ilyen alakoknak, ez is afelé mutat, hogy a betoldásnak csupán artikulációs szerepe van, ellenben nem hat a toldalékolás mikéntjének megválasztására – hiszen ezekben az esetekben akkor a toldalék kötőhangzós allomorfjának kellett volna megjelennie.)

A másik kézenfekvő megoldás az, hogy a mássalhangzósság kerekedik felül a szillabikusságon, a kötőhangzós allomorfot kiváltva, és ezzel azt eredményezve, hogy a CVCC szerkezetből $CVCC+VC = CVC.CVC$ keletkezik. Betoldásról itt nem beszélhetünk, és mivel megtörtént az *l*-nek a rag szótagjába történő átszótákolása (szótagkezdetté vált), így a harmonikus alternánst a tő első szótagjának magánhangzója határozza meg: *lid.len* (2).

Eddigi elemzésünkbe a most vázolt két lehetséges javítóstratégiát alkalmazó alakokat nem vontuk be, ugyanis ezek kétszótagosak lévén feltehetően másképp hatnak az *l* tartamára, illetve a betoldás megjelenésére, a harmonikus viselkedés szempontjából pedig mellékesek: az első stratégia esetén nem beszélhetünk ilyenről, a második esetén a tő első magánhangzójából pedig triviálisan következik bármely harmonikusalternáns-választás kimenetele. Mindemellett ilyen alakok viszonylag kis számban is fordultak elő.

Vegyük a további lehetőségeket: tegyük föl, hogy fönn akarjuk tartani a tő kétszótagúságát, úgy, hogy az például az előbb említett módokon ne sértse a paradigma egységességét: tehát a superessivusi alak esetében legyen egy első szótag ($CV. := li.$), egy, az *l*-t megelőző mássalhangzót és az *l*-t valamilyen módon, legalább részben tartalmazó szótag ($CV. := dV.$ vagy $dL.$), és a rag szótagja, annyi megkötéssel, hogy a hiátus is szűnjön meg az *l* és a rag magánhangzója között ($CVn := lön$). A hiátus feloldásához (ld. még a 3.2.1.-beli modell eredményeit) elengedhetetlen, hogy az *l* részben átkerüljön a rag szótagjába, és annak szótagkezdetévé váljon. Arra, hogy mi töltsen ki a második szótag szótagmagját, két lehetőség van: vagy betold egy V-t a beszélő, vagy pedig megnyújtja az *l*-t, és az így keletkező hosszabb *l* első fele alkotja a szótagmagot, míg a második

fél a következő szótag kezdetét.¹³ E két megoldás ekvivalenciáját pedig épp az imént láttuk, mikor megmutattuk, hogy a *Vl*-szekvencia attól függetlenül ugyanolyan hosszú, hogy szerepel-e betoldás, vagyis hogy mi a szótagmagja a második szótagnak. Ezzel beláttuk, hogy a hosszú *l*-es megoldás ugyanazt reprezentálja, mint a betoldást igénylő verzió. Így a kapott szóalak szótagolása vagy *li.dV.lön* (3), vagy *li.dL.lön* (4).¹⁴ Az ilyen és ezek aleseteként a következő bekezdésben tárgyalt javítóstratégiát alkalmazó alakok harmonikus viselkedéséről a következő szakaszban lesz szó. (Az egyszerűség és a megkülönböztethetőség kedvéért a toldalék-magánhangzót nem *V*-vel jelöltük, hanem a leggyakrabban, semleges tövek esetében szinte kizárólagosan tapasztalt *ö*-vel.) Annyit viszont itt újra elmondhatunk, hogy az *ö* megjelenését, mint kiderült, egyáltalán nem kondicionálja, hogy az ebben a bekezdésben tárgyalt két ekvivalens stratégia közül melyiket alkalmazzák a beszélők, vagyis hogy van-e betoldás.¹⁵

Az előbb tárgyalt javítóstratégiák egy-egy alese az (néhány beszélő következetesen ezeket használta!), amikor a korábban *Vl*-lél jelölt szekvencia második felében, tehát a *tő* második szótagjának szótagmagját követően még a toldalék-magánhangzó előtt az *l* megnyúlik: *li.dVl.lön* (5) és *li.dLl.lön* (6). A 13. ábrán a superessivusi *l*-tartam sűrűségfüggvényeit ábrázoló ablakban jól látható, hogy a betoldott alakok közt van némi további variáció az eloszlás felső végén, ami a leírt javítóstratégiáknak lenne megfeleltethető. Bár az allativusi alakok esetében (3) és (4) előfordulhat (*li.dVl.höz* (3h), ill. *li.dLl.höz* (4h)), az (5) és (6) szerinti alakok analógiájaként elképzelt **li.dVll.höz* (5h), ill. **li.dLll.höz* (6h)

¹³E gondolatmenetünk kizárólag fonológiai. Ahogy pl. a *pattan* szó szótagszerkezeti elemzésében sem volna különös a hosszú *t* szétválasztása (*pat.tan*), még ha fonetikailag egy zárképzésről-felpattanásról beszélhetünk is, úgy a *li.dL.lön* esetében sem volna példanélküli így eljárni; az elemzés szokatlanságát az adja, hogy a megfelelő hosszú mássalhangzót nem kódává és szótagkezdeté választjuk szét, hanem szótagmaggá és szótagkezdeté, ami ellenben szükség-szerű következménye az *l* (legalább részben) szótagalkotó voltának. Fonetikai szempontból ugyanakkor nem vizsgáltuk külön-külön az *l* első és második felét, ez egy további vizsgálat tárgya lehet.

¹⁴Mint már korábban is jeleztük, az, hogy az *l*-nek a közvetlenül a toldalék-magánhangzó előtti része valóban szótagkezdet-e, az *l*-ek vokalizáltságának vizsgálatából nyerhetne megerősítést, mivel szótagkezdetben ez a folyamat nem megy végbe.

¹⁵Ennek kapcsán fölmerül az is, hogy ha nincs jelentősége a betoldásoknak, akkor az *l* betoldás esetén nem szótagmag-pozícióból váltja ki a harmóniát, amely igen meglepő eredmény volna. Ld. még: A szillabikus *l* és a magánhangzó-harmónia

alak a magyar mássalhangzó-rövidülés miatt szükségszerűen (3h)-vá és (4h)-vá egyszerűsödik.

Ebből fakad az, mint (5) és (5h), valamint (6) és (6h) léte különbségének leképeződése, hogy a toldaléktípusonként standardizált *l*-tartam toldaléktípusonkénti összehasonlítása során mind a Levene-, mind a robusztusabb Brown–Forsythe-próba szignifikáns eltérést mutat az *l* -*hOz* és -*On* előtti varianciája közt mind a betoldásos, mind a betolás nélküli alakok körében (a pontos *p*-értékekért ld. a Függelék 5.2. táblázatát). Ráadásul ha kiszámoljuk az idevágó ferdeségeket (Függelék 5.1. táblázat), azt találjuk, hogy míg a betoldás nélküli alakok közt a -*hOz*-zal ellátott alakokban az *l*-tartam eloszlása csak kevésbé jobbferde (0,20), addig az -*On*-nal ellátottakban jóval inkább az (0,59). A különbség kétszer ugyanekkora a betoldásos alakok között: 0,43, ill. 1,23, tehát az -*On*-nal ellátott alakokban az *l*-tartam variációja sokkal inkább az eloszlás felső végén történik, ami az (5)-szerű alakok létre utalhat.¹⁶

Összesítve azt látjuk, hogy a superessivus esetén a hat fentebb vázolt lehetséges javítóstratégia közül választottak a beszélők. Az allativusi alakok esetében a két kiinduló elv ütközése nem releváns, viszont a tapasztalt szegmentális mintázatok megfeleltethetők a superessivus esetén a javítóstratégiák alkalmazása során megvalósult mintázatok némelyikének, ilyenek a (3h), a (4h), illetve a (2)-nek megfeleltethető [ibl.hez] (sőt, [ibl.høz]!) típusú alakok.

¹⁶A tőalakok *l*-tartamának varianciája szintén szignifikánsan különbözik az -*On*-nal ellátott alakokétól (szintén ld. a függelékbeli 5.2. táblázatot): a betoldás nélküli alakok esetén kisebb és a -*hOz*-zal ellátott alakokétól nem különbözik szignifikánsan, a betoldásos alakok esetén mindkettőnél szignifikánsan nagyobb, ami megintcsak visszavezethető a frázisvégi nyúlásra. Ugyanakkor fontos látni, hogy a tőalakok *l*-tartamának ferdesége mindkétszer kisebb az -*On*-nal ellátott alakokéénál: rendre 0,55 és 1,01 betoldás nélküli és betoldásos alakokban. Utóbbi nagyfokú variációra mutat az eloszlás felső végén, de itt ez nem lehet egy (5)-szerű javítóstratégia megnyilvánulása, hanem feltehetőleg a nagy mértékű variancia kapcsán is éppen említett frázisvégi nyúlásra utal.

4. A szillabikus *l* és a magánhangzó-harmónia

Az alábbiakban azt mutatjuk be, milyen problémát okoz a szillabikus *l* viselkedése a magyar magánhangzó-harmónia leírásának szempontjából. Ehhez először összefoglaljuk az *opcionális átlátszóság* és a *kerekségi harmónia* jelenségeit.

Átlátszóságról akkor beszélhetünk, ha a toldalék harmóniáját nem az utolsó, hanem az azt megelőző magánhangzó hangrendje határozza meg. Az átlátszóság a magyarban kizárólag a semleges magánhangzókra (vagyis az *i*, *í*, *é* és *e* hangokra) jellemző tulajdonság. Azonban míg az *i* és az *í* szinte teljes mértékben átlátszó (bár ld. Fejes & Rebrus 2020), az *é* és az *e* nem mindig az. Ez egyaránt megnyilvánul lexikális variációban, tehát azonos harmonikus mintájú lexikális elemek eltérő viselkedésében (*koncerten*/**koncerton*, de *have-rok*/**haverek*), és ingadozásban, amikor egy lexikai elem többféle viselkedést is mutat (*bohémme*/*bohémme*, *hotelben*/*hotelben*).

Ettől független jelenség az 1.2 szakaszban is tárgyalt kerekségi harmónia. Ennek kiváltó osztályát az elülső magánhangzók, vagyis az *ö*, *ü*, *ő*, *ű* képezik. Ha ezen magánhangzók a *tő* utolsó magánhangzói, a háromalakú toldalékok közül az *ő*-t tartalmazó fog megjelenni.

Milyen harmonikus osztályba tudjuk besorolni a szillabikus *l*-t?¹⁷ Az eddigiekben részletezett legfontosabb tulajdonsága, hogy háromalakú toldalékoknál kiváltja a kerekségi harmóniát (ld. 9. táblázat, L jelöli a szillabikus *l* osztályát). Ebből a szempontból az elülső magánhangzókkal viselkedik azonosan, hiszen a semleges magánhangzók után sosem jelenik meg elülső alternáns.¹⁸

Csak hogy a szillabikus *l*-re emellett jellemző az opcionális átlátszóság is. A kísérlet során a direkt összehasonlíthatóság kedvéért nem vizsgáltunk kétalakú

¹⁷A fenti statisztikai elemzés eredményei alapján ugyanis joggal hívhatjuk szillabikusnak, hiszen a betoldás jelen(nem)léte nem befolyásolja a harmóniát, erősen valószínű, hogy egy fonetikai szinten lejátszódó beszélő- és lexikalizáltságfüggő folyamatról van szó, amelynek kevésbé lehet fonológiai reprezentációt tulajdonítani. Így az *l* még betoldás esetén is kváziszillabikusként viselkedhet (értsd: képes kiváltani a harmóniát), holott a szótagszerkezet önmagában nem ezt jósolná.

¹⁸A kísérlet során előfordultak *Yblhez* alakok, ezek viszont olyan esetek voltak, mikor a beszélő egyszótagos tövel ejtette a szót, ilyenkor valójában nem beszélhetünk szillabikus hangról. Az antiharmonikus ejtést (pl. *Yblhoz*) pedig produkciós esetlegességnek tekintjük. (Vö. 2 db *Yblhoz* és 0 db *Stroblhez*, de 18 db *Stroblhöz*, ld. még a 9. ábrát.)

	előlségi harmónia	
NN	zsiliphez	*zsiliphöz
NL	*Yblhez	Yblhöz
NF	*tribünhez	tribünhöz

9. táblázat. Kerekségi harmónia

toldalékokat, de mind anyanyelvi intuíciónk, mind internetes keresések alapján elmondható, hogy kétalakú toldalékoknál bizonyos szillabikus *l*-t tartalmazó tövek toldalékolása ingadozik (pl. *Schadlnek/Schadlnak*). Ez a tulajdonsága harmonikus szempontból egyértelműen a semleges magánhangzókhoz köti a szillabikus *l*-t, elülső magánhangzókra egyáltalán nem jellemző az átlátszóság (ld. 10. táblázat).

	előlségi harmónia	
BN	bohémnek	bohémnak
BL	Schadlnek	Schadlnak
BF	parfümnek	*parfümnak

10. táblázat. Opcionális átlátszóság

A szillabikus *l* ezek alapján tehát nem sorolható be egyik harmonikus osztályba sem, hiszen nincsenek olyan magánhangzók, amelyek kiváltják a kerekségi harmóniát, de előlség szempontjából átlátszóak lehetnek. Ezt foglalja össze a 11. táblázat.

	átlátszóság	kerekség
N	✓	✗
L	✓	✓
F	✗	✓

11. táblázat. A szillabikus *l* és a magánhangzók csoportjai

Ennek sajátos megnyilvánulása a *Strobl* tő allativusi alakjának ingadozása (*Stroblhoz/Stroblhöz*). A második szótag lehet átlátszó, ilyenkor a hátsó alternáns jelenik meg, de ha nem átlátszó, akkor elülső (kerekített) alternáns követel. Az így létrejövő *o~ö* ingadozás szinte sehol máshol nem figyelhető meg

a magyar magánhangzó-harmónia rendszerében,¹⁹ egyedül a szillabikus *l*-hez kötődő jelenség.

Hogy miért éppen hátsó magánhangzó után lehet átlátszó, semleges után viszont kerek ségi harmóniát kiváltó a szillabikus *l*, annak kapcsán visszautalnánk a betoldás redukáltságát taglaló 3.2. szakaszra. Amikor nem az *l* a szótagmag, hanem a betoldott V, bár magánhangzónak már magánhangzó áll a szótagmagban, ennek a minősége közel sem egységes, mint azt a V minőségének bemutatásakor (ld. a 3. és a 4. ábrát) láttuk. Tehát nem köthető egyetlen harmonikus osztályhoz sem, ezáltal nemcsak betoldás nélkül, de betoldással együtt is bizonytalan ennek a szótagnak a harmonikus szerepe a hátsó/semleges/elülső osztályok viszonylatában.

Viszont a nem utolsó szótagban csak semleges magánhangzót tartalmazó tövek esetében mindig az utolsó szótag határozza meg a harmonikus viselkedést (*királyhoz*, *zsiliphez*, *tribünhöz*; *vityillóhoz*, *biciklihez*, *retikülhöz*), így az említett szótag magja kénytelené válik harmonikus osztályt választani (vagy újat létrehozni!), hiszen ő maga határozza meg az egész szó harmonikus viselkedését. Hogy ebben az esetben miért éppen az elülső osztálynak megfelelően viselkedik, arra vonatkozóan az *l* akusztikai hasonlóságával érveltünk korábban.

Ám utolsó előtti szótagi hátsó magánhangzók esetén, amelyek – ellentétben a nem utolsó szótagbeli semlegesekkel – az utolsó szótag harmonikus minőségétől függően ugyan, de képesek meghatározni a harmóniát, nem triviális, hogy a hátsó magánhangzó vagy az utolsó, *l*-t tartalmazó szótag lesz a meghatározó a szó harmonikus viselkedése szempontjából. Ebből adódik az L-osztálynak az a különleges tulajdonsága, hogy még átlátszósági *o/ö*-váltakozást is kiválthat.

¹⁹Toldalékallomorfiában is egyedül a *-nok/-nök* képzőben jelenik meg *o~ö* váltakozás.

