Nyelvtudományi Intézet

BESZÉDTUDOMÁNY

SPEECH SCIENCE

2020

Budapest

Hungarian Research Institute for Linguistics

BESZÉDTUDOMÁNY – SPEECH SCIENCE

2020 (1)

Nyelvtudományi Intézet

Hungarian Research Institute for Linguistics

Budapest

Technikai szerkesztés/Typesetting: Ligeti-Nagy, Noémi

Borítóterv/Cover design: Gyarmathy, Dorottya ©

Korrektúra/Proofreading: Vakula, Tímea

A folyóiratszám kiadását az MTA Könyv- és Folyóiratkiadó Bizottsága támogatta.

This volume was supported by the Committee of Book and Journal Publications, Hungarian Academy of Sciences.

1068 Budapest, Benczúr u. 33.

Szerkesztői előszó

Kedves Kollégák!

Megújul a korábbi Beszédkutatás folyóirat, új neve: *Beszédtudomány – Speech Science*. Célunk a beszédtudomány különböző területeiről érkező kutatások ismertetése. A jelen kötet tanulmányai foglalkoznak többek között a beszéd szegmentális és szupraszegmentális szerkezetének vizsgálatával, a diskurzusjelölők leírásával, a beszédhangok különböző realizációival, illetve az idegennyelvi jellemzőkkel.

A jövőben is várjuk a tanulmányokat, pl. az artikuláció, akusztikum és percepció; a beszédtechnológia, a beszédfelismerés, a beszédszintézis, a kriminalisztikai kutatások és alkalmazások, a fonológiai folyamatok érvényesülése a beszédben; az anyanyelv és idegen nyelvek-elsajátítása, a két- és többnyelvűség, a prozódia, a szintaxis, a pragmatikai vonatkozások, a klinikai kutatások, a beszéd- és nyelvi zavarok, a korpuszok, adatbázisok fejlesztése, a diszharmóniás jelenségek a beszédben témaköreiből, valamint további, a beszéd jellemzőivel, feldolgozásával, létrehozásával kapcsolatos területekről. 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 találhatók: `https://fonetika.nytud.hu/journal`. Amennyiben emailben is szeretne értesítést kapni a felhívásról, kérjük, jelezze a fonetika@nytud.hu címen!

Üdvözlettel: a kötet szerkesztői
Gráczi Tekla Etelka, Gyarmathy Dorottya, Horváth Viktória,
Krepsz Valéria, Mády Katalin

Editorial foreword

Dear Colleagues,

The Hungarian journal Beszédkutatás (Speech Research) has been renewed. Its new name is *Beszédtudomány – Speech Science*. One of our goals is to increase the number of international publications, as is signalled by the change of the journal's title that has become bilingual. The present issue deals with the study of the segmental and suprasegmental structure of speech, the description of discourse markers, the different realisations of speech sounds and foreign language characteristics among others. Papers in all areas of speech science are welcome, among others in the field of: 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 corpus, databases; disharmonic phenomena in speech. 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 the paper length. Papers will be published based on a double blind peer-reviewing process. The submission and reviewing process will take 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 Gráczi, Dorottya Gyarmathy, Viktória Horváth,
Valéria Krepsz, Katalin Mády

Tartalomjegyzék/Table of contents

Nadia Hajjej – Tamás Gábor Csapó: Realistic Ultrasound Tongue Image Synthesis using Generative Adversarial Networks 7

Tekla Etelka Gráczi – Tamás Gábor Csapó – Márton Bartók – Andrea Deme – Alexandra Markó: The realization of voicing opposition in alveolar fricatives in Hungarian. Preliminary study on articulation and acoustics 22

Rácz Bianka – Csapó Tamás Gábor: Ajakvideó alapú beszédszintézis konvolúciós és rekurrens mély neurális hálózatokkal 57

Száraz Bettina – Gráczi Tekla Etelka: Explozívák realizációja fiatal és öregedő felnőttek beszédében 73

Kata Baditzné Pálvölgyi: The intonation of lengthenings in northern and southern dialects of Spanish 99

Surányi Balázs – Pintér Lilla: A prozódiailag jelölt fókusz azonosításának elsajátítása 124

Gyarmathy Dorottya: A néma szünetek sajátosságai az életkor és a beszédtípus függvényében 152

Horváth Viktória: Az egyszerre beszélések jellemzői háromfős társalgásokban 187

Judit Bóna – Tímea Vakula: Disfluent whole-word repetitions across the lifespan: Durational patterns and functions 214

Schirm Anita: A diskurzusjelölők a telefonos ügyfélszolgálati beszélgetésekben 237

Andrea Götz: Discourse markers and connectives in interpreted Hungarian discourse: A corpus-based investigation of discourse properties and their interdependence 259

Realistic Ultrasound Tongue Image Synthesis using Generative Adversarial Networks

Nadia Hajjej[1], Tamás Gábor Csapó[1,2]

[1] *Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics*
[2] *MTA-ELTE „Lendület" Lingual Articulation Research Group*

Abstract

Ultrasound Tongue Imaging (UTI) is a technique suitable for the acquisition of articulatory data, showing the motion of the tongue. When the subject is speaking, the ultrasound transducer is placed below the chin, resulting in mid-sagittal images of the tongue movement. The typical result of 2D ultrasound recordings is a series of grayscale images in which the tongue surface contour has a greater brightness than the surrounding tissue and air. UTI has been used for many years in phonetic research on speech production. However, these studies are mostly based on manually annotated articulatory data, and reliable extraction of high-level features from ultrasound data remains a challenge. In this paper, we propose a method to generate realistic ultrasound images from a database of midsagittal images of the tongue. First, we explain the principle of Generative Adversarial Networks (GAN), which is a subset of generative models, where deep neural networks are applied. Then, we detail our method, starting with the properties of the dataset, to the conception of the convolutional neural network model. The model consists of a generator and a discriminator network, which are trained against each other in the task of realistic image generation: the generator tries to fool the discriminator. The experiments demonstrate the efficiency of the GAN in creating realistic images when the training is run long enough, in order that the generator network can learn the properties of ultrasound images. The GAN-generated images were tested with a subjective test, and it supported our hypothesis that the synthesized ultrasound tongue images are of high quality and are difficult to distinguish from real images of the tongue. The results can be exploited for data augmentation, for predicting the next frame in a UTI sequence or for motion detection of tongue contours within images.

1. Introduction

Ultrasound tongue imaging (UTI) is a technique suitable for the acquisition of articulatory data. Stone (2005) summarized the typical methodology of investigating speech production using ultrasound. Usually, when the subject is

Email addresses: hajjej.nadia@gmail.com (Nadia Hajjej), csapot@tmit.bme.hu (Tamás Gábor Csapó)

Beszédtudomány – Speech Science 2020. 7–21. DOI: 10.15775/Besztud.2020.7-21

speaking, the ultrasound transducer is placed below the chin, resulting in mid-sagittal images of the tongue movement. The typical result of 2D ultrasound recordings is a series of gray-scale images in which the tongue surface contour has a greater brightness than the surrounding tissue and air (for a sample, see Figure 1). Although a large number of linguistic studies are applying 2D ultrasound (Stone, 2005), there are not many freely available databases with a large number of images. Eshky et al. (2018) introduced a database that is related to the Ultrax2020 project (`http://www.ultrax-speech.org/ultrasuite`), but it contains ultrasound images of children only. This UltraSuite repository currently contains tongue ultrasound data from 5–12 year old children who are typically developing or have a speech sound disorder. Another UTI dataset is related to Silent Speech Interfaces (Ji et al., 2018), but it contains processed ultrasound images and not the original raw data.