5. Összefoglalás

Összefoglalva azt láttuk, hogy a beszélők szinte kivétel nélkül két szótagban ejtették a célszavak töveit, vagyis beszélhetünk szillabikus *l*-ről. Abban az értelemben biztosan, hogy a realizációja vagy annak egy része (a redukált magánhangzó betoldása vagy pedig a hosszú *l* első fele) egy önálló szótag magjává lesz, amely a beszélők nagy többségénél minden vizsgált alakban (tő, superessivus, allativus) – és feltehetőleg így a célszavak, valamint a hozzájuk hasonló egyéb szavak paradigmájának egészében is – megjelenik.

Eldöntendő kérdésben gyakrabban toldottak be egy redukált magánhangzót az *l* elé, mint kijelentésben. Ráadásul maga a betoldás is hosszabb volt, nemcsak a redukált magánhangzó, hanem a teljes *Vl*-szekvencia tekintetében is (amelyet a szillabikus *l* realizációjaként értelmezhetünk, akár volt betoldás, akár nem). Mindez – azzal összhangban, hogy az említett szekvencia tartama független volt attól, hogy képezi-e részét betoldás – afelé mutat, hogy az egész szekvencia nyúlik meg az intonációjelölés szükségességénél fogva, anélkül, hogy elkülönülne szótagmagra és kódára. A betoldás és az *l* még szorosabb egységét bizonyítja az is, hogy nem találtunk összefüggést a betoldás jelenléte és a harmónia között (különös tekintettel a célszavak első szótagjai alapján egyáltalán nem motivált *ő*-re). Így bár vannak tendenciák a betoldás megjelenésére különböző tényezők alapján (pl. a betoldási arány nagyobb volt a superessivusi toldalék előtt, mint egyébként, ezt szótagszerkezeti okokkal magyaráztuk), fontos látni, hogy ennek nem tulajdoníthatunk nagyobb fonológiai jelentőséget, különösképpen a harmónia kérdésében nem.

Ez fölveti azt a kérdést, hogy akusztikailag milyen minőségűek a betoldott redukált magánhangzók. Erre az elemzésünk nem tért ki, de érintőleges vizsgálatok azt mutatják, hogy a célszavak első szótagjával erősen koartikuláltak (egyes beszélőknél különösen azok), de nem határoznak meg egy konkrét magánhangzó-minőséget, amelynek alapján meg lehetne jósolni a szavak harmonikus viselkedését, ebben az irányban még mindenképp érdemes további vizsgálatokat folytatni. Ellenben ha a betoldás nem jelentős, akkor az *l*-nek

kell kiváltania a harmóniát még akkor is, ha a betoldás jelen van (vagyis az *l* nem szótagmag)! Ez az eddigi (különösen a magyar) harmóniáról alkotott elképzeléseket merőben új megvilágításba helyezné. Ugyanakkor figyelembe véve azt, hogy pl. a délszláv nyelvekben megfigyelhető szillabikus *r*-nek is létezik egy megelőző svával és anélkül ejtett allofónja is, amennyiben az *l*-nél is egy hasonló jelenségről volna szó, úgy a probléma egy része meg is oldódna azzal, hogy fonológiai reprezentációt egyáltalában nem tulajdonítunk a betoldásnak, ellenben az *l*-hez egy új harmonikus osztályt rendelünk.

Azt is láttuk, hogy hátsó tövek esetén gyakrabban jelent meg a *tő* első szótagjának harmonicitását követő toldalék, miképp az allativusi toldalék esetén szintén nagyobb volt ez az arány, mint superessivusi alakok esetén. Ugyanakkor minden célszó esetében igen jelentős arányban fordult elő az elülső alternáns. Így az *l*, mint azt leírtuk elemzésünk során, egyfelől viselkedhet átlátszóként (miként a semleges magánhangzók) hátsó magánhangzók után, másfelől kiváltja a kerek ségi harmóniát, mint az elülső magánhangzók. Ez semelyik eddig leírt harmonikus osztállyal nem összeegyeztethető, így egy új osztályt kell rendelnünk az *l*-hez.

Azzal kapcsolatban is érdemes további vizsgálatokat végezni, hogy ha a betoldás nem játszik szerepet, az *l*-nek milyen fonetikai tulajdonsága okozza azt, hogy ki tudja váltani a kerek ségi harmóniát. Elöljáróban annyit jelenthetünk ki, hogy az, hogy az alveolárisok (így az *l*) lókusza közel áll az *ő* második formánsához (és hogy specifikusan az *l* formánsszerkezete igen hasonló az *ő*-éhez), elképzelhetővé teszi, hogy az (esetlegesen betoldással kiegészült) *l* elülső magánhangzóként való percepciója váltja ki az elülső toldalékalternánsokat. Ugyanakkor ennek bizonyítása további kísérletek feladata lesz.

A tanulmányhoz tartozó függelék a következő linken érhető el:

https://osf.io/mzgrd/?view_only=120e4fd17150472293a08b5e352b2cbe.

Köszönetnyilvánítás

A tanulmány megszületését az NKFI-139271 számú pályázat (*A paradigma-szerkezet szerepe a magyar fonológiában és alaktanban nyelvtipológiai szempontból*) támogatta.

A kutatás a Kulturális és Innovációs Minisztérium ÚNKP-22-1 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.



Hivatkozások

- Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer (version 6.2.03). URL: <http://www.praat.org>.
- Draxler, C., & Jänsch, K. (2004). Wikispeech – a content management system for speech databases. In *Proc. of Interspeech, Brisbane, 2008* (pp. 1646–1649).
- Fejes, L., & Rebrus, P. (2020). Mennyire átlátszó az i? *Hungarológiai Közlemények*, 20, 49–60. doi:10.19090/hk.2019.4.49–60.
- Jurjec, P., & Bjorkman, B. M. (2018). Indexation to stems and words. *Phonology*, 35, 577–615. doi:10.1017/s0952675718000210.
- Kenstowicz, M. (2003). Review article: The role of perception in loanword phonology. *Studies in African Linguistics*, 32, 96–99.
- Markó, A. (2017). Hangtan. In G. Tolcsvai Nagy (Ed.), *Nyelvtan* (pp. 75–206). Budapest: Osiris Kiadó.
- Nádasdy, Á. (1989). Consonant length in recent borrowings into hungarian. *Acta Linguistica Hungarica*, 32, 195–213. URL: <https://www.jstor.org/stable/44362748>.

- Rebrus, P., & Törkenczy, M. (2019). Magyar harmónia: a dolgok állása. *Általános Nyelvészeti Tanulmányok*, XXXI, 233–333.
- Siptár, P., & Törkenczy, M. (2000). *The Phonology of Hungarian*. Oxford/New York: Oxford University Press.
- Vendelin, I., & Peperkamp, S. (2006). The influence of orthography on loanword adaptations. *Lingua*, 116, 996–1007. doi:10.1016/j.lingua.2005.07.005.
- Zsilinszky, É. (2003). Szókészlettörténet. In J. Kiss, & F. Pusztai (Eds.), *Magyar nyelvtörténet* (pp. 372–392). Budapest: Osiris Kiadó. URL: <https://www.szaktars.hu/osiris/view/kiss-jeno-pusztai-ferenc-szerk-magyar-nyelvtortenet-osiris-tankonyvek-2003/>.

Audiovizuális beszédszintézis nyelvultrahang alapon

Csapó Tamás Gábor¹

¹*Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék*

Abstract

In this study, we present our initial results in audiovisual text-to-speech synthesis (AV-TTS), which is a subfield of the more general areas of speech synthesis and computer facial animation. The goal of visible speech synthesis is typically to generate face motion or articulatory-related information (e.g., lip, tongue movement, or velum position). We conduct experiments in text-to-articulation prediction, using ultrasound tongue image targets. We extend a traditional deep neural network-based text-to-speech synthesis (DNN-TTS) framework by predicting ultrasound tongue images, of which the continuous tongue motion can be reconstructed in synchrony with synthesized speech. The final output is speech and ultrasound tongue video in 'wedge' orientation. We use the data of eight English speakers (roughly 200 sentences from each) from the UltraSuite-TaL dataset, train several types of deep neural networks, and show that simple DNNs are the most suitable ones for the prediction of sequential articulatory data, as we have limited training material. Objective experiments and visualized predictions show that the proposed solution is feasible and the generated ultrasound videos are mostly close to natural tongue movement but are sometimes oversmoothed. A specific application of audiovisual speech synthesis and text-to-articulation prediction is computer-assisted pronunciation training/computer-aided language learning, which can be beneficial for learners of second languages. With such an AV-TTS, by giving an arbitrary input text, one is able to hear the synthesized speech and, in synchrony with it, see (in 2D or 3D) how to move the tongue to produce target speech sounds. This visual feedback can be helpful for pronunciation training in L2 learning, especially when the target language contains speech sounds that are difficult to articulate (e.g., significantly different from the speaker's mother tongue).

Keywords: AV-TTS, mély neurális hálózatok, DNN, beszédtechnológia

1. Bevezetés

A beszédszintézis (más néven gépi szövegfelolvasás, text-to-speech, TTS) célja, hogy írott szöveget alakítsunk át emberihez hasonló beszéddé. Az audiovizuális beszédszintézis esetén nem csak beszéd a rendszer kimenete, hanem

Email address: csapot@tmit.bme.hu (Csapó Tamás Gábor)

valamilyen más kiegészítő információt, videó formájában megjeleníthető tartalmat is láthatóvá kívánunk tenni az emberi artikulációról. Ez a kiegészített tartalom lehet például beszélő fej (Czap & Mátyás, 2003, 2005), száj- és arcmozgás (Massaro et al., 2012; Schabus et al., 2014), vagy a nyelv mozgására utaló információ (Steiner et al., 2017; Le Maguer et al., 2017; Yu et al., 2019).

A beszéd-szintézisben napjainkban legtöbbször valamilyen gépi tanulási módszer alkalmaznak. A 2010-es évek elején a rejtett Markov-modell (HMM) alapú megoldások voltak elterjedtek (Tóth & Németh, 2008; Csapó & Németh, 2011), azonban az évtized közepétől ezeket fokozatosan felváltották a mély neurális hálózatot (Deep Neural Network, DNN) alkalmazó rendszerek (Zen et al., 2013; Zainkó et al., 2017). A neurális hálózatok rétegei igen hatékonyan képesek az adatokra jellemző tulajdonságokat megtanulni. Ez azt jelenti, hogy magukból a nyers adatokból tanulja meg a rendszer, hogy milyen absztrakcióval írhatóak le azok; nem pedig ember által megalkotott szabályokat követ. A TTS kutatások iránya a legutóbbi években az 'end-to-end' irányba halad, amely azt jelenti, hogy nyelvészeti tudás nélkül, csak a szöveges tartalom alapján rendeli hozzá a gépi rendszer az adott szöveghez legjobban illeszkedő beszédet, és minden belső komponens neurális hálózat alapú – erre példa a Tacotron2 rendszer (Shen et al., 2018). Ugyanakkor a hagyományos, nem end-to-end típusú, 'klasszikus' DNN-TTS megoldásokat is érdemes használni, amennyiben kevés a rendelkezésre álló adat, hiszen a Tacotron2 rendszer betanításához például több 10 óra nagyságrendű beszédfelvétel szükséges. Főleg akkor beszélhetünk kevés adatról, ha nem csak beszéddel, hanem valamilyen egyéb kiegészítő biológiai jellel is dolgozunk, például nyelvultrahang (Csapó et al., 2017a,b; Hajjé & Csapó, 2020), ajakvideó (Rácz & Csapó, 2020; Arthur & Csapó, 2021), vagy agyi jel (Arthur & Csapó, 2022). A jelen cikk témája, az audiovizuális beszéd-szintézis esetén szintén tipikusan csak kevés adat (néhány 10 percnyi felvétel) áll rendelkezésre.

1.1. Audiovizuális beszéd-szintézis

Az audiovizuális beszéd-szintézis a beszéd-szintézis és a számítógépes animáció általánosabb területeinek egy része (Massaro et al., 2012). A vizuális

beszédszintézis területe meglehetősen jól megalapozott, és az elmúlt években számos megközelítést fejlesztettek ki (beleértve a szabályalapú (Perrier, 2014) és az adatvezérelt módszereket (Schabus et al., 2014)). A szabályalapú rendszerek magukban foglalják a beszédhangsorozat megtervezését, az izommechanizmusok és a fizikai beszédképző rendszerek modelljét is. A vokális traktus biomechanikai modelljében a nyelv egy végeselemhálóként ábrázolható (Perrier, 2014), és komplex biomechanikai szimulációk szükségesek az emberi artikulációs szervek mozgása során fellépő belső izomfeszültségi állapotok becsléséhez (Stavness et al., 2014). Az adatvezérelt megközelítésekben két fő kategória van: a képalapú rendszerek célja a valós személyről készült videó szintézise, míg a mozgásrögzítésen ('motion capture') alapuló rendszerek az arcpontokból származtatott jellemzők időbeli múlását jelenítik meg (Schabus et al., 2014). Az adatvezérelt, gépi tanulás alapú rendszerekhez olyan beszédkorpuszra van szükség, amely a beszédadatok mellett az arc vagy egyéb artikulációs szervek mozgásának adatait is tartalmazza, szinkron módon rögzítve.

1.2. Artikulációsmozgás-becslés szöveg alapján

Az audiovizuális beszédszintézis egyik típusa az az eset, amikor a szöveg alapján artikulációs mozgást (pl. ajak- vagy nyelvmozgás) szeretnénk megbecsülni a szintetizált beszéddel párhuzamosan. Ehhez speciális biojelek rögzítésére van szükség, amelyek nyomon követik az artikulációs szervek mozgását (pl. elektromágneses artikulográfia, röntgen, mágnesesrezonancia-képalkotás, ajakvideó és nyelvultrahang). Egy ilyen rendszerrel tetszőleges bemeneti szöveghez meghallgatható a szintetizált beszéd, és ezzel szinkronban megtekinthető (2D / 3D-ben), hogyan lehet mozgatni a nyelvet, hogy adott beszédhangokat állítsunk elő. Ez a vizuális visszacsatolás nagy eredményt hozhat az idegennyelv-tanulásban, különösen akkor, ha a cél nyelv nehezen artikulálható beszédhangokat tartalmaz.

A legtöbb korábbi tanulmány ezen a területen pontkövető eszközöket, például elektromágneses artikulográfiát (EMA) használt (Ling et al., 2010a,b; Wei et al., 2016; Steiner et al., 2017; Le Maguer et al., 2017; Yu et al., 2019). Ling és munkatársai (2010a) egy HMM-alapú szövegből artikulációs mozgást előre-

jelző rendszert javasoltak, amely képes szintetizálni a beszélő szájmozgását. Itt még nem modellezték az időtartamokat, de egy későbbi tanulmányban az időzítési szempontokat is vizsgálták, és elemezték a kritikus artikulátorokat (Ling et al., 2010b). Wei és munkatársai (2016) DNN-eket használtak a szövegből EMA-ba történő előrejelzéshez. Steiner és munkatársai (2017) hasonlóképpen a szöveg-EMA előrejelzéssel kísérleteztek HMM-ek segítségével (szinkron szöveg-felolvasóval), és célként egy geometriai 3D nyelvmodellt is beépítettek. Ezután összehasonlították a HMM-eket és a DNN-eket a szöveg-nyelv modell előrejelzéséhez (Le Maguer et al., 2017). Az eredmények szerint a DNN-ek már 2 óránál kevesebb adattal is felülmúlták a HMM-eket.

Ahogy fent látható, számos tanulmány vizsgálta a szöveg-artikulációs mozgást HMM-ekkel vagy DNN-ekkel, de mindegyik pontkövető berendezést (elektromágneses artikulográfia) használt. Csapó (2021) kezdeti kísérleteket mutatott be nyelvultrahanggal, amit a jelen cikkben az eredmények részletesebb elemzésével egészítünk ki.

1.3. A nyelvultrahang

Az ultrahangot nemzetközi szinten az 1980-as évek kezdete óta használják beszédkutatóra (Stone et al., 1983). Magyarországon az artikulációs vizsgálatok új korszaka nyílt meg az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport 2016-os megalakulásával. Attól függően, hogy a vizsgálófejet (transzdúcer) milyen helyzetben (orientációban) helyezzük az állkapocs alá, többféle irányból is vizsgálható a nyelv. A leggyakrabban a midszagittális orientációt használjuk. A midszagittális felvételek során az ultrahangtranszdúcort az áll alá helyezik; így az ultrahangjelben a legnagyobb változást a nyelv izomzatának felső határa okozza, ami az ultrahangos képeken ideális esetben jól kivehető fehér sávot eredményez.