Ultrasound imaging of the tongue has been used for many years in research on speech production (Stone, 2005). For some relevant experiments of the MTA-ELTE „Lendület" Lingual Articulation Research Group, see Markó et al. (2017, 2018, 2019a) and Markó et al. (2019b). However, these studies are based on manually annotated articulatory data, and reliable extraction of high-level features from ultrasound data (e.g. automatic tongue contour tracking) remains a challenge (Csapó & Lulich, 2015; Csapó & Csopor, 2015; Xu et al., 2017b). The topic of the current study, i.e. the realistic synthesis of ultrasound images can be a starting point for exploiting the higher-level representation of the tongue in a variety of applications in speech research.

Ever since computers were developed, scientists and engineers thought of artificially intelligent systems that work and react like humans. In the past decades, the increase of generally available computational power provided a helping hand for developing fast learning machines. Meanwhile, the internet supplied an enormous amount of data for training. These two developments boosted the research on smart self-learning systems, with neural networks among the most promising techniques.

Recently, deep neural networks have produced high accuracy scores in speech and ultrasound-related tasks, such as articulatory-to-acoustic mapping (Csapó et al., 2017b), articulation-to-text mapping (Xu et al., 2017a; Tóth et al., 2018), articulation-to-F0 prediction (Grósz et al., 2018; Csapó et al., 2019), acoustic-to-articulatory inversion (Porras et al., 2019) and also edge (contour) detection (Csapó & Csopor, 2015; Xu et al., 2017b). For these problems, usually, regression models are used, which can be trained for mapping from the input to the target feature. Typical networks are fully-connected feed-forward deep neural networks, convolutional neural networks, and recurrent neural networks.

1.1. Generative models

A branch of deep learning methods deals with generative models, i.e. how to generate new data that is similar to the properties of the training data. In general, the goal of the generative models is to estimate or to learn the data distribution of the training data and generate new data points with some variations by modeling a distribution, which is as much as possible close to the real data distribution. Most of the generative models use the maximum likelihood method to define a model that estimates the parametrized probability distribu-

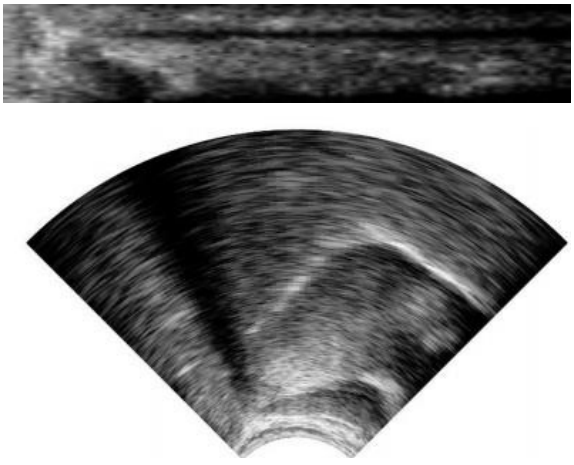


Figure 1: Ultrasound image of the tongue. Top: raw scanline data. Bottom: 'wedge' format (the tongue root is on the left, while the tongue tip on the right).

tion (Goodfellow, 2017). Those models differ mainly in the approximation of maximum likelihood. There are two main types: 1) models that aim to represent the probability distribution over the space where the data lies explicitly, and 2) models that interact implicitly with the probability distribution and try to generate samples from it. Since the introduction of Generative Adversarial Networks (GANs) by Goodfellow et al. (2014), they have proven a vast potential to automatically learn the natural features of a particular dataset, to mimic any data distribution and generate data like it. Typical examples include generated digits, flowers, realistic human faces (Karras et al., 2018), or speech data for emotion recognition (Chatziagapi et al., 2019).

In this paper, we aim to synthesize realistic ultrasound tongue images using Generative Adversarial Networks, that belong to the second type of the above generative models. The GAN-generated ultrasound images can be useful for data augmentation, which might be necessary for scenarios with limited data, e.g. for motion detection of tongue contours within images (Xu et al., 2017b) or articulatory-to-acoustic mapping (Csapó et al., 2017b).

2. Methods

2.1. GAN framework

The purpose of GANs is to create samples, which are able to deceive humans and even computers. Thus, the main idea of GAN is to set up a game between two players: the generator and the discriminator (Goodfellow et al., 2014; Goodfellow, 2017). The generator is the player that creates samples. Those samples are intended to come from the same distribution as the training data. The discriminator is the second player that examines samples to decide whether they are real or fake. The discriminator usually learns using traditional supervised learning techniques to divide inputs into two classes (real or fake). Figure 2 shows the block diagram of a GAN with sample real and generated ultrasound tongue images. Each of our players has its own differentiable function with respect to its parameters and its inputs. The discriminator has a function D

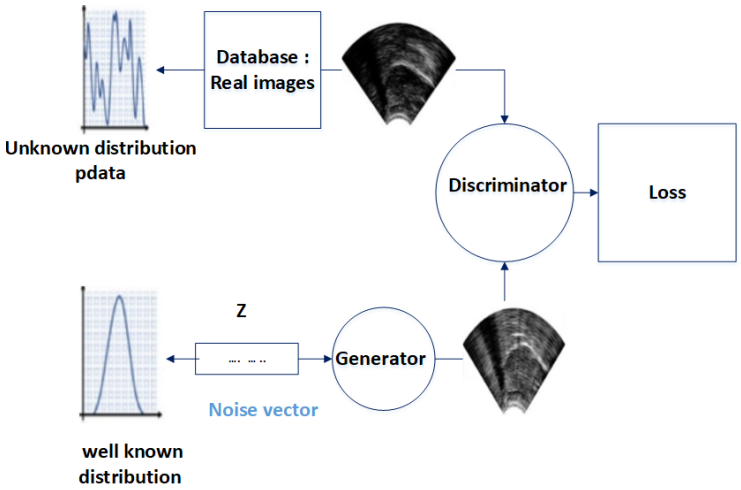


Figure 2: Block diagram of a GAN with discriminator and generator networks, used for the ultrasound tongue image synthesis task.

that takes x as input and uses θ_D as parameters. The generator is defined by a function G that takes z as input and uses θ_G as parameters. Both players have cost functions that are defined in terms of both players' parameters. This cost function is defined by the following equation:

$$min_{\theta_G} max_{\theta_D} (E_x log(D_{\theta_D}(x) + E_z log(1 - D_{\theta_D}(G_{\theta_G}(z)))$$

Now let us explain this relation. We have a real image x that will be examined by the discriminator D. For this image, it will give a value close to zero. Hence, for a fake image, it will give a higher value close to one. For the generator G, he will take a randomly generated vector from a very simple and well-known distribution and produce an image that will also be used to train the discriminator. The latter will be alternatively shown real and fake images. The generator's role is to minimize the output of D by providing more realistic images, while D tries to maximize the same thing. Each player's cost depends on the other player's parameter, but they can only control their own parameter. This scenario is most straightforward to describe as a 'minimax' game where the solution is a Nash equilibrium (Goodfellow et al., 2014; Goodfellow, 2017).