Az ultrahangos módszer előnye a többi artikulációs rögzítési technikához képest, hogy egyszerűen használható, non-invazív, elérhető árú, valamint nagy felbontású (akár 800 x 600 pixel) és nagy sebességű (akár 100–150 képkocka/másodperc) felvétel készíthető vele. A jó térbeli felbontás azért fontos, hogy a

nyelv alakjáról minél pontosabb képet kapjunk; míg a jó időbeli felbontás ahhoz szükséges, hogy a beszédhangok képzésének gyors változását (pl. zár-felpattanás; koartikuláció) is vizsgálni tudjuk. Az ultrahang hátránya ugyanakkor, hogy a hagyományos beszédkutatói kísérletekhez a rögzített képsorozatból ki kell nyerni a nyelv körvonalát ahhoz, hogy az adatokon további vizsgálatokat lehessen végezni. Ez elvégezhető manuálisan, ami rendkívül időigényes, vagy automatikus módszerekkel, amelyek viszont ma még nem elég megbízhatóak (Csapó & Csopor, 2015; Csapó & Lulich, 2015; Whalen et al., 2019). Az ultrahang használatának bizonyos mértékig hátránya az is, hogy csak a nyelv középső részéről ad információkat, gyakran a nyelvgyök és/vagy a nyelvhegy nem látszik. Emellett előfordulhat, hogy ha a nyelv felülete közel párhuzamos az ultrahangsugárral, akkor a középső részből is hiányos az információ.

1.4. A jelen kutatás célja

A cikk célja az audiovizuális beszéd-szintézishez való hozzájárulás, azaz a DNN-TTS kiterjesztése artikulációs mozgás-előrejelzéssel, a nyelv ultrahangos képeinek felhasználásával. Megmutatjuk néhány beszélő adatain, hogy a kombinált TTS és szintetizált artikulációs mozgás megvalósítható és elfogadható artikulációs mozgás-videót eredményez. A szöveg alapján történő artikulációs mozgás-előrejelzés hasznos lehet a számítógéppel segített kiejtéstanulás (Computer-Assisted Pronunciation Training, CAPT) alkalmazásokhoz és az artikuláció vizualizációjához.

2. Módszerek

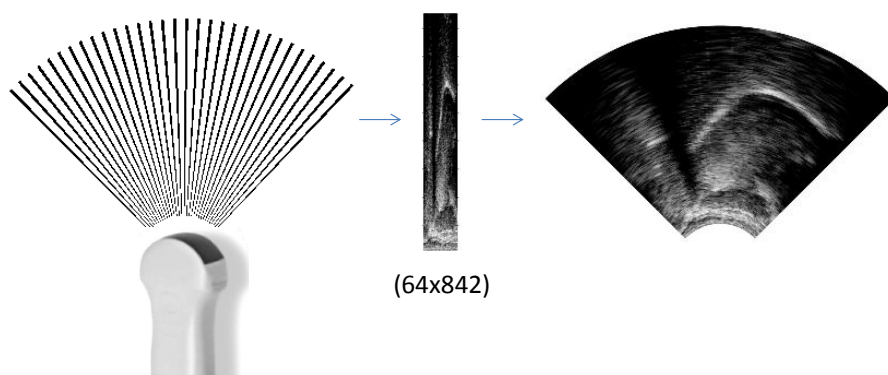
2.1. Adatok

Adatként az UltraSuite-TaL80 adatbázist használtuk fel (Ribeiro et al., 2021) (https://ultrasuite.github.io/data/tal_corpus/). Négy angol nyelvű férfi (03mn, 04me, 05ms, 07me) és négy női beszélőt választottunk (01fi, 02fe, 06fe, and 09fe). A beszéddel párhuzamosan a nyelv mozgását midszagittális orientációban rögzítették az Articulate Instruments Ltd. „Micro” ultrahangos

rendszerével, 81,5 képkocka/másodperc sebességgel. Az UltraSuite-TaL80-ban ajakvideót is rögzítettek, de ezt az információt nem használtuk fel a jelenlegi tanulmányban. Az ultrahangadatok és a beszédjelek szinkronizálása az Articulate Instruments Ltd. által biztosított eszközzel (Articulate Assistant Advanced, V219.08) történt. Minden beszélő közel 200 mondatot olvasott fel – a felvételek időtartama beszélőnként kb. 15 perc volt, amit 85-10-5 arányban bontottunk fel tanuló, validációs és teszt adatokra a gépi tanulási kísérletekben.

2.2. Az ultrahangos adatok feldolgozása

Kísérleteinkben a 'nyers' ultrahangos adatokból számított artikulációs jellemzőket használtuk a gépi tanulás további célpontjaként. A 'nyers' adat azt jelenti, hogy az ultrahangeszközből érkező intenzitásinformációt közvetlenül bináris formátumba mentettük (így nem vészett el adat a képpé konvertálás során), és így is dolgoztuk fel. Az 1. ábra mutatja, hogy a letapogatás hogyan történik a „Micro” rendszerrel: az ultrahangfej 64 radiális vonalon (bal oldalon), minden vonalon 842 helyen méri az intenzitást (azaz a szürkeárnyaltos színskálát), és a nyers adatban minden intenzitásértéket 8 biten tárol (ennek eredménye látható középen). Ha ezt a szokásos ultrahangképpé akarjuk alakítani, akkor az adatokat poláris koordinátarendszerben lehet ábrázolni szürkeárnyaltos képként, mely a jobb oldalon látható.



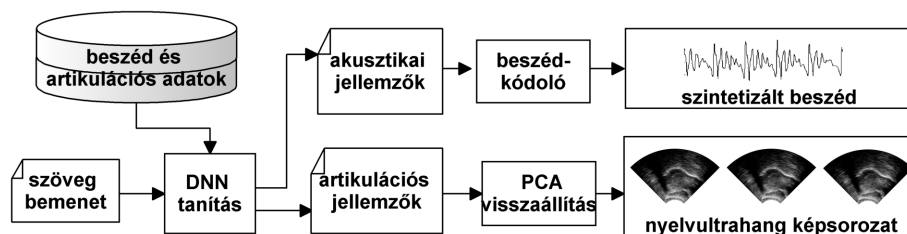
1. ábra. Nyers adatokból ultrahangkép előállítása

A 64×842 pixeles képeket átméreteztük 64×128 pixelre. Az ultrahangos kép viszonylag redundáns, ezért hatékonyan tömöríthető, ami a gépi tanuláshoz előnyös lehet, hiszen csak kisebb dimenziójú adattal kell dolgoznunk. Emiatt az 'EigenTongues' (Hueber et al., 2007) módszert követve főkomponens-analízist (Principal Component Analysis, PCA) végeztünk a képeken, melynek során a képpontok varianciájának 70%-át hagytuk meg, így 128 együtthatóra tömörítve az egyes képeket. Ahhoz, hogy az artikulációs adatok szinkronban legyenek az akusztikai adatokkal, újramintavételeztük az előbbi 200 Hz-re, ami 5 ms lépésközzel felel meg.

2.3. DNN-TTS rendszer

A 2. ábra szemlélteti a javasolt megközelítést, azaz a szöveges bemenetből gépi tanulóval (mély neuronhálóval) becsült kombinált akusztikus és artikulációs jellemzőket, valamint a kimeneten megjelenő szintetizált beszédet és nyelvultrahang-képsorozatot (videót). A kísérleteket a Merlin DNN-TTS keretrendszert (<https://github.com/CSTR-Edinburgh/merlin>) felhasználva végeztük (Wu et al., 2016), amely angol nyelvre kidolgozott recepteket tartalmaz. A szöveges bemenetet a rendszer először nyelvi jellemzőkké alakítja: ún. környezetfüggő címkék készülnek (Tóth & Németh, 2008; Tóth, 2013). A gépi tanulás bemenete tehát egy-egy beszédhang reprezentációja 425-dimenziós címkéként. A kimenethez a beszédből akusztikai jellemzőket számítunk (60-dimenziós spektrális együtthatók, 5-dimenziós aperiodicitás, és 1-dimenziós alaphérfekvencia), míg az artikulációs adatokat a 128-dimenziós PCA komponensekkel reprezentáljuk. A kimeneti paraméterekhez delta és delta-delta jellemzőket is számít a rendszer, így a kimeneti vektor teljes dimenziója 582. A bemeneti és kimeneti jellemzőkből a Merlin rendszer idősort készít, és külön hálózatot tanítunk az időzítésekre, illetve az akusztikai/artikulációs paraméterekre. Egyszerre az idősortnak egy-egy eleme megy a hálózatba, és így egy-egy 5 ms-os blokknak megfelelő jellemzőt generál.

A kísérletekben két neurális hálózat-architektúrát hasonlítottunk össze: 1) FC-DNN (fully connected deep neural network), és 2) LSTM (Long-Short Term



2. ábra. A javasolt módszer blokkdiagramja

Memory). Az FC-DNN esetén hat rejtett réteget alkalmaztunk, rétegenként 1024 neuronnal, 256-os batch mérettel, SGD optimalizálóval, batch normalization és dropout nélkül. Az LSTM esetén a háló elején négy teljesen kapcsolt réteg volt (mindegyik 1024 neuronnal), melyet egy LSTM réteg követett (512 neuronnal), batch normalization és dropout nélkül. Optimalizációnak az ADAM algoritmust alkalmaztunk, és a batch méret 256 volt. Minden esetben az MSE hibafüggvényt alkalmaztuk, és a tanítás a hibavisszaterjesztés (backpropagation) algoritmussal történt. Az előbbi, FC-DNN egyszerűbb modell és gyorsabban tanítható; az utóbbi, LSTM bonyolultabb és általában lassabb a betanítása, de 'rekurens' típusú hálózat, azaz jobban tudja modellezni az adatok időbeliségét, mint az FC-DNN. A két hálózat összpáraméterszáma hasonló volt (FC-DNN: 6,28 millió, LSTM: 6,03 millió). Mindegyik hálózatot külön-külön tanítottuk a négy férfi és négy női beszélő adatain, beszélőfüggetlen modelleket létrehozva.

A szintézis során a PCA-tömörített artikulációs adatokból inverz PCA transzformációval állítjuk elő a 64×128 pixeles 'nyers' elrendezésű nyelvultrahangképeket, melyeket újra átméretezünk az eredeti 64×842 pixeles méretre. Vizualizációs célból ezt a nyers adatot a fonetikai megjelenítésben is használt 'ék'/'legyező' formátumba alakítjuk át, amely megmutatja a nyelv felületének valós arányait (lásd a 4. ábrát). Utóbbi transzformációt az 'ultrasuite-tools' eszközzel véghezvük (<https://github.com/UltraSuite/ultrasuite-tools>). Végül a szintetizált beszédet és a nyelvultrahang-képsorozatot folyamatos videóvá fűzzük össze, amely az audiovizuális beszédsszintetizátor rendszer kimenete lesz.

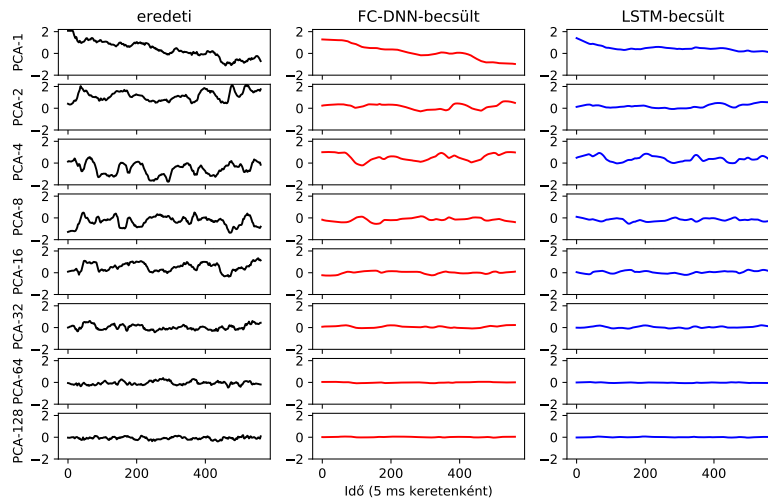
3. Eredmények

A fenti modellek betanítása után beszédet és ultrahangvideót szintetizáltunk az adatbázisok tesztelésre félretett mondatainak szövegei alapján. A validációs és teszthiba mérésére egyrészt spektrális jellemzőkkel kapcsolatos hibát (mel-kepsztrális távolság, Mel-Cepstral Distance, MCD), másrészt artikulációs jellemzőkkel kapcsolatos hibát (átlagos négyzetes hiba, Root Mean Square Error, RMSE) számoltunk. A gépi tanulás során mind az időtartam, mind az akusztikai és artikulációs modelleket betanítottuk, de a hibaszámításokhoz a tesztmondatokat az eredeti időzítésükkel szintetizáltuk. Így a hibamértékek kiszámításához nem volt szükség a jellemzők időbeli vetemítésére.

3.1. Demonstrációs minták

A becsült artikulációs jellemzőkre a 3. ábrán mutatunk egy példát ('01fi' beszélő). Mivel az artikulációs adatokat 128-dimenziós PCA jellemzőkként reprezentáljuk a gépi tanulás során, az ábrának nincs kézzelfogható, artikulációs mozgásokkal összevethető értelmezése. Amit érdemes megfigyelni, hogy az FC-DNN hálózattal becsült és az LSTM hálóval becsült görbék is követik az eredeti artikulációs jellemzők tendenciáit, de a finom részletek a DNN reprezentáció során kisimultak, eltűntek. Ez a túlsimítás jelenség gyakori a statisztikai parametrikus beszédszintézis (HMM-TTS és DNN-TTS) esetén is. Az alacsonyabb számú komponensek (pl. PCA-1, PCA-2, PCA-4) még tartalmazznak az eredeti adathoz hasonló részeket, de a magasabb dimenziójú komponensek (pl. PCA-64, PCA-128) közel konstansak, azaz utóbbiakat a neurális hálózatok nem tudták jól modellezni. Összességében a fenti ábra azt mutatja, hogy a szintetizált nyelvultrahangadatok a nagyobb változásokat (pl. a nyelv vízszintes mozgása) mutatják majd, de a finom részletek várhatóan elvesznek.