2.2. Dataset

Before building the required neural network, we chose the dataset from which we are aiming to generate similar samples. This dataset contains tongue-ultrasound images. For training the GAN, we used ultrasound tongue images from a previously collected database (Csapó et al., 2017a), which applied the "Micro" ultrasound system (Articulate Instruments Ltd., UK). The database contains raw midsagittal ultrasound images of the scanline data (see Fig. 1 left) recorded at 82 fps from several speakers, of which we chose one Hungarian female speaker for our experiments and used 209 sentences altogether. Each pixel is stored as a 1-byte unsigned integer, which is actually a grayscale pixel intensity. Using the extracted raw ultrasound images, we can convert and visualize them as ultrasound frames in the 'wedge' format (see Fig. 1 right). Those frames can be used to produce a video illustrating the movement of the tongue.

In our case, we are interested in using the raw ultrasound images, as they contain the information before any image processing. Therefore, after successfully extracting those images, we built a data set of 27 925 raw images having the dimension of 64×842 pixels. We split the image set into two groups. The first group is made of 2/3 of images, which is used for training, and the second one made of 1/3 of the dataset used for testing. Figure 1 illustrates a sample of raw ultrasound images and ultrasound frames as well.

2.3. Proposed method for ultrasound tongue image generation

As we have mentioned, the principle of Generative Adversarial Networks is managing a game between the two networks (i.e. the generator and the discriminator). We implemented this in Python using a DC-GAN implementation as a starting point (`https://github.com/carpedm20/DCGAN-tensorflow/`). For our project, we chose to use a deep convolutional neural network for both the generator and the discriminator. In our model, we fixed the number of hidden layers to nine for both discriminator and generator. As hyperparameter, we fixed 64 as batch size and 25 for the number of epochs. After every 100 iterations, we generated 64 images.

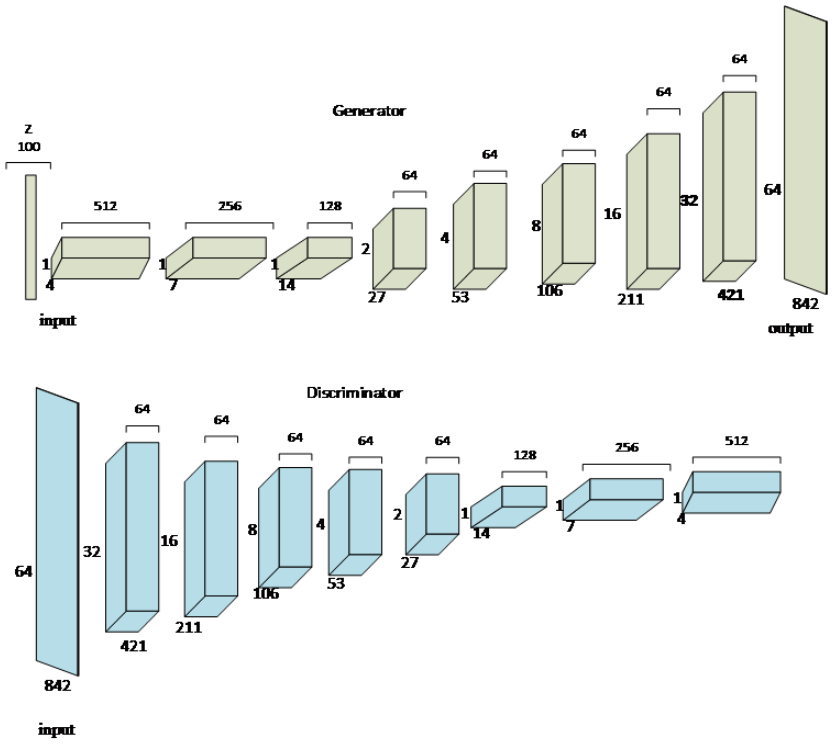


Figure 3: The architecture of the Generator (top) and Discriminator (bottom) networks within the GAN.

- for number of iterations: Maximise G
- for k steps:
  - Sample minibatch of noise samples and generate images
  - Sample minibatch from the image dataset
  - Update the discriminator using gradient ascent
- Sample minibatch of noise samples and generate images
- Update the generator Minimise D

Figure 4: Pseudo code of GAN training.

Figure 3 illustrates the GAN network architecture. The discriminator (Fig. 3 bottom) is a downsampling network using strided convolutional layers. Its role is to check real images and save variables in order to use them in fake image checking. The discriminator has a last convolutional layer before applying the cross-entropy. On the other hand, the generator (Fig. 3 top) is an upsampling neural network that takes as input a vector z randomly generated from a known distribution. After linearly transforming z, it will be fed through the layers in order to get as a result an image with the same size as our original image. The generated image will be the input of the discriminator, and according to its result, the networks will improve their performance. There is one discriminator update per each generator update. The process can be summarized with the following pseudo-code shown in Figure 4.

Obviously, the quality of the GAN-generated images was low in early epochs and continuously improved during the training until the final epochs, because as we have written, during the training, the generator improves its performance.

3. Experiments and results

As we have previously mentioned, after every 100 iterations during the GAN training, we generated 64 ultrasound images. The raw images were converted to the 'wedge' format for visualization. In order to assess the quality of our images, we created an internet-based test (`http://leszped.tmit.bme.hu/gan2018/`). In this test, we used 100 hand-selected samples, including 20 real and 80 generated images, chosen by visual inspection to ensure that there are different images in the experiment. The latter is made of 20 'early' samples created after the early iterations of the training, and 60 'late' ones generated from the last iterations. The task of the participants was to assess the quality and reality of the images on a scale between 0–100, without knowing whether they are real or generated. Thus, as a result, we will have numbers associated to images describing their quality. Figures 5–7 show several examples from the subjective evaluation process, while Figure 8 presents a sample image with the question provided.

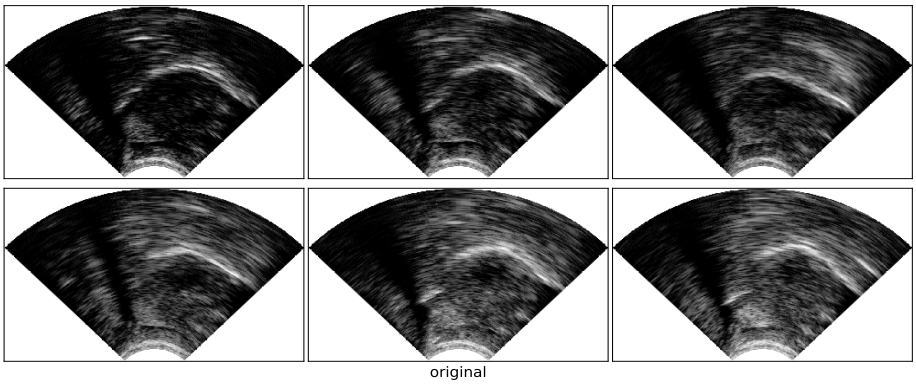


Figure 5: Original Ultrasound Tongue Images.

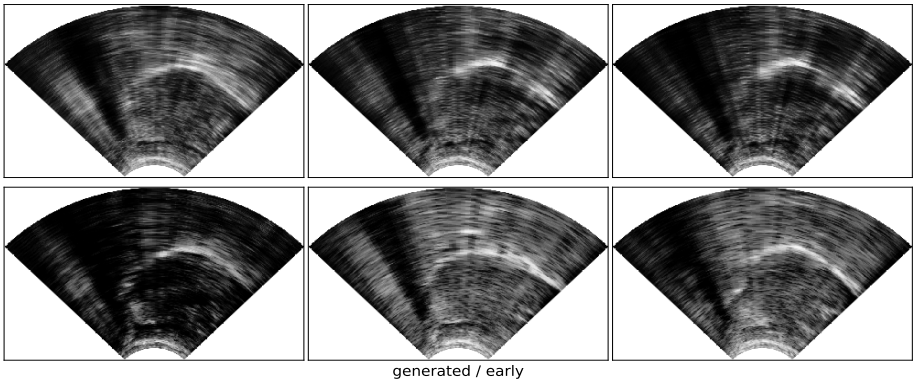


Figure 6: Generated Ultrasound Tongue Images from early epochs.