Ahhoz, hogy a generált artikulációs adatokat vizuálisan is megjelenítsük, a PCA adatokból nyelvultrahangképeket generáltunk, hiszen ezeken már láthatóvá válik a nyelv mozgása és alakja a szintetizált mondatok esetén is. A 4. ábrán az eredeti videókból és a szintetizáltakból néhány ultrahangképkockát ábrázoltunk az idő függvényében. A '01fi' beszélő esetén a bal oszlopban (eredeti-PCA)

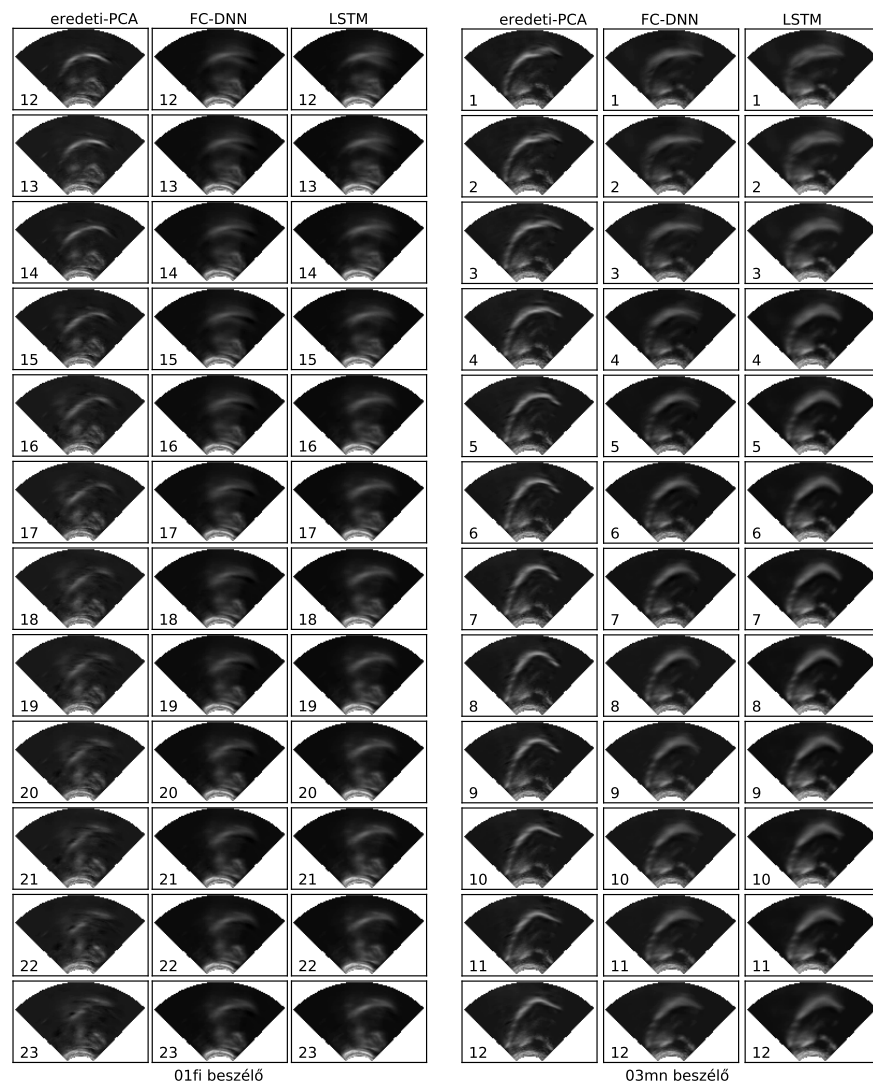


3. ábra. Eredeti és becsült artikulációs jellemzők PCA reprezentációban, '01fi' beszélőtől. Mondat: *"I leave it to nobody," said Shakespeare, sulkily.*

láthatjuk, hogy jelentős nyelvmozgás van, azaz a nyelvcsúcs (jobb oldalon) az idő múlásával feljebb mozdul: míg a kezdő, 12. képkockán még körív jellegű a nyelv látható kontúrja, a 16. képkockától a nyelv első része emelkedni kezd, majd a 20. képkockánál eléri a legmagasabb helyzetet. Mind az FC-DNN, mind az LSTM hálózat előrejelzései követik az eredeti artikulációs mozgást, de a képek helyenként elkentek, és a nyelv kontúrjának éles láthatósága csökken – ismét a statisztikai túlsimítás eredményeként. A '03mn' beszélő esetében hasonló tendenciák figyelhetők meg: a nyelv mozgásának görbülete változik az idő függvényében az eredeti felvételen, de az FC-DNN és az LSTM által előrejelzett képeken a nyelv felülete nem olyan tiszta, mint az eredetiben. Mivel a nyelv szintetizált mozgása jobban látható a valós idejű videókban, néhány mintát elérhetővé tettünk: <http://smartlab.tmit.bme.hu/besztud2022>.

3.2. Objektív mérések

Az 1. táblázat összegzi az MCD eredményeket (kisebb MCD hiba jelenti azt, hogy a szintetizált beszéd közelebb van az eredeti beszédhez). Mivel az MCD értékek logaritmikus skálájúak, ezért az átlagokat az értékek tízes alapú



4. ábra. Eredeti és becsült artikulációs jellemzők nyelvultrahangképként ábrázolva. A bal alsó sarkokban lévő számok a videó képkockáját jelentik

1. táblázat. MCD hibák a validációs/teszt halmazon

Beszélő	MCD	
	FC-DNN	LSTM
01fi	6,995 / 6,971	6,647 / 6,588
02fe	6,095 / 5,803	6,486 / 6,259
03mn	5,781 / 5,785	5,977 / 5,948
04me	5,896 / 6,024	6,318 / 6,312
05ms	6,244 / 6,256	7,235 / 7,083
06fe	5,758 / 5,582	6,444 / 6,330
07me	6,589 / 6,562	6,831 / 6,749
09fe	6,516 / 6,844	7,197 / 7,472
átlag	6,440 / 6,486	6,821 / 6,861

exponenciálisa alapján számítottuk, majd logaritmusra alakítottuk vissza. A tesztmondatok MCD értékei FC-DNN esetén 5,8-7,0 dB (átlag: 6,5 dB), míg LSTM esetén 6,0–7,5 dB (átlag: 6,9 dB) között vannak, ami azt jelzi, hogy a rekurrens neurális hálózat nem javított az akusztikai jellemzők becslésében. A 6 dB körüli MCD értékek is arra utalnak, hogy az így szintetizált beszéd minősége és természetessége gyenge. Ennek az lehet az oka, hogy csak korlátozott méretű artikulációs-akusztikus adatbázisunk van (nagyjából 200 mondat minden beszélőre), ami túl kicsi az LSTM modell betanításához.

Az artikulációs jellemzőre számított RMSE hiba eredményeit a 2. táblázat foglalja össze (kisebb RMSE hiba jelenti azt, hogy a szintetizált artikulációs mozgás közelebb van az eredetihez). A legalacsonyabb hibát a '09fe' beszélő adataival érték el: az FC-DNN-nél 2,9, míg az LSTM-nél 3,1 a tesztadatokon mért hiba. A tendencia hasonló az MCD esetéhez: az LSTM hálózat nem javított az artikulációs jellemzők előrejelzésében, valószínűleg az adatbázisok kis mérete miatt.

Az objektív mérések konklúziója tehát az, hogy a jelen kutatáshoz rendelkezésre álló, beszélőnként kb. 200 mondatot tartalmazó adathalmazok esetén az

2. táblázat. ULTPCA/RMSE hibák a validációs/teszt halmazon

Beszélő	ULTPCA128/RMSE	
	FC-DNN	LSTM
01fi	3,292 / 3,223	3,319 / 3,208
02fe	3,533 / 3,732	3,753 / 3,904
03mn	3,147 / 3,660	3,289 / 3,680
04me	3,849 / 3,985	4,031 / 4,033
05ms	3,133 / 3,233	3,249 / 3,405
06fe	3,439 / 3,250	3,743 / 3,451
07me	3,544 / 3,595	3,498 / 3,461
09fe	3,022 / 2,864	3,234 / 3,133
átlag	3,370 / 3,443	3,515 / 3,534

egyszerű FC-DNN hálózattal jobb eredményt lehet elérni, mint a bonyolultabb LSTM hálózattal.

4. Összefoglalás és következtetések

A fenti kísérletekben bemutattuk, hogy a szöveg alapján történő nyelvultrahangvideó-előrejelzés megvalósítható a hagyományos DNN-alapú beszéd-szintézis kiterjesztéseként, a viszonylag kis mennyiségű tanítóadat ellenére. Bár a vizuális és a beszédkimenet közötti szinkronizálást a modell nem kényszeríti ki semmilyen módon, a DNN tanítása során az akusztikai és artikulációs jellemzők összekapcsolása biztosítja, hogy a beszéd és a vizuális jellemzők szinkronban legyenek, azaz a generált ultrahangos videóknak a nyelv megfelelően mozogjon a szintetizált beszédhez képest. Az akusztikai és artikulációs paraméterek összekapcsolása a gyakorlatban azt jelenti, hogy a tanításuk együttesen történik.

Bár korábban több kísérlet is történt a gépi szövegfeldolvasás artikulációs adatokkal való kiterjesztésére, ezen vizsgálatok mindegyike EMA-t használt, amely egy pontkövető berendezés, és kevesebb térbeli információt tartalmaz a nyelvről, mint az ultrahang (Ling et al., 2010a,b; Wei et al., 2016; Steiner et al.,

2017; Le Maguer et al., 2017; Yu et al., 2019). Az ultrahang előnye ebben a kontextusban, hogy az eredményül kapott videó a nyelv nagyobb részét mutatja az EMA-hoz képest.

A szöveges bemenetből származó artikulációsmozgás-előrejelzés hasznos lehet az audiovizuális beszédszintézisben. Egy konkrét alkalmazás a számítógéppel segített nyelvtanulás/számítógéppel segített kiejtésgyakorlás (Katz et al., 2014; Jones, 2017; Agarwal & Chakraborty, 2019), amely hasznos lehet az idegen nyelvet tanulók számára. Egy ilyen kombinált TTS és szöveg-artikuláció becselő rendszerrel tetszőleges bemeneti szöveg megadása után hallhatóvá válik a beszéd, és ezzel szinkronban láthatóvá válik (2D-ben vagy 3D-ben), hogyan kell a nyelvet mozgatni a célhangok képzéséhez. Ez a vizuális visszajelzés hasznos lehet az idegen nyelvek kiejtésének tanulásában, különösen akkor, ha a cél nyelv nehezen artikulálható (pl. a beszélő anyanyelvétől lényegesen különböző) beszédhangokat tartalmaz.

A forráskódok, a Merlin receptek, és a betanított modellek elérhetőek a következő oldalon: <https://github.com/BME-SmartLab/txt2ult>.

Köszönetnyilvánítás

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal OTKA programja (FK 142163 projekt), az MTA Bolyai János kutatói ösztöndíja, valamint az Új Nemzeti Kiválóság Program Bolyai+ (ÚNKP-22-5-BME-316) pályázata támogatta.

A kutatást az MTA Bolyai János kutatói ösztöndíja támogatta. A kutatás az Innovációs és Technológiai Minisztérium ÚNKP-21-5 kódszámú (azonosító: ÚNKP-21-5-BME-352) Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült. A mély neuronhálós kísérletekhez használt Titan X GPU az NVIDIA Corporation adománya. Köszönjük a CSTR kutatócsoportnak a Merlin eszköz és az UltraSuite-TaL adatbázis rendelkezésre bocsátását.

Hivatkozások

- Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies*, 24, 3731–3743. URL: <https://doi.org/10.1007/s10639-019-09955-7>. doi:10.1007/s10639-019-09955-7.
- Arthur, F. V., & Csapó, T. G. (2021). Szájról olvasás automatizálása mély neurális hálózatok és mobilalkalmazás-kezelőfelületet alkalmazásával. *Beszédtudomány - Speech Science*, 2, 7–23. doi:10.15775/Besztud.2021.7-23.
- Arthur, F. V., & Csapó, T. G. (2022). Deep learning alapú agyi jel feldolgozás és beszédszintézis előkészítő munkálatai. In *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY)* (pp. 185–198). online.
- Csapó, T. G. (2021). Extending Text-to-Speech Synthesis with Articulatory Movement Prediction using Ultrasound Tongue Imaging. In *Proc. ISCA SSW11* (pp. 7–12). Budapest, Hungary. doi:10.21437/SSW.2021-2. arXiv:2107.05550.
- Csapó, T. G., & Csopor, D. (2015). Ultrahangos nyelvkontúr követés automatikusan: a mély neuronhálókra alapuló AutoTrace eljárás vizsgálata. *Beszédkutatás*, 23, 176–186.
- Csapó, T. G., Deme, A., Grácsi, T. E., Markó, A., & Varjasi, G. (2017a). Szinkronizált beszéd- és nyelvultrahang-felvételek a SonoSpeech rendszerrel. In *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY)* (pp. 339–346). Szeged.
- Csapó, T. G., Grósz, T., Tóth, L., & Markó, A. (2017b). Beszédszintézis ultrahangos artikulációs felvételekből mély neuronhálók segítségével. In *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY)* (pp. 181–192). Szeged.
- Csapó, T. G., & Lulich, S. M. (2015). Error analysis of extracted tongue contours from 2D ultrasound images. In *Proc. Interspeech* (pp. 2157–2161). Dresden, Germany.

- Csapó, T. G., & Németh, G. (2011). Prozódiai változatosság rejtett Markov-modell alapú szövegfelolvasóval. In *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY)* (pp. 167–177). Szeged.
- Czap, L., & Mátyás, J. (2003). Beszélő fej. In *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY)* (pp. 196–201). Szeged. URL: http://acta.bibl.u-szeged.hu/59393/1/msznykonf_001_196-201.pdf.
- Czap, L., & Mátyás, J. (2005). Hungarian Talking Head. In *Proceedings of Forum Acusticum 4th European Congress on Acoustics* (pp. 2655–2658). Budapest, Hungary.
- Hajjej, N., & Csapó, T. G. (2020). Realistic Ultrasound Tongue Image Synthesis using Generative Adversarial Networks. *Beszédtudomány - Speech Science*, 1, 7–21. doi:10.15775/Besztud.2020.7-21.
- Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Rousset, P., & Stone, M. (2007). Eigentongue feature extraction for an ultrasound-based silent speech interface. In *Proc. ICASSP* (pp. 1245–1248). Honolulu, HI, USA.
- Jones, D. (2017). *Development of Kinematic Templates for Automatic Pronunciation Assessment Using Acoustic-to-Articulatory Inversion*. Master's thesis Marquette University. URL: https://epublications.marquette.edu/theses_open/433.
- Katz, W., Campbell, T., Wang, J., Farrar, E., Eubanks, J., Balasubramanian, A., Prabhakaran, B., & Rennaker, R. (2014). Opti-speech: A real-time, 3D visual feedback system for speech training. In *Proc. Interspeech* (pp. 1174–1178). Singapore, Singapore.
- Le Maguer, S., Steiner, I., & Hewer, A. (2017). An HMM/DNN comparison for synchronized text-to-speech and tongue motion synthesis. In *Proc. Interspeech* (pp. 239–243). Stockholm, Sweden. doi:10.21437/Interspeech.2017-936.

- Ling, Z.-H., Richmond, K., & Yamagishi, J. (2010a). An Analysis of HMM-based prediction of articulatory movements. *Speech Communication*, 52, 834–846. URL: <http://linkinghub.elsevier.com/retrieve/pii/S0167639310001147>. doi:10.1016/j.specom.2010.06.006.
- Ling, Z.-H., Richmond, K., & Yamagishi, J. (2010b). HMM-Based Text-to-Articulatory-Movement Prediction and Analysis of Critical Articulators. In *Proc. Interspeech* (pp. 2194–2197). Makuhari, Japan. URL: https://www.isca-speech.org/archive/interspeech_2010/i10_2194.html.
- Massaro, D. W., Cohen, M. M., Tabain, M., Beskow, J., & Clark, R. (2012). Animated speech: research progress and applications. In G. Bailly, P. Perrier, & E. Vatikiotis (Eds.), *Audiovisual Speech Processing* (pp. 309–345). Cambridge, UK: Cambridge University Press. URL: <http://ebooks.cambridge.org/ref/id/CB09780511843891A024>. doi:10.1017/CB09780511843891.014.
- Perrier, P. (2014). "GEPPETO": A target-based model of speech production including optimal planning and physical modeling. In *Adventures in Speech Science*. Tokyo, Japan. URL: <https://hal.archives-ouvertes.fr/hal-01057251>.
- Rácz, B., & Csapó, T. G. (2020). Ajakvideó alapú beszéd-szintézis konvolúciós és rekurrens mély neurális hálózatokkal. *Beszédtudomány – Speech Science*, 1, 57–72. doi:10.15775/Besztud.2020.57–72.
- Ribeiro, M. S., Sanger, J., Zhang, J.-X. X., Eshky, A., Wrench, A., Richmond, K., & Renals, S. (2021). TaL: a synchronised multi-speaker corpus of ultrasound tongue imaging, audio, and lip videos. In *2021 IEEE Spoken Language Technology Workshop (SLT)* (pp. 1109–1116). Shenzhen, China. URL: <https://arxiv.org/abs/2011.09804>. doi:10.1109/SLT48900.2021.9383619. arXiv:2011.09804.
- Schabus, D., Pucher, M., & Hofer, G. (2014). Joint audiovisual Hidden Semi-Markov Model-based speech synthesis. *IEEE Journal on Selected Topics in*

- Signal Processing*, 8, 336–347. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6589946>. doi:10.1109/JSTSP.2013.2281036.
- Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Chen, Z., Zhang, Y., Wang, Y., Skerrv-Ryan, R., Saurous, R. A., Agiomvrgiannakis, Y., & Wu, Y. (2018). Natural TTS Synthesis by Conditioning Wavenet on MEL Spectrogram Predictions. In *Proc. ICASSP* (pp. 4779–4783). Calgary, Canada. doi:10.1109/ICASSP.2018.8461368. arXiv:1712.05884.
- Stavness, I., Nazari, M. A., Cormac, F., Perrier, P., Payan, Y., Lloyd, J., & Fels, S. (2014). Coupled Biomechanical Modeling of the Face, Jaw, Skull, Tongue, and Hyoid Bone. In N. Magnenat-Thalmann, O. Ratib, & H. F. Choi (Eds.), *3D Multiscale Physiological Human* (pp. 253–274). Springer London. doi:10.1007/978-1-4471-6275-9_11.
- Steiner, I., Le Maguer, S., & Hewer, A. (2017). Synthesis of Tongue Motion and Acoustics from Text Using a Multimodal Articulatory Database. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 25, 2351–2361. doi:10.1109/TASLP.2017.2756818. arXiv:1612.09352.
- Stone, M., Sonies, B., Shawker, T., Weiss, G., & Nadel, L. (1983). Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system. *Journal of Phonetics*, 11, 207–218.
- Tóth, B. P. (2013). *Rejtett Markov-modell alapú gépi beszédkeltés*. Phd thesis BME TMIT.
- Tóth, B. P., & Németh, G. (2008). Rejtett Markov-modell alkalmazása magyar nyelvű gépi szövegfelolvasóhoz. *Beszéd kutatás*, 16, 182–193.
- Wei, Z., Wu, Z., & Xie, L. (2016). Predicting articulatory movement from text using deep architecture with stacked bottleneck features. In *Proc. APSIPA* (pp. 1–6). Jeju, South Korea. doi:10.1109/APSIPA.2016.7820703.
- Whalen, D. H., Kang, J., Iwasaki, R., Shejaeya, G., Kim, B., Roon, K. D., Mark, K., Tiede, Preston, J., Phillips, E., McAllister, T., & Boyce, S. (2019).