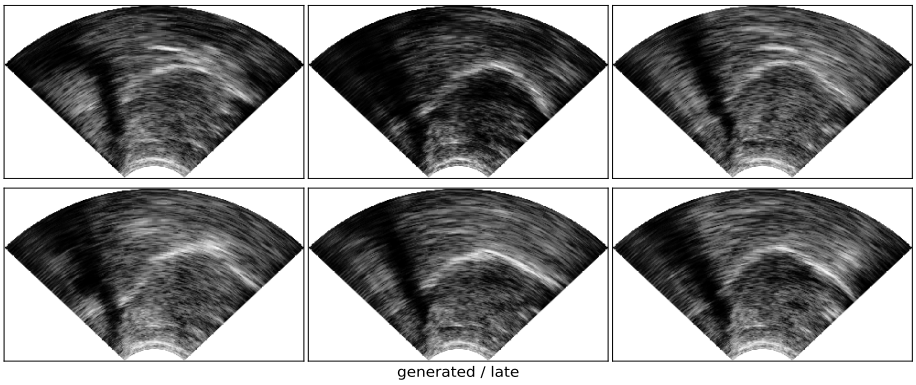


Figure 7: Generated Ultrasound Tongue Images from late epochs.

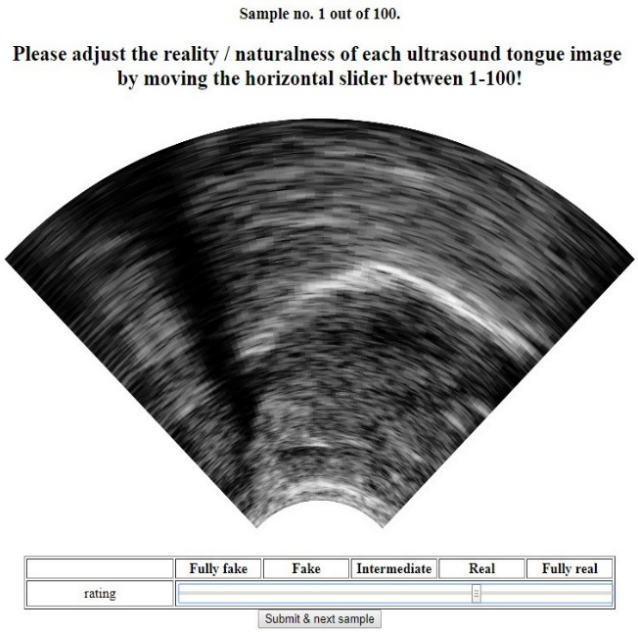


Figure 8: A sample generated image from the subjective test.

In Fig. 6, the 'early' generated images can be mostly distinguished from the 'late' images of Fig. 7, as in the early epochs, the generator within the GAN was not able to produce realistic images. On the other hand, the 'late' generated images (Fig. 7) are visually close to the original ultrasound images of the tongue (Fig. 5). In the figures, the shapes of the tongues are different (as they are isolated examples of real or synthetic images), and would produce different sounds. This means that while the model generates realistic looking ultrasound images, they are not constrained linguistically, which could be addressed in future work.

A total of 8 subjects, blinded to the approaches, participated in the subjective test, three of them being speech researchers and the remaining five being university students. The test took, on average, 13 minutes to complete. The test results are summarized in Table 1. Analyzing the results, we can see that the images generated from the 'early' epochs were evaluated with low scores (around 29%). In contrast, the tongue images from the 'late' epochs reached 59%, which

Table 1: Mean and standard deviation result of the subjective test for the 'reality / naturalness' question. Higher scores are better.

	early images	late images	real images
No. of images	20	60	20
Average result (experts)	17.57 (14.63)	63.28 (23.54)	68.77 (25.09)
Average result (non-experts)	35.65 (29.69)	56.79 (22.13)	68.06 (19.29)
Average result (all)	28.87 (26.61)	59.23 (22.89)	68.33 (21.66)

is close to the quality of the real ultrasound images (being 68%). The three experts were more strict: they evaluated the 'early' images with lower scores, and the 'late' images with higher scores than the non-experts. Therefore, we can say that the GANs are efficient in ultrasound tongue image generation and deceive humans, as the results showed how hard it is to differentiate between real and generated images.

4. Discussion

Generative Adversarial Networks, being a subfield of generative models within machine learning, are suitable to synthesize new images which are similar to the training data. Typical example uses of GANs include generated digits, flowers, or human faces (Karras et al., 2018). In this study, we presented a pioneering work in ultrasound tongue image synthesis using GANs.

According to the experiments, the Generative Adversarial Networks are able to generate realistic tongue ultrasound images. Therefore, the results can be useful for data augmentation. This might be important for scenarios with limited data, e.g. for motion detection of tongue contours within images (Xu et al., 2017b) or articulatory-to-acoustic mapping (Csapó et al., 2017b). One potential issue is that errors in synthetic data can propagate into future models trained on this data, something we need to be careful about not just in medical applications but in general. Besides, a conditional GAN with a similar architecture

could be used for predicting the next frame in a UTI sequence (Wu et al., 2018), or can be useful for acoustic-to-articulatory inversion (Porras et al., 2019).

5. Conclusion and future work

In this paper, we have shown in detail our method aiming to synthesize realistic ultrasound images. First, we have explained the principle of Generative Adversarial Networks. Then, we have detailed our method, starting with the creation of the dataset, to the conception of the network model, and finally, the investigation of the obtained results.

The performance shown by the GANs in generating realistic tongue ultrasound images encourages us to improve the used model by taking into consideration the time dimension, to be able to predict the next input for the generator (e.g. as a form of a recurrent neural network) which may enhance the performance and get better and accurate results. In the future, we plan to train generative networks conditioned on the linguistic content, and test the GAN-based methods on other types of articulatory data (e.g. vocal tract MRI and lip images).

Acknowledgement

The authors were partially funded by the National Research, Development and Innovation Office of Hungary (FK 124584 and PD 127915) and by the MTA „Lendület" program. We would like to thank the subjects taking part in the subjective experiment. The Titan X GPU used for this research was donated by NVIDIA Corporation.

References

Chatziagapi, A., Paraskevopoulos, G., Sgouropoulos, D., Pantazopoulos, G., Nikandrou, M., Giannakopoulos, T., Katsamanis, A., Potamianos, A., & Narayanan, S. (2019). Data Augmentation Using GANs for Speech

Emotion Recognition. In *Proc. Interspeech* (pp. 171–175). Graz, Austria. URL: http://www.isca-speech.org/archive/Interspeech{_}2019/abstracts/2561.html. doi:10.21437/Interspeech.2019-2561.

Csapó, T. G., Al-Radhi, M. S., Németh, G., Gosztolya, G., Grósz, T., Tóth, L., & Markó, A. (2019). Ultrasound-based Silent Speech Interface Built on a Continuous Vocoder. In *Proc. Interspeech* (pp. 894–898). Graz, Austria. doi:10.21437/Interspeech.2019-2046. arXiv:1906.09885.

Csapó, T. G., & Csopor, D. (2015). Ultrahangos nyelvkontúr követés automatikusan: a mély neuronhálókon alapuló AutoTrace eljárás vizsgálata [Automatic tongue contour tracking based on ultrasound: investigation of the Deep Neural Network based AutoTrace method] (in Hungarian). *Beszédkutatás 2015 [Speech Research 2015]*, *1*, 177–187.

Csapó, T. G., Deme, A., Gráczi, T. E., Markó, A., & Varjasi, G. (2017a). Synchronized speech, tongue ultrasound and lip movement video recordings with the "Micro" system. In *Challenges in analysis and processing of spontaneous speech*.

Csapó, T. G., Grósz, T., Gosztolya, G., Tóth, L., & Markó, A. (2017b). DNN-Based Ultrasound-to-Speech Conversion for a Silent Speech Interface. In *Proc. Interspeech* (pp. 3672–3676). Stockholm, Sweden. URL: http://dx.doi.org/10.21437/Interspeech.2017-939. doi:10.21437/Interspeech.2017-939.

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.

Eshky, A., Ribeiro, M. S., Cleland, J., Richmond, K., Roxburgh, Z., Scobbie, J. M., & Wrench, A. (2018). UltraSuite: A Repository of Ultrasound and Acoustic Data from Child Speech Therapy Sessions. In *Proc. Interspeech* (pp. 1888–1892). Hyderabad, India: ISCA. URL: http://www.isca-speech.org/archive/Interspeech{_}2018/abstracts/1736.html. doi:10.21437/Interspeech.2018-1736.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)* (pp. 2672–2680). URL: https://papers.nips.cc/paper/5423-generative-adversarial-nets.

Goodfellow, I. J. (2017). NIPS 2016 Tutorial: Generative Adversarial Networks. *CoRR*, *abs/1701.0*. URL: http://arxiv.org/abs/1701.00160. arXiv:1701.00160.

Grósz, T., Gosztolya, G., Tóth, L., Csapó, T. G., & Markó, A. (2018). F0 Estimation for DNN-Based Ultrasound Silent Speech Interfaces. In *Proc. ICASSP* (pp. 291–295). Calgary, Canada.

Ji, Y., Liu, L., Wang, H., Liu, Z., Niu, Z., & Denby, B. (2018). Updating the Silent Speech Challenge benchmark with deep learning. *Speech Communication*, *98*, 42–50. doi:10.1016/j.specom.2018.02.002. arXiv:1709.06818.

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive Growing of GANs for Improved Quality, Stability, and Variation. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net. URL: https://openreview.net/forum?id=Hk99zCeAb.

Markó, A., Bartók, M., Csapó, T. G., Deme, A., & Gráczi, T. E. (2019a). The effect of focal accent on vowels in Hungarian: Articulatory and acoustic data. In *Proc. ICPhS* (pp. 2715–2719). Melbourne, Australia.

Markó, A., Bartók, M., Gráczi, T. E., Deme, A., & Csapó, T. G. (2018). Mondathangsúlyos és hangsúlytalan helyzetű magánhangzók néhány artikulációs és akusztikai jellemzője a magyarban. *Beszédkutatás*, *26*, 85–109.

Markó, A., Csapó, T. G., Deme, A., Gráczi, T. E., & Bartók, M. (2019b). Gyermekek lingvális artikulációjának variabilitása magánhangzós nyelvkontúrok

alapján. In *Az anyanyelv-elsajátítás folyamata hároméves kor után* (pp. 165–190). ELTE Eötvös Kiadó.

Markó, A., Csapó, T. G., Deme, A., Gráczi, T. E., & Varjasi, G. (2017). A gyermeki artikuláció vizsgálata – Új lehetőségek a hazai kutatásban. In *Új utak a gyermeknyelvi kutatásokban* (pp. 65–95). Budapest, Hungary: ELTE Eötvös Kiadó. URL: http://www.eltereader.hu/media/2017/11/Bona{_}Gyermeknyelv{_}READER.pdf.

Porras, D., Sepúlveda-Sepúlveda, A., & Csapó, T. G. (2019). DNN-based Acoustic-to-Articulatory Inversion using Ultrasound Tongue Imaging. In *International Joint Conference on Neural Networks* (pp. N–19221). Budapest, Hungary. URL: http://arxiv.org/abs/1904.06083. arXiv:1904.06083.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, *19*, 455–501. doi:10.1080/02699200500113558.

Tóth, L., Gosztolya, G., Grósz, T., Markó, A., & Csapó, T. G. (2018). Multi-Task Learning of Phonetic Labels and Speech Synthesis Parameters for Ultrasound-Based Silent Speech Interfaces. In *Proc. Interspeech* (pp. 3172–3176). Hyderabad, India. doi:10.21437/Interspeech.2018-1078.

Wu, C., Chen, S., Sheng, G., Roussel, P., & Denby, B. (2018). Predicting tongue motion in unlabeled ultrasound video using 3D convolutional neural networks. In *Proc. ICASSP*. Calgary, Canada.

Xu, K., Roussel, P., Csapó, T. G., & Denby, B. (2017a). Convolutional neural network-based automatic classification of midsagittal tongue gestural targets using B-mode ultrasound images. *The Journal of the Acoustical Society of America*, *141*, EL531–EL537. doi:10.1121/1.4984122.

Xu, K., Roussel, P., & Denby, B. (2017b). Is Speckle Tracking Feasible for Ultrasound Tongue Images? *Acta Acustica united with Acustica*, *103*, 365–368. doi:10.3813/AAA.919065.

The realization of voicing opposition in alveolar fricatives in Hungarian. Preliminary study on articulation and acoustics

Tekla Etelka Gráczi[1,4], Tamás Gábor Csapó[2,4], Márton Bartók[3,4], Andrea Deme[3,4], Alexandra Markó[3,4]

[1] *Hungarian Research Institute for Linguistics*
[2] *Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics*
[3] *Eötvös Loránd University, Department of Applied Linguistics and Phonetics*
[4] *MTA-ELTE „Lendület" Lingual Articulation Research Group*

Abstract

The simultaneous articulation of the turbulent noise of fricatives and vocal fold vibration poses difficulties due to their conflicting pressure requirements. Previous studies found advanced tongue root and narrower obstacle in voiced fricatives than in voiceless ones. The first helps to maintain vocal fold vibration, while the latter helps to achieve the appropriate amount of turbulence.

In our study 12 subjects produced /izi/ and /isi/ sequences in pre-focal position. Headset microphone-, EGG- and tongue ultrasound (US)-signals were recorded. Cessation and restart points of voicing, and the voiceless part ratio (VR) were measured in the EGG-signal. CoG, SD, skewness and kurtosis were measured in the acoustic signal at 11 equally distanced time points in the fricatives. The midsagittal tongue contours were analyzed in the US signal in the closest image to the 0%, 50% and 100% points of the fricatives' total duration. Voicing characteristics of /z/ and /s/ were compared by LMM, the further spectral features were analyzed by GAMM, and the tongue contours were analyzed by polar GAMM.

The VR, the cessation and restart point of voicing were distinctive, although some of them had large VR in /z/ realizations. That may be resulted not only by the laryngeal settings but also by the supraglottal settings. The present study found tongue contour differences between the two fricatives at 50%, of the fricatives, and also at 0% and 100% point, but in less subjects' speech: suggesting advanced tongue root and narrower constriction in /z/ realizations and speaker dependent timing of gestures. The spectral measures did not reflect the US results in one-on-one way. That is explicable by the quantal relations of the two domains (Stevens, 1968), and we suggest that they are also a result of further articulatory maneuvers that are applied in the voiced and voiceless fricative pairs (see Liker & Gibbon, 2011, 2013, 2018).