- Accuracy assessments of hand and automatic measurements of ultrasound images of the tongue. In *Proc. ICPHS* (pp. 542–546). Canberra, Australia.
- Wu, Z., Watts, O., & King, S. (2016). Merlin: An Open Source Neural Network Speech Synthesis System. In *9th ISCA Speech Synthesis Workshop* (pp. 202–207). Sunnyvale, CA, USA. doi:10.21437/ssw.2016-33.
- Yu, L., Yu, J., & Ling, Q. (2019). BLTRCNN Based 3D Articulatory Movement Prediction: Learning Articulatory Synchronicity From Both Text and Audio Inputs. *IEEE Transactions on Multimedia*, 21, 1621–1632. doi:10.1109/TMM.2018.2887027.
- Zainkó, C., Tóth, B. P., & Németh, G. (2017). Magyar nyelvű WaveNet kísérletek. In *Magyar Számítógépes Nyelvészeti Konferencia (MSZNY)*. Szeged.
- Zen, H., Senior, A., & Schuster, M. (2013). Statistical parametric speech synthesis using deep neural networks. In *Proc. ICASSP* (pp. 7962–7966). Vancouver, Canada. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6639215>. doi:10.1109/ICASSP.2013.6639215.

A FilmHír adatbázis fejlesztése régi filmhíradókban elhangzó beszéd vizsgálatához

Gocsál Ákos^{1,2}

¹*PTE Művészeti Kar, Zeneművészeti Intézet*

²*ELKH Nyelvtudományi Kutatóközpont*

Abstract

This paper presents the first steps of the development of a speech database that supports the research of the spoken material in the early sound newsreels produced in Hungary. Since there was no recording protocol, and the recordings were not made in a controlled environment, the development of this database differs from that of other, purpose-built databases, such as the BEA Hungarian Spontaneous Speech Database, in several ways. There is no balanced selection of speakers or speech styles, and the speech material may contain elements unpredictable to the developer. In the first phase of development, speech materials of 140 newsreels, made in 1931 and 1932, were extracted and annotated (only pause-to-pause sections). A table was created with basic metadata. It was revealed that the newsreels contain a variety of speech styles, such as conversations, public speeches, narrations, commands, and interviews. In the second phase, further metadata were added, including background noises, music, utterances in foreign languages, distortions, etc. Research experiences and possibilities offered by this database are also discussed as well as directions for its further development.

1. Bevezetés

Ebben a tanulmányban egy olyan adatbázis összeállításával kapcsolatos eddigi eredményeket és tapasztalatokat ismertetjük, amely a magyar filmhíradók – nyelvészeti, fonetikai szempontból mindeddig feldolgozatlan – beszédanyagának kutatását teszi lehetővé.

A nemzetközi és a hazai szakirodalomban is sok évtizedes hagyománya van a fonetikai kutatásokat segítő adatbázisok, korpuszok használatának. Ezek többféle szempont alapján is jellemezhetők és csoportosíthatók. Néhány példát idézünk a magyar szakirodalomból. Életkor, nem, iskolai végzettség szerint

Email address: `gocsal.akos@pte.hu` (Gocsál Ákos)

nagy változatosságot mutató felnőtt beszélők közreműködésével készült a magyar spontánbeszéd-adatbázis (BEA), amely többféle beszéd típus tanulmányozását is lehetővé teszi (Gósy et al., 2012). Más adatbázisok specifikus csoportok, például gyerekek, tinédzserek beszédének jellegzetességeinek kutatását segítik, magyarra a GABI (Bóna et al., 2014) és a Tini BEA (Gyarmathy & Neuberger, 2015) adatbázis említendő. Más adatbázisok fonációs rendellenességgel beszélők, depressziósok, Parkinson-kórosok beszédét tartalmazzák (Sztahó et al., 2019), de ismeretesek olyanok is, amelyeket speciális célokra, például zajos, vagy érzelmes beszédet automatizáltan felismerő, feldolgozó módszer fejlesztéséhez állítottak össze (Sztahó et al., 2011). Szintén hazai fejlesztők munkája a HuComTech multimodális korpusz, amely a hanggal együtt videofelvételeket is tartalmaz a beszédeseményekről. Az igen gazdag, 47 szintű annotáció az akusztikai és nyelvi jellemzők mellett a nemverbális jelek lejegyzését – testtartás, gesztusok, mimi-ka, tekintet stb. – is tartalmazza, így az egyes közlések, társalgások még tágabb kontextusban kutathatók, mint ha csak a hang állna rendelkezésre (Hunyadi et al., 2018).

Az érzelmes beszéd kutatásához készített adatbázisokkal kapcsolatban egy igen lényeges kérdéssel találkozott a kutatók. Felmerült ugyanis, hogy jelentősen eltérnek egymástól a mesterséges – azaz eljátszott vagy tudatosan előidézett – helyzetben, illetve valós körülmények között, spontán módon keletkezett adatok (Campbell, 2000; Campbell et al., 2006), így ennek megfelelően az adatbázisok is két fő csoportra – „mesterséges” és „valós életbeli” – oszthatók (Pittermann et al., 2010, 42). Hasonló, de további részletekkel kiegészített felosztást közölt Aubergé et al. (2003). Ők az érzelmes beszédet tartalmazó korpuszokat három, egymásra merőleges dimenzió szerint osztályozták. A korpuszok készülhetnek *in vivo* vagy *in vitro* módszerekkel. Az előbbi a kutatói beavatkozás nélküli, az utóbbi a laboratóriumi, tudatosan beállított körülmények közötti adatrögzítésre vonatkozik. A második jellemző a beszédjellemzők – interakciós helyzet, nyelvi tartalom, fonetikai tulajdonságok – fölötti kutatói kontroll jelenléte vagy hiánya. A harmadik jellemző pedig az, hogy a beszélők eljátszák-e az adott érzelmet, vagy pedig a beszéd autentikusan tartalmazza-e azt. Pittermann

et al. (2010, 43) mindezek mellett további csoportosítási szempontként megemlíti, hogy professzionális színészek vagy naiv beszélők bemondásában készült-e a felvétel, illetve maga a nyilatkozat tartalmaz-e érzelmi töltetet vagy szemantikailag semleges-e.

A mesterséges vagy valós életbeli beszédesemények használata, és az ebből adódó különbségek problematikája azonban túlmutat az érzelmes beszédből épített adatbázisokon. Más céllal is állítanak össze valós életből vett beszédekbe – például parlamenti vitákból (Solberg & Ortiz, 2022), telefonos ügyfélszolgálatok beszélgetéseiből (Vicsi & Sztahó, 2009) vagy televíziós műsorokból (Teleki et al., 2005; Lleida et al., 2020) – adatbázisokat, és rendszeresen felmerülő kérdés, hogy mennyire tekinthetők természetesnek az adatbázisba beemelt beszédmin-ták. Campbell (2000) szerint a televíziós műfajok esetében a hírműsorokban bemutatott interjúk általában természetesebbek, mint a talk-show-típusú műsorokban és valóságshow-kban elhangzó beszédek, ugyanis ezekben az esetekben előfordulhat a helyzetek tudatos megjátszása. Grimm et al. (2008) viszont beszámolnak egy érzelmes beszédet tartalmazó, német talkshowadásokból összeállított korpuszról, amely előre nem megírt, a műsor vendégei között megvalósult autentikus társalgásokat tartalmaz. A természetesség problémájával foglalkozott még Devillers & Vidrascu (2007) az általuk összeállított, 688, segélyhívó vonalon lezajlott beszélgetésből álló korpusz bemutatásakor. A szerzők rámutattak, hogy vészhelyzetben, amikor valaki mentőt hív, a beszédben tükröződő érzelmek természetesen adódnak, nem megjátszottak.

A valós életből vett, autentikus helyzetekben rögzített beszédesemények természetességének azonban „ára” van, mégpedig éppen a kontroll, azaz előzetesen meghatározott szempontok, egységes körülmények, felvételi protokoll hiánya, s ezt a kutatónak figyelembe kell vennie (Douglas-Cowie et al., 2003). Mivel ennek vannak különféle technikái, a kétféle – „mesterséges” és „valós életbeli” – adatokból felépített adatbázisok között nincs olyan éles határ, mint amilyenre a fentiek alapján következtetni lehetne. A kontrollált körülmények között összeállított adatbázisok fejlesztői számos olyan módszert alkalmaznak, amellyel a felvételi körülményeket közelítik az autentikus, természetes helyzetekéhez. Ide

sorolhatók például azok a meghatározott témából kiinduló, de kötetlen társalgások, amelyek jól modellálják a laboratóriumon kívül megvalósuló helyzeteket. Ilyen társalgásokat találunk a BEA adatbázisban (Gósy et al., 2012; Neuberger et al., 2014), vagy egy külföldi példát is említve, a Cieri et al. (2004) által összeállított telefonos adatbázisban, amelynek egymást nem ismerő beszélői egy véletlenszerűen kiválasztott témáról folytattak rövid beszélgetéseket. Ugyanakkor az előzetes kutatói kontroll, tudatos tervezés nélkül rögzített beszédmintákból is összeállítható olyan adatbázis, amelyek több szempontból is kontrolláltnak tekinthetők. Ilyenek például a rádióbemondók hírolvasásai, amelyeknél – még akkor is, ha a kutató nincs ilyen információk birtokában – feltételezhető, hogy minden esetben hangszigetelt helyiségben, professzionális eszközökkel, zavaró külső hatások nélkül történt a felolvasás rögzítése.

Mindezek alapján belátható, hogy elengedhetetlen feladat az adatbázis fejlesztésének dokumentálása, az egyes elvi megfontolások, döntések, fejlesztési szempontok precíz rögzítése és mindezek közzé tétele felhasználók számára. Jól példázza ezt a BEA adatbázis ismertetése (Gósy et al., 2012). Munkájukban a szerzők részletesen bemutatták, hogy a beszédanyagok rögzítése milyen körülmények között, milyen protokoll szerint történt. A beszélők az interjúkészítőtől többféle feladatot kaptak, így különféle beszéd típusok rögzítése vált lehetővé. A fejlesztők törekedtek a nők és a férfiak, illetve a különböző életkorú beszélők arányos megjelenítésére, továbbá tendenciaszerűen figyeltek az iskolázottság alakulására is. A hangrögzítést minden esetben ugyanabban a zajszigetelt helyiségben, professzionális felszereléssel végezték, így a felvételi körülmények minden adatközlőnél megegyeztek. Különbségek legfeljebb az egyes beszélők között lehettek. Ezeknek az információknak a birtokában a felhasználó megismerheti az adatbázis által nyújtott lehetőségeket, annak esetleges korlátait, így ezek tudatában választhatja ki a kutatási céljainak megfelelő beszédmintákat, értelmezheti kutatási eredményeit.

Az előzőekben felvázolt problémákkal, dilemmákkal a FilmHír adatbázis fejlesztésének kezdeti szakaszától kezdődően találkoztunk. A filmekben talált beszédanyag jelentős részét valós életbeli események során rögzítették, így a kuta-

tói kontroll hiánya, az előadódó adatok esetlegessége és rendezetlensége, a beszéd típusok sokfélesége miatt egyedi megoldást kellett kidolgozni ahhoz, hogy a tervezett adatbázis jól átlátható, könnyen lekérdezhető legyen. Az alábbiakban a fejlesztési folyamat részleteit mutatjuk be, amelyek nem csak a felhasználás és a továbbfejlesztés lehetőségeit vetítik előre, hanem kiindulópontként szolgálhatnak más, hasonlóan nem kutatási céllal rögzített beszédanyagok adatbázisban való feldolgozásához is.

2. A FilmHír adatbázis felépítése

A magyar filmhíradók az interneten bárki számára szabadon hozzáférhetők (<https://filmhiradokonline.hu/>). A filmhíradókon végzett kutató-fejlesztőmunkát megelőzően tájékoztunk azok felhasználhatóságáról. A honlapon közzétett tájékoztató (<https://filmhiradokonline.hu/intro.php>), illetve a Magyar Nemzeti Filmintézet munkatársával történt konzultáció megerősítette, hogy oktatási, kutatási, közismereti célra felhasználhatók a Filmhíradók Online anyagai.

Az első filmhírek előzetes megtekintése során világossá vált, hogy az adatbázis fejlesztési feladatai és alkalmazási lehetőségei sok szempontból jelentősen eltérnek a tervezett, protokoll szerint rögzített adatbázisokéitól. Négy olyan szempontot emelünk ki, amelyek mentén a legjelentősebb különbségek megfogalmazhatók:

1. A FilmHír adatbázis felépítése nem előre meghatározott protokoll szerint történik, így számítani kell arra, hogy véletlenszerűen bármilyen beszéd típus és bármilyen demográfiai jellemzővel rendelkező beszélő előfordulhat. Ebből adódóan a beszélőket, beszédmintákat illetően nem beszélhetünk reprezentativitásról, de még kiegyensúlyozottságról sem.
2. A filmhíradóknál számos, a felvételek akusztikai minőségét kedvezőtlenül befolyásoló tényezővel kell számolni. A kezdetleges hangrögzítési technológia, a hangsávot ért sérülések, filmszakadások, torzulások, a különféle, ki nem szűrt háttérzajok következtében akusztikai szempontból is rendkívül heterogén a filmhírek beszédanyaga.

3. Maga a rögzítés is változatos körülmények között, nem kutatói kontroll mellett történt, emiatt a spontaneitás és természetesség szempontjából sem kezelhetők egységesen a beszédminták. Több filmben megfigyelhető, hogy a beszélők vélhetően nem is voltak tudatában annak, hogy hangjukat rögzítik, míg máskor kifejezetten a kamerába beszéltek. Más esetekben pedig, például az ünnepi szónoklatok esetében az feltételezhető, hogy akkor is ugyanígy, vagy nagyon hasonló módon valósult volna meg a beszédesemény, ha a híradó nem rögzítette volna. Ezekben az esetekben tehát a beszédesemények a rögzítés tényétől függetlenül adottak voltak, autentikus, természetes helyzetekben valósultak meg, a kamera csak mellékesen volt jelen.
4. A meghatározott protokoll alapján készült, előzetesen megtervezett adatbázisok általában tetszőlegesen bővíthetők további beszélőkkel vagy beszéd típusokkal. A FilmHír számára azonban véges számú felvétel áll rendelkezésre. A fejlesztés során azonban törekszünk a teljességre, azaz valamennyi fellelhető filmhír beszédanyagának feldolgozására. Adatbázisunk ebből a szempontból hasonlít az ómagyar korpuszra, amelyek fejlesztői célul tűzték ki, hogy korpuszuk valamennyi fennmaradt nyelvemléket tartalmazza, ugyanakkor ők is említették a reprezentativitással, kiegyensúlyozottsággal kapcsolatos, korábban említett problémát a korlátozott nyelvi anyag miatt (Simon & Sass, 2012).

Az adatbázis felépítéséhez a filmhírek feldolgozását kronologikus sorrendben terveztük, így a jelen állapotban a legkorábbi, 1931. szeptember – 1932. december közötti hangos híradókkal foglalkozunk. Az egyes hírek megkülönböztetése két számadattal történt. Az első számadat az adott híradó (általában 6–9 hírből álló hírblokk) száma, a második pedig a híradón belüli filmhír száma. A legelső hangos hír ennek megfelelően a 398/1. számmal azonosítható. A teljes azonosító tartalmaz még néhány karaktert (MVH, Magyar Világhíradó), azonban ezt nem mindenhol tüntettük fel, mivel 1944 októberéig minden híradó esetében azonos volt.