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

Beszédtudomány – Speech Science 2020. 22–56. DOI: 10.15775/Besztud.2020.22-56

1. Introduction

Fricatives are produced with turbulent airflow through a narrow constriction in the oral cavity. The participating articulators and the size of the vocal tract in front of the constriction determine the resulting acoustic patterns (Fant, 1960; Shadle, 1991).

In order to produce a high intensity turbulent noise, the cross-sectional area of the oral obstacle must be smaller than that of the glottis, and the intraoral pressure needs to be larger than the atmospheric. The high intraoral pressure also means high supraglottal pressure, therefore the transglottal pressure differential decreases. The continuous increase of the pressure above the glottis also leads to an increase in the area of the glottis. The loss of the transglottal pressure differential and the increasing pressure in the glottis hinders vocal fold vibration. The vocal folds are first forced to vibrate slower than to stop vibration and stay apart (Bickley & Stevens, 1986; Stevens, 1997). As a result, voiced obstruents, hence also voiced fricatives may become partially (or totally) devoiced (Smith, 1997).

Although the intraoral pressure rises during the production of obstruents before full articulatory closure or constriction is reached (Müller & Brown Jr., 1980), and decreases only when the obstacle starts opening (or other articulatory strategies cause its decrease via the initiation of volume expansion of the oral cavity), vocal fold vibration does not cease in all voiced obstruents. There is a narrow range of pressure in which both voicing and friction could be maintained (Ohala & Solé, 2010). This can be reached by adjusting the cross-sectional area of the glottis and the constriction to approximately equal values (Stevens et al., 1992).

Voicing may also be maintained by other articulatory maneuvers / articulatory strategies through active or passive enlargement of the vocal tract during voicing. For instance, slower pressure build up can be achieved by expanding the area behind the obstacle (Docherty, 1992; Fuchs & Perrier, 2003). Further, lowering the larynx, enlarging the oral cavity, lowering the tongue, or forward-

ing the radix may also be used to expand the oral cavity, and thus to decrease intraoral pressure. The relaxation of the larynx with the resulting supraglottal area increase cause the slackening of the muscles close to the tongue surface that results further passive expansion of the cavities (Svirsky et al., 1997). While most of the earlier studies investigating articulatory maneuvers that may aid the maintenance of voicing concentrated on stops, more recent studies also aim to describe articulatory maneuvers in fricatives. Narayanan and colleagues (1995) studied /f θ s ʃ v ð z ʒ/ in American English in four speakers by MRI. They found advanced tongue root and larger pharyngeal area in voiced fricatives than in voiceless ones. Fuchs and her colleagues (2007) found that voicing during the frication interval was a less reliable discriminator of the voicing contrast, especially in Southern speakers of German and in word final position. Their results also showed that the relative voicing duration and the amount of tongue palate contact correlated who did not devoice, and that voiced fricatives showed more anterior articulation than voiceless ones (especially postalverolars).

The articulatory timing of voiced and voiceless fricatives were also found to be different in English and Croatian. In EPG-studies /s/ realizations were found to need longer time to achieve the largest contact surface at the constriction in both languages than the realizations of /z/ (Liker & Gibbon, 2013, 2018).

Studies on the area of contact using EPG showed contradictory results. Fletcher (1989) and Tabain (2001) did not find differences between the alveolar and the postalveolar voicing counterparts, while McLeod and her colleagues (2006) and Fuchs and her colleagues (2007) found a greater percentage of anterior tongue palate contact in the production of voiced fricatives than that of voiceless ones. Since vocal fold vibration results in lower intraoral pressure that would lead to lower turbulence intensity, narrower constriction may appear in order to avoid turbulence to be low in intensity. Additionally, the greater amount of contact may be attributed to a narrower medial groove. This hypothesis is also supported by the absence of this tendency in subjects, who devoiced voiced fricatives in their production (Fuchs et al., 2007).

Liker and Gibbon (2011) also studied the groove of the constriction in the anterior and the posterior part of the constriction in /ʃ/ and /ʒ/. Although the anterior groove was smaller in most speakers than the posterior, no consistent difference was found between the two counterparts. The anterior groove did not show major differences, while the posterior groove was larger in voiced fricatives in 3 out of 5 subjects' pronunciation. This shows that the parallel maintenance of the targets of voicing and friction exhibits differences across speakers.

The maintenance of vocal fold vibration in voiced fricatives and thus articulatory patterns were found to vary not only across speakers but also across languages (Shih et al., 1999).

The motivation of the research that the present study belongs to was to analyze the articulatory differences across the consonant duration. Our goal was to detect if the distinction that is present at the mind point of the fricative is also present already at the start and still at the end of the fricatives. We also aimed to analyze how the acoustic distinction is apparent during the consonants.

In the present study, we analyzed articulatory and acoustic patterns of voiced and voiceless alveolar fricatives in Hungarian. Our goals were (i) to describe the articulatory and acoustic distinction of voiced and voiceless alveolar fricatives in Hungarian, and (ii) to describe the timing relations of the articulatory and acoustic features of the voicing contrast in these fricatives.

2. Methods

2.1. Subjects

Twelve native female speakers of Hungarian were recorded. None of them had any speech or hearing impairments. Their age was between 20 and 27 years (mean: 22.25 years, sd: 1.5 years). All were given information on the procedure before the recordings both orally and written. All of them signed an informed consent before the recording.

2.2. Speech material

The analysed material is a part of a larger corpora we recorded previously (Markó et al., 2019). During the recordings, mini dialogues were introduced one by one to the subjects. Their task was to read the first utterance only for themselves (this part served as a context to the target sentence), and then to read aloud the next utterance (the target sentence) as an answer to the first one. The target sentences started with VCV# /nɛ/. Here we analysed only the /isi/ and /izi/ targets (5 per speaker). All target words occurred in focus position.

2.3. Recordings

The speech signal was recorded with a Beyerdynamic TG H56c tan omnidirectional condenser microphone at 44.1 kHz sampling rate and the tongue movement was recorded in midsagittal orientation using the "Micro" ultrasound system (Articulate Instruments Ltd.) with a 2–4 MHz / 64 element 20 mm radius convex ultrasound transducer at 83 fps. The vocal fold activity was captured by an electroglottograph (D200, Laryngograph LtD.) at 44 kHz. The speech signal was also recorded by the electroglottograph through a clipped microphone that was placed on the helmet used to stabilize the ultrasound probe at a fix distance (10-15 cm) away from the mouth. This speech signal was used to time-align the EGG- and the speech signal of the ultrasound recordings.

The segmentation of vowels was carried out by forced alignment (Mihajlik et al., 2010) and corrected manually in Praat (Boersma & Weenink, 2019), on the basis of the F2 trajectory.