Az adatbázis felépítésének volt egy első, feltáró jellegű fázisa, amely konkrét kutatási feladathoz szolgáltatott adatokat. A régi filmhíradókban hosszú évtizedeken keresztül alkalmaztak bemondókat a hírek felolvasására, így kézenfekvőnek tűnt a hírolvasás tempójának vizsgálata. A beszéd tempóértékei egyébként is könnyen vizsgálhatók más célból készített felvételeken, mivel tempó minden beszédeseményben jelen van, szemben egyes ritkán előforduló nyelvi struktúrákkal, továbbá előhívásukhoz nem szükségesek külön kutatói instrukciók, illetve viszonylag kevés annotációs munkát igényelnek (Seifart, 2021). A legkorábbi, 1931-32-ben készült hangos filmhírek megtekintése során azonban bebizonyosodott, hogy azok igen heterogén beszédanyagot tartalmaznak, a narrációs szövegek felolvasásán kívül számos más beszédtypus is előfordul bennük. Ezért az egyes bemondások tempóértékein kívül azonosítottuk a beszédtypusokat, és más adatokat is rögzítettünk, amelyekből világossá vált, hogy egy így felépített adatbázis számos további kutatási feladat megoldására lehet még alkalmas (Gocsál, 2022).

A fejlesztés jelenleg is zajló második fázisában a teljes eddigi adatbázist átvizsgáltuk és kiegészítettük olyan adatokkal, amelyek segítségével jóval pontosabban szűrhető az adatbázis, így kutató-fejlesztő feladatokhoz célzottan kereshetők beszédminták.

A két fázis ismertetéséhez Drude et al. (2014) munkáját vesszük alapul, akik a korpuszfejlesztés során keletkező adatok három típusát különböztették meg. Az elsődleges adatok maguk a hang- vagy videófelvevételek. A másodlagos adatokat az annotáció során rögzített szövegek, illetve olyan időfüggő szimbólumok alkotják, amelyek a vizsgált beszédminta releváns tulajdonságait kódolják. A harmadik adattípushoz pedig a metaadatok tartoznak. Ezek az adatok az elsődleges és a másodlagos adatok nem időfüggő, általános tulajdonságait jelentik, például a rögzített esemény mi volt, hol, mikor, történt, ki beszél rajta stb., de ide tartozik a felvétel formátuma, kódolása, struktúrája, forrása, azonosító kódja és a helye az adatbázisban.

2.1. A fejlesztés első fázisa

2.1.1. Az elsődleges adatok

Az adatbázist alkotó elsődleges adatok maguk a filmhíradók. Mivel az adatbázis célja elsősorban a nyelvi-kommunikációs jelenségek tanulmányozásának támogatása, a filmhíradókból kinyerhető, megállapítható adatoknak csak egy meghatározott, az említett kutatási terület szempontjából releváns részhalmazával foglalkozunk. Értelemszerűen ide tartoznak a beszédet tartalmazó hangszávok, de a kommunikatív aktust kísérő, és a filmen megjelenített minden olyan szöveges és képi információ is, amely segítségével megállapítható az adott beszédeseemény kontextusa, azonosítható a beszéd típusa, esetleg a beszélő. Az adatbázis felépítése során fontos teendő volt az információhordozó egységek meghatározása, amelyek az adatbázis egyes rekordjait alkotják. Más kutatásban az egyedi híreket tekintették egységnek (Pincemin et al., 2020), azonban a jelen esetben ez egyelőre nem tűnt célszerűnek, ugyanis egy híren belül gyakran többféle beszédeseemény is megvalósult, és több beszélő is megszólalt. Így az adatbázis számára definiáltuk a „közlési egység” fogalmát, azaz az (1) egy filmhíren belül (2) ugyanazon beszélő (3) ugyanazon beszéd típusához tartozó bemondását tekintettük az adatbázis elemi egységének, azaz rekordnak.

Döntést igényelt az is, hogy szerepeljen-e minden előforduló beszédeseemény az adatbázisban. A fejlesztés első fázisában kizártuk az idegen nyelvű és a csoportosan, tömegesen mondott szövegeket (pl. éljenzés), illetve a gyenge technikai minőségű, érthetetlen vagy erősen zajos bemondásokat, tehát csak a viszonylag tiszta, érthető bemondásokkal foglalkoztunk.

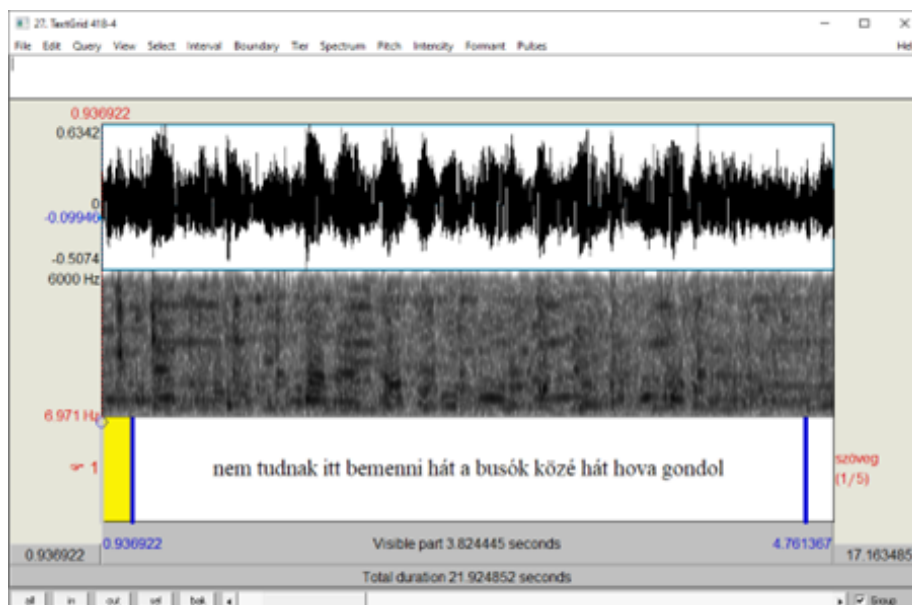
A vizsgált időszakból 61 híradó 441 filmhíre maradt fent. Ezekben a fenti elvek alapján összesen 258 közlési egységet találtunk. A filmhírek beszédet tartozó részleteit külön hangfájlokban, *.wav* formátumban elmentettük. Ezek formátuma *xxx-y.wav* volt, ahol *xxx* a híradó száma, *y* pedig a híradón belüli filmhír száma volt.

2.1.2. A másodlagos adatok

Az egyes közlési egységek annotálását a *Praat* programban végeztük (Borsma & Weenink, 2019). A fejlesztés első fázisában csak a szünettől szünetig tartó beszédszakaszokat jelöltük, és csak elhangzó szövegeket jegyeztünk le, felhasználva, de több esetben kiegészítve, korrigálva a filmhiradokonline.hu oldalon közzölt szövegeket. Ahol filmszakadás miatt ugrás volt az elhangzó szövegben, ott megjelöltük a szakadás előtti, illetve utáni, egészben elhangzó szó határait, ezzel a szakadást tartalmazó rövid részt kihagytuk a további feldolgozásból. A szöveg lejegyzéshez a BEA spontánbeszéd-adatbázis annotációjának első változatát vettük alapul (Gyarmathy et al., 2014), azaz fonológiai alapon, a helyesírási szabályoknak megfelelően rögzítettük az elhangzó beszédet. Az idegen szavakat, illetve a feltűnően sajátos ejtésű magyar szavakat kiejtés szerint írtuk, nagybetűket és írásjeleket nem használtunk. Ez az egyszerűsített lejegyzés már lehetővé tette a beszédszakaszok időtartamainak és az elhangzó beszédhangok számának meghatározását, így első kutatási feladatunkhoz az artikulációs tempó értékeit ki lehetett számítani. Ugyanakkor részletesebb annotálásra egyelőre nem vállalkoztunk, részben annak időigényessége miatt, részben pedig azért, mert nem ismerve a beszédanyagot, bizonytalan volt, hogy milyen mélységben, miket érdemes egyáltalán lejegyezni. Az 1. ábrán látható a 418/1. sz. filmhírből kiemelt *Nem tudnak itt bemenni, hát a busók közé, hát hova gondol* mondat megjelenítése. Az oszcillogram és a spektrogram is jól mutatja az igen erős háttérzaj jelenlétét, azonban a szöveg érthetőségét ez nem akadályozta. A beszédminőségnek ez a szakasza olyan kutatási feladatokra jól használható, amelyek nem igénylik az alacsony zajszintet, vagy épp ellenkezőleg, kifejezetten szükségesek hozzá, például beszédfelismerő rendszerek teszteléséhez.

2.1.3. A metaadatok

Ibrus & Ojamaa (2018) szerint a metaadatok a filmhírt és az általa megjelenített történeti eseményeket kulcsszavakból álló verbális nyelvvé fordítják át. A metaadatok sajátos metanyelvet alkotnak. Ez a metanyelv azonban nem szigorúan standardizált elemekből épül fel, ugyanis ebben az esetben nem lehetne



1. ábra. A 418/4. sz. filmhír lejegyzésének részlete.

új jelentéseket konstruálni, illetve magát az adatbázist sem lehetne fejleszteni különféle kutatási célokra. Mindebből az következik, hogy egy filmhír-adatbázis fejlesztésekor a metaadat-állomány létrehozása nagyban függ a fejlesztő szubjektív megítélésétől – mit tart fontosnak –, illetve attól, hogy a későbbi felhasználók igényeiről hogyan gondolkodik. A metaadatoknak számos típusa létezik. Egy más céllal, a spanyol köztelevízió adásaiból összeállított adatbázis fejlesztésekor rögzítették a beszélőket, beszédfordulókat, továbbá azt is, hogy a felvétel városi vagy falusi környezetben, illetve belső térben vagy külső helyszínen készült, osztott képernyő (több kamera képe is látszik-e), meleg vagy hideg időjárás volt, illetve nappal vagy éjszaka készült-e a felvétel (Lleida et al., 2020).

A jelen adatbázis metaadat-struktúrájának kialakításakor a korábbiakhoz hasonló okok miatt egyszerű, de kutatási célra már használható rendszert igyekeztünk kialakítani. Egy Excel-táblában minden egyes, közlési egységre vonatkozóan az alábbi adatokat rögzítettük:

- a közlési egység azonosítói
 - sorszáma az adattáblában
 - a honlapon közölt azonosító
 - a filmhíradó száma
 - a filmhír száma
- a filmhír megjelenésének ideje
 - évszám
 - hónap
- a beszélő személy
 - neve, ha ismert
 - születési éve, ha ismert
 - életkora, ha ismert
- az annotálás állapota
 - x, ha elkészült
- beszéd típus
- a filmhír témája néhány szóban
- megjegyzés
- link a filmhírhoz

A beszéd típus feltüntetése csak első megtekintés alapján történt. A megjegyzéseknél azokat a lényeges információkat tüntettük fel, amelyek nem szerepeltek külön oszlopban, például azt, ha gyermek volt a beszélő, vagy hogy feltételezhetően megtanult szöveget mondott el.

Az első fázis során 441 filmhírt tekintettünk meg, ezek közül 140-ben találtunk feldolgozható beszédeseményeket. Összesen 258 közlési egységet adatoltunk a fenti módszer alkalmazásával, ezek szünetek nélküli, összesített időtartama 3221,37 másodperc volt.

2.2. A fejlesztés második fázisa

A tizenöt hónapnyi híryanagy feldolgozása során számos olyan megfigyelést tettünk és tapasztalatot szereztünk, amely birtokában felmerült, hogy érdemes a teljes adatbázist újragondolni, más elvek alapján felépíteni.

2.2.1. Az elsődleges adatok

Az első fázisban kizárólag a jól érthető, magyar nyelvű beszédrészleteket emeltük ki a filmhírekből, és ezekből állítottuk össze az adattáblát. A fejlesztés során azonban több olyan adatfajta is előadódott, amelyekről úgy véltük, hogy rögzítésük jelentősen bővítheti az adatbázis felhasználhatóságát. Ide tartozik például az idegen nyelvű bemondások, a különböző torzítások, hibák, vagy az aláfestő zenék, különféle háttérzajok megjelenítése. Így a második fázis során azt az alapelvet követtük, hogy az adatbázis minden filmhírral kapcsolatban tartalmazzon valamilyen információt, akkor is, ha beszéd nincs benne, mivel ez is informatív lehet a kutató számára. Felülvizsgálva az előző gyakorlatot, ebben az esetben az adatbázist alkotó egységnek, rekordnak az egyedi filmhírt tekintettük, a jövőbeni kutatóra bízva azt, hogy saját kutatási feladatának megfelelően kiemelje abból a szükséges közlési egységeket.

Ebben a fázisban kétféle hangfájl használtunk. A teljes híradót tartalmazó fájl *xxxx.wav* formátumban mentettük el, ahol *xxxx* a híradó száma volt, pl. *0405.wav*. A magyar beszédet tartalmazó filmhíreket pedig a könnyebb kezelhetőség miatt – mivel itt az elhangzó szövegeket részletesen lejegyeztük – külön fájlokban mentettük el *xxxx-y.wav* formátumban, ahol *y* az adott filmhír híradón belüli száma volt, pl. *0405-1.wav*.

2.2.2. A másodlagos adatok

A teljes híradókat tartalmazó hangfájlok annotálása két sávban történt. Az elsőben jelöltük a filmhír számát a híradón belül, a másodikban pedig a hallható hangeseményre jellemző címkéket jegyeztünk le az alábbiak szerint.

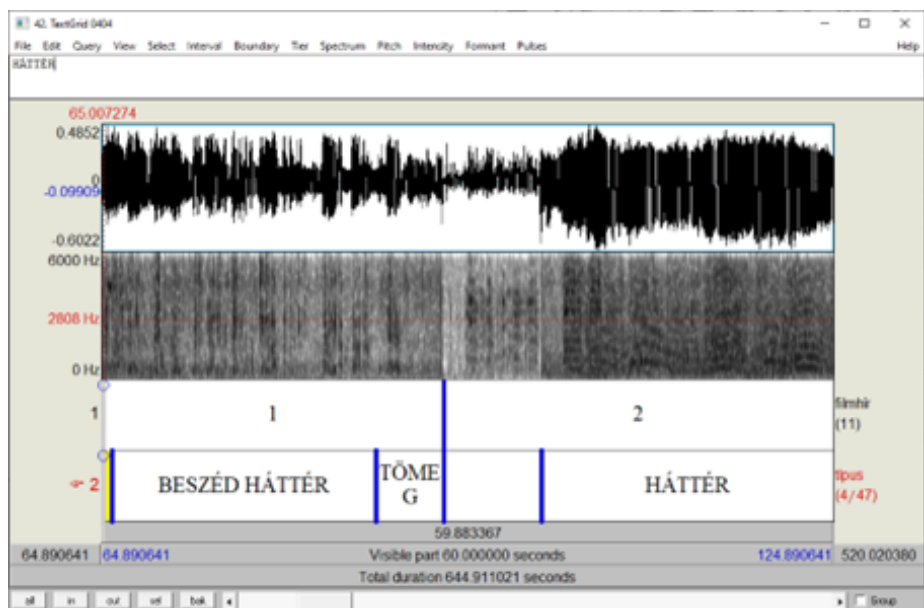
- BESZÉD: magyar nyelvű beszédesemény

- IDEGEN: idegennyelvű beszédesemény
- ÉNEK
- ZENE: a hír alatt megszólaló, a helyszínen rögzített vagy utólag alákevert zene (a hír főcíme alatti felvezető zenét nem jegyeztük le)
- TÖMEG: egyszerre, tömegesen mondott beszéd, pl. éljenzés, bravózás, skandálás
- TORZ: sérült, erősen zajos, torzult, érthetetlen felvétel
- SZAKAD: ugrás a hangsávban, filmszakadás következtében hiányzó rész miatt (ennek határait az előtte, ill. utána egészben elhangzó szó határaival jelöltük ki)
- HÁTTÉR: háttérzörejek, zajok

A teljes híradókhöz készült annotáció a hangokat nagyobb blokkokban, az elhangzó szöveg lejegyzése nélkül mutatja, és elsősorban a híradón belüli tájékozódást szolgálja, a megfelelő részlet kiválasztását segíti. Pontos mérések ezek finomításával a konkrét kutatói igények alapján lehetségesek. A 2. ábra szemlélteti mindezt a 404. sz. híradónál. Az 1. filmhírben Ripka Ferenc főpolgármester tart beszédet szűkebb körben, így a beszéd mellett az utcai zajok miatt a HÁTTÉR címkét is megjelenítettük. A TÖMEG címke a közönség éljenzésére utal. Ezt követően látható a 2. filmhír, amely rendőrlovak utcazajhoz szoktatásáról szól. Az elején nem jegyeztük le a főcím alatti zenét, ezért látható egy rövid kihagyás a 2. sávon. A HÁTTÉR címke itt is utcazajt jelez, de ebben a hírben nincs beszéd. Az első hírt külön fájlként (0404-1.wav) is elmentettük és a szövegét lejegyeztük, míg a másodikat nem, mivel abban nem volt beszéd.