2.4. Analyses

The EGG signal was analyzed in Praat (Boersma & Weenink, 2019). Voiced fragments within the target vowels were labelled automatically and corrected manually. Cessation point of voicing and restart point of voicing were also labelled. We also calculated the voiceless part ratio as the ratio of the duration of the unvoiced part to the total duration of the fricative. Additionally, the

cessation point and the restart point of the vocal fold vibration were also calculated as shown below, where V_{end} is the cessation point of voicing, $V_{restart}$ is the restart time of voicing, Fr_{start} is the start point of the fricative, and Fr_{end} is the endpoint of the fricative:

- voiceless part ratio: $((V_{restart} - V_{end}) / (Fr_{end} - Fr_{start})) * 100$
- cessation point of voicing: $((V_{end} - Fr_{start}) / (Fr_{end} - Fr_{start})) * 100$
- restart point of voicing: $((V_{restart} - Fr_{start}) / (Fr_{end} - Fr_{start})) * 100$

Figure 1 shows a partially devoiced /z/ realization. It can be seen that the intensity of the vocal fold vibration decreases as frication is superimposed on it towards the middle of the consonant, and that it ceases (first arrow) as the target is reached in the frication component. The vocal fold vibration restarts after the frication target is reached, which is supposedly caused by the release of the constriction which allows the air pressure to decrease above and in the glottis, and to reach the transglottal pressure differential that allows vocal fold vibration.

The midsagittal tongue contours were manually traced in the AAA software (Articulate Instruments Ltd.) and then extracted in the Cartesian coordinate system. Ultrasound recordings insist of tongue contour images at every 12^{th} s, therefore any tongue contour in between these time points are averaged images from the closest before and the closest after images. Therefore we did not take the tongue contours at the exact 0%, 50% and 100% time points of the consonants, but we selected the closest "real" image to these points. At the starting and endpoints, the first/last image frame within the fricative was chosen.

The present analyses addressed the "anterior", "mid", and "posterior" parts of the tongue which terms refer to the parts of the tongue that can be seen in the ultrasound tongue contour. The real parts of the tongue that these terms approximately refer to are the following. The "anterior part" corresponds to the tongue tip and/or the tongue blade. The "posterior part" corresponds to the back and the root of the tongue. The "mid part" corresponds to the tongue

body. These are, however, only approximations of the denoted regions, since (except from very few cases, where the tongue contour is very short) it cannot be reliably decided if the tongue contour seen in the ultrasound images is in fact the entire contour, or not. In the present study, several items had to be excluded from the analysis for technical reasons (no tongue contour could be detected or only a very short line appeared that was evidently only a small part of the subject's tongue surface): one /s/ realization by sp09 (all three measurement points), one 100% tongue contour in one /s/ realization by sp11, and seven 100% point tongue contours in /z/ realizations (one in the production of each of the following speakers: sp02, sp05, sp06, sp07, sp08, sp09, and sp11).

The acoustic analyses were also carried out in Praat (Boersma & Weenink, 2019). The spectral measurements were done both on the total duration of the consonant and at the three measurement points we listed above.

In order to measure the spectral characteristics for the total duration of the fricative, the fricative was extracted from the speech recording with rectangular window and transformed to spectrum slice with fast Fourier transformation. In order to measure the spectral features, the total speech recording was transformed to spectrogram using the Burg algorithm with a window length of 0.005 s, time step of 0.002 s, frequency step of 20 Hz, in the range of 0 to 21000 Hz using Gaussian window, and the spectral slice was taken at the time points to be measured with fast Fourier transformation.

The center of gravity (CoG), the standard deviation of the spectral shape (SD), the skewness and the kurtosis was calculated (2^{nd} power) at each 10% of the consonant duration between 0% and 100%.

First, the actual manner of articulation of the /z/ realizations were grouped based on the CoG and EGG-data, then the fully voiced tokens with low CoG values were checked visually to separate the voiced fricatives and the approximant-like realizations. Partially devoiced /z/ realizations were not grouped further into minor groups.

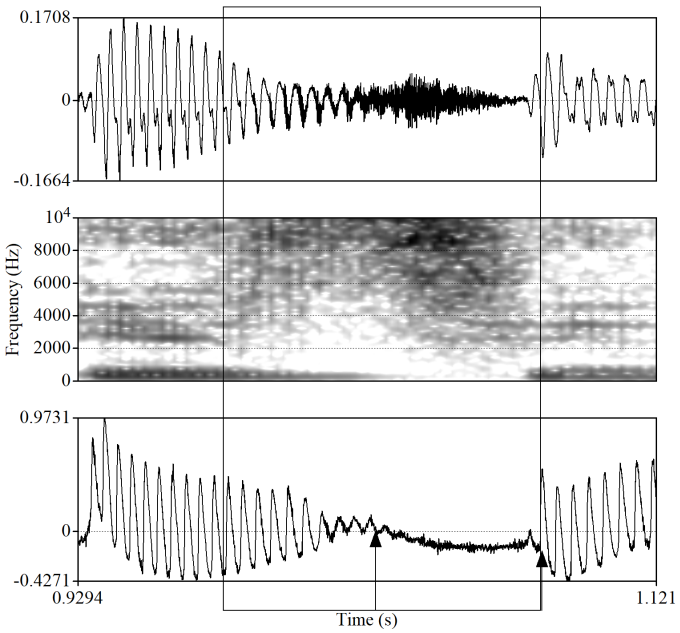


Figure 1: Partially devoiced /z/ in an /izi/ sample and its labeling. The top most box of the figure includes the oscillogram of the speech signal, the middle box includes the spectrogram of the speech signal, while the lowest includes the oscillogram of the EGG-signal. The rectangle shows the increase and decrease of the intensity of the frication noise in the speech signal. The decrease of the intensity of the vocal fold vibration can be observed. The first arrow shows the cessation of the vocal fold vibration, while the second marks the restart of voicing.

2.5. Statistical analyses

The statistical analyses were carried out in R (R Core Team, 2019).

Linear mixed models (Bates et al., 2015) were built in order to test the difference of the voiceless part ratio between the two fricatives, the cessation time, and the restart time of vocal fold vibration. In these models, voiceless part ratio, cessation time and restart time of voicing were used as the dependent variables. First a random intercept model was fitted (using the speakers) in a base model. The second model also included the consonant as fixed effect. The third one was further expanded with random slope for the consonants. The best fitting model was selected as the final model as determined on the basis of the Akaike Criterion (AIC-number) (Akaike, 1974) by using the anova() function. P-values were calculated by anova() in the lmerTest package (Kuznetsova et al., 2017).

The four spectral measures were analyzed by generalized additive mixed models (GAMM; Wood, 2017), the tongue contour was analyzed by polar GAMM (Coretta, 2019b) which is a modified version of GAMM especially for ultrasound tongue imaging. GAMM is a model that was elaborated for non-linear data, that are better described by fitting any function on the fixed effect (Wieling, 2018). This statistical approach determines the non-linear pattern automatically.

Tongue curves were analyzed by polar GAMMs in rticulate package (ultrasound tongue imaging in R; Coretta, 2019b). The models were built and compared separately for all speakers based on the suggestion of Coretta (2019a). The models were built for the temporal midpoints of the fricatives. Three models were built with maximum likelihood estimation. The horizontal placement of the measurement point (x coordinate value) was analyzed as a function of the vertical measurement point (y coordinate value). The models included a reference smooth by x-axis value and the consonant as fixed effect with the interaction between the x-axis value and the consonant.

The results of both GAMM and polar GAMM models include the comparison of the factor groups in general across the entire time interval/tongue contour, while the estimated difference can also be traced back along the time

interval/tongue contour and the phases/parts can be detected where there are differences between the factor groups. Therefore if the difference consequently appears but only in a smaller time interval/region of the tongue contour and does not lead to an overall significant difference, it still can be detected.

The statistical analysis of the acoustic data (CoG, SD, SK, and KU) was also carried out by means of GAMMs with maximum likelihood estimation (mgcv package: Wood, 2017) in R. The models included the reference smooth of time, and the consonant as fixed effect. Random effect smooth of time was included in the model. We also fitted a separate model that included the random effect smooth of the time by consonant as the fixed effect. Autocorrelation was found in the data for CoG, SD and skewness, therefore it was incorporated in the model to remove its effects.