A külön elmentett filmhírek beszédanyagát a korábbiakban ismertetett elvek alapján, azaz elsősorban a helyesírási szabályok szerint, de egyedi esetekben a kiejtést követve jegyeztük le. Alapesetben egy annotációs szintet és egy „egyéb” feliratú sávot alkalmaztunk, amelyben a beszéd alatti zenét, háttérzajt jeleztük. Egyszerre beszélés esetén további annotációs sávot adtunk hozzá.

A 3. ábra egy beszédet tartalmazó, külön elmentett, részletesebben annotált fájlt mutat. A 406/3. sz. hír arról tudósít, hogy dr. Sipőcz Jenő polgármes-

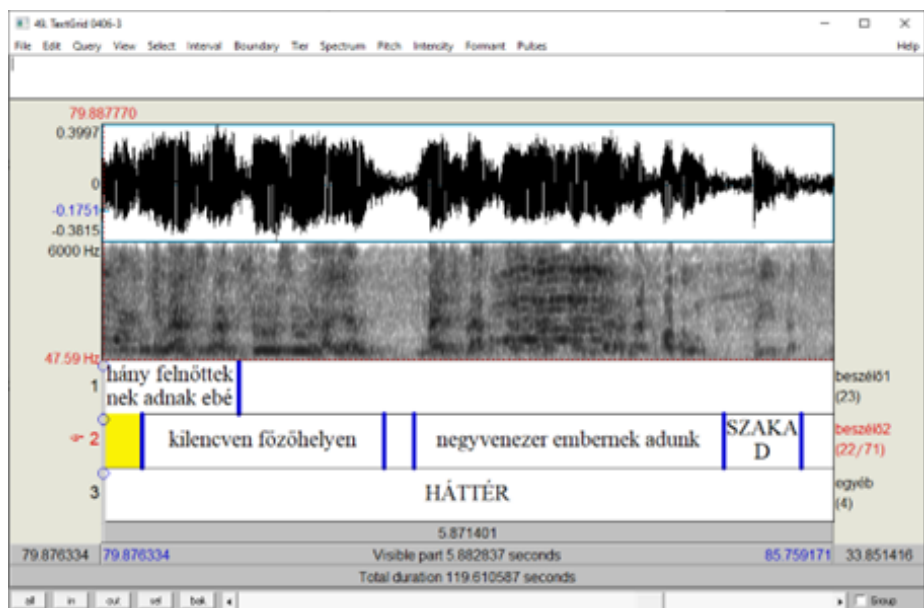


2. ábra. A 404. sz. híradó annotációjának részlete.

ter meglátogatott egy budapesti csecsemőgondozó intézményt, ahol a szociális ellátásról nyilatkozott. Ebben a hírben interjúkészítő is megszólal, és egyszerre beszélés is előfordul, emiatt két beszédsvot használtunk. Háttérzajként végig gyermekzsivaj és egyéb zajok hallatszanak. Megfigyelhető itt egy filmszakadás következtében adódó ugrás. Emiatt az előző szakaszt az utolsó egészben elhangzó szó (*adunk*) határánál lezártuk.

2.2.3. A metaadatok

A fentiek alapján látható, hogy számos új metaadat rögzítése vált szükségesé, így az adattáblát jelentősen át kellett strukturálni. Az adatokat tartalmazó táblázat sorai a rekordok, azaz az egyedi filmhírek. Az első oszlopok az egyes filmhírek azonosítását szolgálják. Rögzítettük a filmhír, a híradó, a filmhír híradón belüli sorszámát, továbbá a hangfájl nevét. Feltüntettük, hogy rendelkezésre áll-e az annotációt tartalmazó textgrid-fájl, illetve a híradó bemutatásának idejét is rögzítettük. Az 1. táblázat az adattábla részletét mutatja. A táblá-



3. ábra. A 406/3. sz. filmhír annotációjának részlete.

zatból megállapítható, hogy a 403. sz. híradó hét filmhírt tartalmazott, ebből kettőben volt magyar nyelvű beszéd.

1. táblázat. A filmhírek azonosítóit tartalmazó adattábla részlete

sorszám	híradó	hír	hangfájl	annotálva	év	hónap
25	403	1	0403	1	1931	november
26	403	2	0403-2	1	1931	november
27	403	3	0403-3	1	1931	november
28	403	4	0403	1	1931	november
29	403	5	0403	1	1931	november
30	403	6	0403	1	1931	november
31	403	7	0403	1	1931	november

A táblázat következő oszlopaiban a beszédesemény jelenlétét, illetve néhány jellemzőjét tüntettük fel. Többféle megoldás kipróbálását követően ezeknek az adatoknak a bináris ábrázolását láttuk legcélravezetőbbnek, ugyanis így világosan átláthatók az egyes tulajdonságok, és ezek kombinációival egyszerűen beál-

líthatók szűrések. A „beszéd” feliratú oszlop azt jelzi, hogy van-e beszédesemény az adott filmhírben. A „csak beszéd” arra utal, hogy háttérzöreje vagy aláfestő zene nélküli-e a beszéd. A beszéd alatti aláfestő zene és a háttérzaj jelenlétét külön oszlopokban tüntettük fel. Rögzítettük, hogy volt-e tömeges beszéd (pl. éljenzés), megszólal-e több beszélő (pl. interjúkészítő és interjúalany), előfordul-e egyszerre beszélés, illetve a beszédesemény magyar vagy idegen nyelvű-e. A 2. táblázat illusztrálja mindezt az előzőekben bemutatott rekordokkal (az első néhány oszlop takarásával).

2. táblázat. A filmhírek metaadatait tartalmazó adattábla részlete

sorsz.	híradó	hír	beszéd	csak beszéd	aláfestő zene	háttérzaj	tömeg	több beszélő	egysz.	magyar	idegen
25	403	1	0	0	0	1	0	0	0	0	0
26	403	2	1	1	0	1	0	0	0	1	0
27	403	3	1	0	0	1	0	1	1	1	0
28	403	4	0	0	0	1	0	0	0	0	0
29	403	5	1	1	0	1	0	1	0	0	1
30	403	6	1	0	0	1	0	0	0	0	1
31	403	7	1	0	0	1	0	1	0	0	1

Megfigyelhető, hogy két filmhír (403/1, 403/3) nem tartalmazott beszédet, csak háttérzajt. A 403/2. és 403/5. sz. filmhíreknél a „csak beszédet” és a „háttérzajt” is megjelöltük. Ennek az az oka, hogy ezekben a hírekben a narrátor hangja tisztán és háttérzajjal is hallható. Három filmhírnél adatoltunk több beszélőt, egy esetben találtunk egyszerre beszélést. Az öt, beszédet tartalmazó filmhír közül kettőben magyar, háromban idegennyelvű szöveget találtunk. A 2. táblázatban terjedelmi okok miatt nem tüntettük fel, de az adattábla tartalmaz még oszlopokat annak megjelenítésére, hogy milyen nyelvű volt a szöveg. Az eddig feldolgozott anyagban német, angol, francia és latin nyelven hangoztak el beszédek.

Rögzítettünk még további kiegészítő információkat. Külön oszlopban megjelenítettük, hogy a filmhír tartalmazott-e a helyszínen játszott zenét (ez nem

azonos az utólag alákevert, aláfestő zenével), a közlés tartalmazott-e paralingvisztikai elemeket (pl. hűmmögést, sóhajtatást, csettintést), illetve a film technikai minőségével kapcsolatban két lényeges információt tüntettünk fel, a filmszakadást és a torzulást (3. táblázat, szintén a korábbi oszlopok kitakarásával).

3. táblázat. A filmhírek azonosítóit tartalmazó adattábla részlete

sorszám	híradó	hír	zene	paralingvisztikai	szakadás	torz
25	403	1	1	0	0	0
26	403	2	0	0	0	0
27	403	3	0	0	1	0
28	403	4	1	0	0	0
29	403	5	1	0	0	1
30	403	6	0	0	0	1
31	403	7	1	0	0	1

Mindezek mellett továbbra is feltüntettük az adattáblában az első fázisnál már említett módon a beszélő személyét, életkorát, a beszédtypust, a filmhír témáját, illetve a filmhírhoz mutató linket.

A jelenleg is folyamatban lévő második fázis során eddig 61 filmhír feldolgozása történt meg, ami az 1931 szeptembere és decembere közötti filmhíradók híranyagát jelenti.

3. Kutatási tapasztalatok

Az eddigiekben egy kutatás készült az első fázisban összegyűjtött adatok, illetve további mérések alapján (Gocsál, 2022). Ebben a kutatásban az elhangzó beszédek artikulációstempó-értékeit vizsgáltuk. Az artikulációs tempó biztonságosan mérhető, mivel a beszédszakaszok hossza és az elhangzó beszédhangok száma könnyen megállapítható. Mivel a protokoll hiánya miatt a beszédanyag strukturálatlan, elvégeztük a felvételek beszédtypusok szerinti, az adatbázisban első megközelítésben rögzítetteknek pontosabb kategorizálását is. Fő kategóriáknak a Wacha (1974) és Krepsz (2016) által közölt típusokat – spontán beszéd,

felolvasás, reprodukív vagy interpretatív beszéd, félreprodukív beszéd – vettük alapul, majd ezekhez kvalitatív módszerrel, iteratív eljárással olyan alkategóriákat dolgoztunk ki, amelyek a konkrét beszédeseményeket pontosabban jellemzik. Így meghatároztuk, hogy a filmhírekben a legnagyobb számban szónoklatok, szűkebb körben elmondott beszédek, társalgások, narrációk, bejelentések, vezényszavak hangzottak el. Ezekre a beszéd típusokra külön-külön kiszámítottuk az artikulációstempó-értékeket, és a vezényszavak kivételével minden esetben 10–12 hang/s közötti mediánértéket kaptunk. A Kruskal–Wallis-próba nem igazolt szignifikáns különbséget a szűkebb körű beszédek, társalgások, narrációk és bejelentések tempóértékei között ($H(3) = 3,968$, $p = 0,265$), a szónoklatok beemelésével azonban szignifikáns eltérés mutatkozott ($H(4) = 13,192$, $p = 0,01$) A vezényszavak esetében valamivel 8 hang/s fölötti medián adódott.

48 beszélő személyazonosságát, és külön adatgyűjtéssel az életkorát is meg tudtuk állapítani, így a mai szakirodalmi adatokhoz hasonlóan kimutatható volt 90 évvel ezelőtti beszélők esetében is az idősebbek lassabb beszéde, azaz az életkor és az artikulációs tempó szignifikáns összefüggése ($r = -0,404$, $p < 0,01$).

Kvalitatív eljárással – kategóriák felállításával, ezek ciklikus felülvizsgálásával – a filmhíreket téma szerint is csoportosítottuk, így hét hírkategóriát (politikai-közéleti, gazdasági, sport, kiállítás-tudomány-kultúra, technikai újdonságok, hétköznapi életképek, külföldi hírek) különböztettünk meg. Ezek a kvalitatív kutatási eredmények – az adatbázisba felvett elemek bármely szempont szerinti kategorizálása – visszaforgathatók, azaz metaadatként beépíthetők az adatbázisba, így további kutatásokhoz szűrési feltételként alkalmazhatók.

Az adatbázis előfordulási gyakoriságok összesítésére is alkalmas. Meg tudtuk vizsgálni, hogy az áttekintett 15 hónapos időszakban hogyan alakult a szóbeli közléseket tartalmazó filmhírek száma, illetve a leggyakrabban előforduló beszéd típusok, a narrációk és a szónoklatok esetében bizonyos ingadozás mellett lassan emelkedő tendenciát figyeltünk meg. Ez arra utal, hogy a híradók szerkesztői egyre tudatosabban használták beszédet a hírekben.

Más akusztikai-fonetikai mérést, adatgyűjtést nem végeztünk, azonban feltűnt, hogy a viszonylag gyengébb minőség ellenére is sok esetben jól megjeleníthető a dallamgörbe, mérhetők az f_0 -értékek, továbbá megállapíthatók a formánsértékek is. Ez további kutatási lehetőségeket nyújt a korabeli beszéd tanulmányozásához. Korlátozó tényező ugyanakkor, hogy a spektrogram csak kb. 6 kHz-ig jelenik meg, így egyes kutatási feladatok elvégzésére – pl. réshangoknál a spektrális eloszlás súlypontjának (Centre of Gravity, CoG) vizsgálata, vö. Grácz et al. (2020) mai beszélőkre – erősen korlátozottan vagy sehogy nincs lehetőség.

4. További lehetőségek, fejlesztési irányok

Tapasztalataink igazolják Wallis (2014) megállapítását, azaz az annotálás során a fejlesztő nem csak információkat – pl. mondathatárokat – tesz hozzá a szöveghez, hanem ezzel párhuzamosan kritikailag reflektál az annotálás kereketeire, módszereire is. A reflexió során nem csak azt vizsgálja, hogy hasznos-e bizonyos adatokat lejegyezni, vagy helyes volt-e a lejegyzés, hanem azt is, hogy milyen kihívásokat jelent az előre nem várt, újabb tények lejegyzése. Ebből adódóan az adatbázis fejlesztése újabb és újabb felülvizsgálatokból álló ciklikus folyamat, különösen egy ilyen, eleve strukturálatlan, protokoll nélkül, más céllal rögzített beszédanyag esetében. Mindezt a fejlesztés itt ismertetett két fázisa is jól illusztrálja.

Mivel előre megjósolhatatlan, hogy milyen konkrét kutatási feladatokhoz fogják használni az adatbázist, gondolni kell arra, hogy a kutatók saját céljainak megfelelően hogyan fejleszthetik azt tovább, ahogy az észt filmhír-adatbázis esetében Ibrus & Ojamaa (2018) közölték. Előfordulhat, hogy a fonáció típusa lesz az érdekes, ebben az esetben külön annotációs sáv hozható létre (Seppänen et al., 2003), míg más esetekben, például a beszélő dialektusának megjelenítéséhez külön metaadat rögzítése lehet célszerű (Solberg & Ortiz, 2022). Az eddig használt Excel-tábla egyelőre megfelel ez utóbbi célnak, hosszabb távon, nagyobb mennyiségű adat kezeléséhez egy rugalmasan alakítható relációs

adatbázis kiépítése tűnik célszerűbbnek. Jelenlegi terveink szerint az adatokat átemeljük egy Access-adatbázisban definiált táblákba. Ezzel a megoldással a lekérdezések átláthatóbbá válhatnak, miközben a felhasználó saját kategóriákkal, vagy akár további adattáblákkal kiegészítve, tetszőlegesen fejlesztheti tovább az adatbázist. Hátránya ugyanakkor, hogy az Access használata kevésbé elterjedt, mint az Excelé, így tervezzük egy olyan felhasználói dokumentáció elkészítését is, amely az alapoktól kezdődően bemutatja az adatbázis működését és fejlesztési lehetőségeit. Sikeres tesztelést követően a Nyelvtudományi Kutatóközpont honlapján tervezzük az adatbázis közzétételét a szakmai érdeklődők számára, a többi nyelvi adatbázishoz hasonló módon.

Az akusztikai-fonetikai kutatások mellett az adatbázis más nyelvészeti szakterületek kutatását is segítheti. Csak néhány példát említve: a korabeli szóhasználat, a sajátos megszólítás- és említésformák, formális és informális élethelyzetek elemzése szociolingvisztikai, dialektológiai, sőt, nyelvtörténeti szempontból is értékes információkat nyújthat.

Számos külföldi kutató alkalmazott hasonló korpuszokat beszédtechnológiai fejlesztésekhez is. Ezek célja olyan szövegfelismerő rendszerek előállítása, amelyek a legváltozatosabb helyzetekben is robosztus megoldást tudnak nyújtani, például filmek feliratozásához (Álvarez et al., 2022). Koržinek et al. (2019) 1945 és 1962 közötti filmhíradók szövegének automatizált lejegyzéséhez állítottak össze adatbázist. Ebben a gyenge hangminőség mellett kihívást jelentett a számos, ma már nem használt kifejezés is. A Carrive et al. (2021) által ismertetett adatbázis fejlesztői pedig francia filmhíradók kutatásához többek között automatizált beszéd- és arcfelismerő szoftvereket alkalmaztak és gazdag metaadat-állományt állítottak elő, amivel az adatbázisuk többféle tudományterületen változatos kutatási feladatok megoldását segítheti.

Nem találtunk erre vonatkozóan említést a szakirodalomban, de lehetőséget látunk kriminalisztikai fonetikai eszközök alkalmazására is, mivel a torzult vagy a lehallgatás során érthetetlen beszédrészekből történeti szempontból adott esetben értékes információk nyerhetők ki. Egy példát említve: a 431. számú, egyetlen hírből álló híradóban 1:57-től azt láthatjuk, hogy az akkoriban min-

denki által jól ismert pilóta, Endresz György tragikusan végződött repülőútja előtt elbúcsúzik a budapesti közönségtől. A hír 2:31–2:35 perc közötti részében érthetetlen beszédfoszlányt hallunk. Ha valamely zaj- vagy torzításcsökkentő eljárás alkalmazásával (Fejes, 2021) egy hangszakértő meg tudja állapítani, hogy mi hangzott el – például a beszélő utalást tett saját egészségi állapotára, fáradtságára, vagy egy elhanyagolhatónak tűnő technikai problémára –, mindez új megvilágításba helyezheti a szerencsétlenség körülményeivel kapcsolatos történettudományi ismereteket. De a kriminalisztikai fonetikai eljárások és a beszédtechnikai fejlesztések tesztelésére is lehetőséget nyújt az adatbázis, mivel a bemondások igen változatos háttérhangok, zörejek, zenék mellett hangoznak el.

Szintén történettudományi jelentősége lehet az orvosi diagnosztikát támogató, beszéd alapú osztályozó eljárások alkalmazásának egyes korabeli közéleti szereplők beszédén (pl. Sztahó et al., 2019). Elképzelhető, hogy ezek segítségével az adott beszélő valamely, mindeddig nem ismert betegsége valószínűsíthetővé válik, ami új támpontot adhat döntéseinek, magatartásának újraértelmezéséhez.

A beszélő személyre és az aláfestő zenére vonatkozó metaadatok használata más jellegű kutatásokat is lehetővé tesz. Lehnert (2019) az 1950-es és 60-as években készült német filmhíradókban alkalmazott narrációs hangokat és aláfestő zenéket vizsgálta. Felfigyelt arra, hogy a II. világháború utáni években a korábbi szigorú, tekintélyt parancsoló, kiabáló hang lassan eltűnt, és helyette kevésbé merev, kevésbé stilizált beszédstílust alkalmaztak, amivel a néző számára hitelesebbé vált a hír. Az NDK-ban két női és három férfi bemondó jelent meg, mindannyiuk hangja jellegzetesen eltérő karakterű volt. A női bemondók nem csak a családdal, gyerekekkel, hanem ipari termeléssel kapcsolatos híreket is felolvastak, ami megfelelt a korabeli ideológiának, ti. nők is dolgoztak ipari munkahelyeken. Az NSZK-ban ugyanakkor csak rövid ideig alkalmaztak női hírolvasókat a filmhíreknél, aminek az oka feltehetően a hagyományosnak tekinthető szerepfelfogás volt.

Lehnert (2019) ugyanebben a munkájában a filmhírek zenei aláfestését is elemezte. Egy 1962-es filmhíradóban a munkások termelésben elért sikereiről

szóló hírben büszkeséget sugalló narrációs hangot és zenei aláfestést használtak, míg ugyanebben a híradóban más híreknél jazzt vagy dél-amerikai zenét alkalmaztak. A zenei aláfestés azonban nem csak a hangulatát adhatja meg a filmhírnek, de sajátos értelmezési kereteket is nyújthat az elhangzó szöveg, illetve képekben közölt információk befogadásához. A szerző talált példát arra, hogy egy, akkoriban jól ismert játékfilm zenéjét vették át a híradó szerkesztői egy filmhírhez, ami tudat alatt sajátos asszociációkra készíthette a nézőket, így rejtett üzenetet közvetíthetett számukra.

Végül még egy példát említve: egy másik kutatásban Gagnepain et al. (2020) televíziós hírekből állítottak össze korpuszt, amelyen egy beszédfelismerő algoritmus, majd szövegbányászati eljárás alkalmazását követően témamodellezést végeztek. Ezzel azt kívánták feltárni, hogy a médiában milyen kollektív sémák, narratívák jelennek meg egy adott, múltbéli témával kapcsolatban, amelyek egy társadalom közös emlékezetét formálják.

A hangos filmhíradók feldolgozása és adatbázisban való rögzítése új lehetőségeket nyit a beszéd, a nyelv, a nyelvhasználat kutatásához. Az itt összegzett eredmények azt is mutatják, hogy egy ilyen adatbázis a beszédeseményeket tágabb kontextusban kezelve szemiotikai, médiatörténeti, történeti, és más irányú kutatások számára is számtalan lehetőséget kínál. Úgy véljük, az itt bemutatott és a későbbiekben továbbfejlesztendő adatbázis is hozzájárulhat ilyen irányú hazai kutatásokhoz.

Köszönetnyilvánítás

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal NKFIH-FK-128814 számú pályázata támogatta.

Hivatkozások

Álvarez, A., Arzelus, H., Torre, I. G., & González-Docasal, A. (2022). Evaluating novel speech transcription architectures on the spanish RTVE2020 database. *Applied Sciences*, 12. URL: <https://doi.org/10.3390/app12041889>.

- Aubergé, V., Audibert, N., & Riiliard, A. (2003). Why and how to control the authentic emotional speech corpora. In *Proc. 8th European Conference on Speech Communication and Technology (Eurospeech 2003)* (pp. 185–188). URL: https://www.isca-speech.org/archive/eurospeech_2003/auberge03_eurospeech.html.
- Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer. version 6.0.52. <http://www.praat.org/>.
- Bóna, J., Imre, A., Markó, A., Váradi, V., & Gósy, M. (2014). GABI – Gyermeknyelvi beszédadatbázis és információtár. *Beszéd kutatás*, 22, 246–251.
- Campbell, N. (2000). Databases of emotional speech. In *Proceedings of ISCA workshop on speech and emotion*. URL: https://www.isca-speech.org/archive_open/archive_papers/speech_emotion/spem_034.pdf.
- Campbell, N., Devillers, L., Douglas-Cowie, E., Aubergé, V., Batliner, A., & Tao, J. (2006). Resources for the processing of affect in interactions. In *Proceedings of the International Conference on Language Resources and Evaluation 2006* (pp. XXV–XXVIII). URL: <http://www.lrec-conf.org/proceedings/lrec2006/>.
- Carrive, J., Beloued, A., Goetschel, P., Heiden, S., Laurent, A., Lisena, P., Mazuet, F., Meignier, S., Pincemin, B., Poels, G., & Troncy, R. (2021). Transdisciplinary analysis of a corpus of French newsreels: The ANT-RACT Project. *DHQ: Digital Humanities Quarterly*, 15. URL: <http://digitalhumanities.org/dhq/vol/15/1/000523/000523.html>.
- Cieri, C., Miller, D., & Walker, K. (2004). The Fisher Corpus: A resource for the next generations of speech-to-text. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC’04)* (pp. 69–71). European Language Resources Association (ELRA). URL: <http://www.lrec-conf.org/proceedings/lrec2004/pdf/767.pdf>.

- Devillers, L., & Vidrascu, L. (2007). Real-life emotions detection on human-human medical call center interactions. In C. Manfredi (Ed.), *Models and analysis of vocal emissions for biomedical applications: 5th international workshop. December 13–15, 2007* (pp. 143–146). Firenze: Firenze University Press. URL: https://www.isca-speech.org/archive_v0/maveba_2007/papers/mv07_143.pdf.
- Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech Communication*, 40, 33–60.
- Drude, S., Trilsbeek, P., Sloetjes, H., & Broeder, D. (2014). Best practices in the creation, archiving and dissemination of speech corpora at the language archive. In S. Ruhi, M. Haugh, T. Schmidt, & K. Wörner (Eds.), *Best Practices for Spoken Corpora in Linguistic Research* (pp. 183–207). Newcastle upon Tyne: Cambridge Scholars Publishing. URL: <https://www.mpi.nl/publications/item2051339/best-practices-creation-archiving-and-dissemination-speech-corpora>.
- Fejes, A. (2021). A kriminálisztikai hangtechnikai szakértői tevékenység. In B. Harmati, E. K. Szitkay, A. L. Pap, & B. Papp (Eds.), *Honestas, Humanitas, Humilitas: Ünnepi kötet Kerecsi Klára tiszteletére* (pp. 49–59). Budapest: L'Harmattan Kiadó. URL: http://real.mtak.hu/135138/1/Honestas_beliv_press.pdf.
- Gagnepain, P., Vallée, T., Heiden, S., Decorde, M., Gauvain, J. L., Laurent, A., Klein-Peschanski, C., Viader, F., Peschanski, D., & Eustache, F. (2020). Collective memory shapes the organization of individual memories in the medial prefrontal cortex. *Nature Human Behaviour*, 4, 189–200. doi:<https://doi.org/10.1038/s41562-019-0779-z>.
- Gocsál, Á. (2022). Beszédtípusok és artikulációs tempóik a korai hangos filmhíradókban. In L. Csárdás, & J. Bóna (Eds.), *Sokszínű beszédtudomány*. Buda-

- pest: Akadémiai Kiadó. URL: <https://mersz.hu/csardas-bona-sokszinu-beszedtudomay>.
- Gósy, M., Gyarmathy, D., Horváth, V., Grácsi, T. E., Beke, A., Neuberger, T., & Nikléczy, P. (2012). BEA: Beszélt Nyelvi Adatbázis. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* (pp. 9–24). Budapest: Akadémiai Kiadó.
- Grácsi, T. E., Csapó, T. G., Deme, A., Bartók, M., & Markó, A. (2020). Az /s/ és /ʃ/ hangok akusztikai és artikulációs elkülönítése gyermekeknél: Két longitudinális esettanulmány. In J. Bóna, & V. Krepsz (Eds.), *Nyelvfejlődés csecsemőkortól kamaszkorig* (pp. 45–66). Budapest: ELTE Eötvös Kiadó.
- Grimm, M., Kroschel, K., & Narayanan, S. (2008). The Vera am Mittag German audio-visual emotional speech database. In *2008 IEEE International Conference on Multimedia and Expo* (pp. 865–868). Hannover, Germany: IEEE. doi:<https://doi.org/10.1109/ICME.2008.4607572>.
- Gyarmathy, D., & Neuberger, T. (2015). Egy hiánypótló adatbázis: A Tini BEA. *Beszédkutatás*, 23, 209–222.
- Gyarmathy, D., Neuberger, T., & Grácsi, T. E. (2014). Lejegyzési útmutató a BEA spontánbeszéd-adatbázis háromszintű annotálásához. *Alkalmazott Nyelvtudomány*, 14, 35–41.
- Hunyadi, L., Váradi, T., Szekrényes, I., Kovács, G., Kiss, H., & Takács, K. (2018). Human-human, human-machine communication: On the HuComTech multimodal corpus. In *Selected papers from the CLARIN Annual Conference 2018*. URL: <https://ep.liu.se/ecp/159/006/ecp18159006.pdf>.
- Ibrus, I., & Ojamaa, M. (2018). Newsreels versus newspapers versus metadata: A comparative study of metadata modelling the 1930s in estonia. *VIEW Journal of European Television History and Culture*, 7, 123–137. doi:<https://doi.org/10.25969/mediarep/14758>.

- Koržinek, D., Wołk, K., Brocki, L., & Marasek, K. (2019). Automatic transcription of the polish newsreel. *Poznan Studies in Contemporary Linguistics*, 55, 183–209. doi:<https://doi.org/10.1515/psic1-2019-0008>.
- Krepsz, V. (2016). Fonetikai hasonlóságok és különbségek a beszédtypusokban. In J. Bóna (Ed.), *Fonetikai olvasókönyv* (pp. 175–188). Budapest: ELTE Fonetikai Tanszék.
- Lehnert, S. (2019). Music and voice in German newsreels of the 1950s/1960s. In *Proceedings of RE:SOUND 2019 – 8th International Conference on Media Art, Science, and Technology* (pp. 91–98). doi:<http://dx.doi.org/10.14236/ewic/RESOUND19.15>.
- Lleida, E., Ortega, A., Miguel, A., Bazán-Gil, V., Pérez, C., Gómez, M., & de Prada, A. (2020). *RTVE2020 Database Description*. URL: <http://catedrartve.unizar.es/reto2020/RTVE2020DB.pdf>.
- 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 P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech and Dialogue* (pp. 424–431). Cham: Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-10816-2_51.
- Pincemin, B., Heiden, S., & Decorde, M. (2020). Textometry on audiovisual corpora experiments with txm software. In *15th International Conference on Statistical Analysis of Textual Data JADT 2020*. Toulouse, France. URL: http://lexicometrica.univ-paris3.fr/jadt/JADT2020/jadt2020_pdf/PINCEMIN_HEIDEN_DECORDE_JADT2020.pdf.
- Pittermann, J., Pittermann, A., & Minker, W. (2010). *Handling Emotions in Human-Computer Dialogues*. Springer.
- Seifart, F. (2021). Combining documentary linguistics and corpus phonetics to advance corpus-based typology. In G. Haig, S. Schnell, & F. Seifart (Eds.), *Doing corpus-based typology with spoken language data: State of the*

- art* (pp. 115–139). Honolulu: University of Hawai’i Press. URL: <https://scholarspace.manoa.hawaii.edu/handle/10125/74659>.
- Seppänen, T., Toivanen, J., & Väyrynen, E. (2003). Mediateam speech corpus: A first large finnish emotional speech database. In *ICPhS-15* (pp. 2469–2472). URL: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_2469.pdf.
- Simon, E., & Sass, B. (2012). Nyelvtechnológia és kulturális örökség, avagy korpuszpépítés ómagyar kódexekből. *Általános Nyelvészeti Tanulmányok*, 24, 243–264.
- Solberg, P. E., & Ortiz, P. (2022). The norwegian parliamentary speech corpus. *ArXiv*, . doi:<https://doi.org/10.48550/ARXIV.2201.10881>.
- Sztahó, D., Imre, V., & Vicsi, K. (2011). Automatic classification of emotions in spontaneous speech. In A. Esposito, A. Vinciarelli, K. Vicsi, C. Pelachaud, & A. Nijolt (Eds.), *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues. Lecture Notes in Computer Science*, vol. 6800 (pp. 222–239). Berli, Heidelberg: Springer-Verlag.
- Sztahó, D., Kiss, G., Tulics, M. G., & Vicsi, K. (2019). Betegségek automatikus szétválasztása időben eltolt akusztikai jellemzők korrelációs struktúrája alapján. In *XV. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 209–212). URL: <http://acta.bibl.u-szeged.hu/59086/>.
- Teleki, C., Velkei, S., Tóth, S. L., & Vicsi, K. (2005). Development and evaluation of a hungarian broadcast news database. In F. Augusztinovicz, A. B. Nagy, & Z. Hunyadi (Eds.), *Proceedings of Forum Acusticum 4th European Congress on Acoustics* (pp. 2659–2662). Budapest: OPAKFI.
- Vicsi, K., & Sztahó, D. (2009). Ügyfél érzelmi állapotának detektálása telefonos ügyfélszolgálati dialógusban. In *VI. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 217–225). URL: <http://acta.bibl.u-szeged.hu/58711/>.

Wacha, I. (1974). Az elhangzó beszéd főbb akusztikus stíluskategóriáiról. *Általános Nyelvészeti Tanulmányok*, X, 203–216.

Wallis, S. A. (2014). What might a corpus of parsed spoken data tell us about language? In L. Veselovská, & M. Janebová (Eds.), *Complex Visibles Out There. Proceedings of the Olomouc Linguistics Colloquium 2014: Language Use and Linguistic Structure* (pp. 641–662). Olomuc: Palacký University. URL: <https://corplingstats.wordpress.com/2014/06/24/corpus-language/>.