In the case of the GAMMs and polar GAMMs, the best fitting models were also selected on the basis of AIC determined using compareML of the itsadug package (van Rij et al., 2017). The smooth curves of the tongue contours were extracted from the fitted GAMMs used for the statistics. The smooth curves of the acoustic measures were extracted from a model separated for the subjects, in that the reference smooth and the consonant were included also allowing for interaction effects.

3. Results and discussion

3.1. Voicing features based on the EGG-signal

The devoicing pattern was different across the speakers (Figure 2). Six speakers did not devoice their /z/ realizations, or only once out of the five repetitions (sp01, sp02, sp03, sp04, sp11, sp12). Five subjects often realized /z/s with devoicing (sp05, sp07, sp08, sp09, sp10), and one subject (sp06) pronounced all 5 /z/s with a voiceless part ratio above 85%. Her /s/ realizations had the lowest amount of voicing, i.e. the highest voiceless part ratio among the 12 subjects. Figure 2 illustrates the voiceless part ratio of /z/ and /s/ realizations. The distinction of these consonants is evident in all speakers, even

in the ones that tend to devoice their /z/ tokens. The linear mixed model with random slope for the consonant was found to describe the results best. The voiceless part ratio was significantly different between the two fricatives ($F(1, 11.095) = 62.820$, $p < 0.001$).

Figure 3 shows the normalized time point of the cessation and the restart of the voicing during the fricative. The voicing in the /z/ realizations of sp06, who devoiced these consonants in almost their entire duration, ceased before the 10% of the total duration was reached. The other speakers, who devoiced their /z/ realizations in the present study (sp05, sp07, sp08, sp09, and sp10) maintained voicing at least until the 25% time point of the fricative was reached. Sp10, however, did not show large variability, while the other subjects' voicing cessation time point varied. Voicing tended to restart at an earlier time point in devoiced /z/ realizations than in the /s/ realizations in those speakers' pronunciation who favored devoicing. The cessation time of the voicing was significantly different between the two fricatives ($F(1, 11.105) = 56.961$, $p < 0.001$). The restart time of voicing was analyzed in a subset of the data which includes only the subjects' who had partially devoiced /z/ realizations. Here, the model not including random slope fitted the data best, and the restart of voicing was significantly different between the two fricatives ($F(1, 52.180) = 12.879$, $p < 0.001$).

3.2. Tongue contours in the fricatives

3.2.1. Tongue contours at the mid point of the fricatives' duration

The analysis of the midsagittal tongue contours during the fricative production may reveal what maneuvers the specific speaker used to maintain vocal fold vibration. The models of polar GAMM including non-linear random slope for the consonant were found to have the lowest AIC in the case of 11 speakers, the second model (without non-linear random slope) was proven to describe data at the midpoint of the consonants the best. Each model had higher r2 than 0.95 which means that they explained at least 95% of the actual data. Table 1 includes the results for the comparison of /s/ and /z/ of these models. As seen in Table 1, the two fricatives were significantly distinct in the pronunciation of the

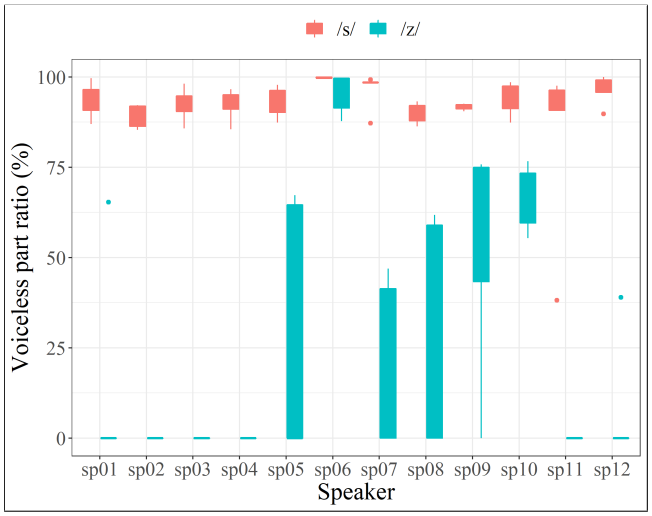


Figure 2: Voiceless part ratio (%) of /z/ and /s/

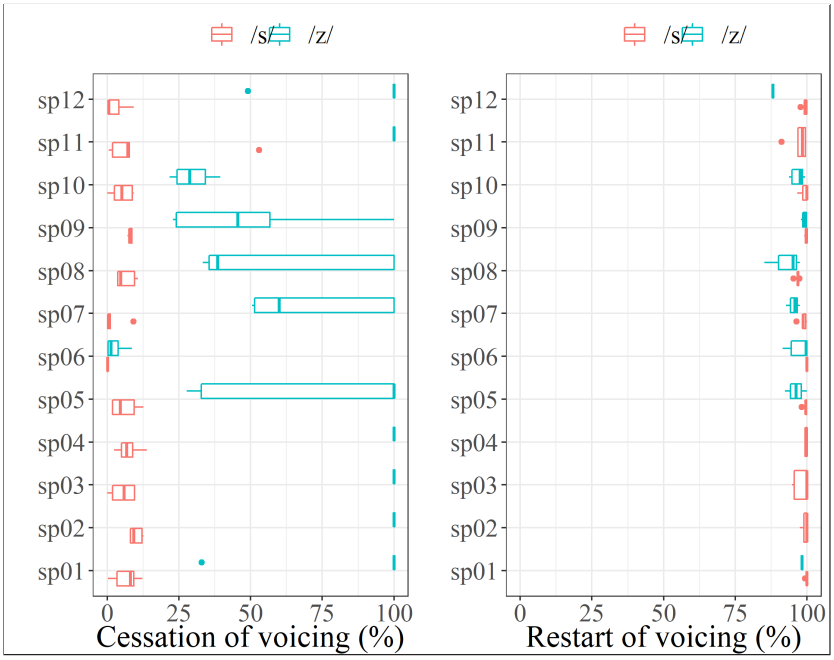


Figure 3: The normalized time of the cessation and the restart of the voicing. (The realizations that were voiced throughout their entire durations appear with a "cessation" of 100% in the left panel, but do not appear in the right panel.)

subjects sp01, sp02, sp05, sp07 and sp11 in their global tongue contours. The smooth of the models fitted on the speakers' data are shown in Figure 4. The estimated difference of the tongue contours is shown in Figure 5. The intervals in that the mean and the 95% confidence intervals of the estimated difference is not equal to zero are shown by red dashed lines. The posterior part of the tongue is on the right side of both figures. Figure 4 shows that even if only 5 out of the 12 subjects had global or large tongue contour differences between the two fricatives, there were consequent differences in some regions of the tongue contours in most subjects' pronunciation.

Table 1: The t- and p-values of the polar GAMM models for the tongue contour differences at the midpoint of the fricatives between /s/ and /z/ realizations.

	sp01	sp02	sp03	sp04	sp05	sp06	sp07	sp08	sp09	sp10	sp11	sp12
t	-2.770	-5.742	0.091	0.953	-2.213	-1.431	-7.394	1.347	-1.071	3.626	3.183	-1.317
p	*0.039*	*<0.001*	0.928	0.341	*0.028*	0.154	*<0.001*	0.179	0.286	*<0.001*	*0.002*	0.190

The posterior part of the tongue was lower in /z/ realizations, and the mid region and/or the anterior region of the tongue contours were also different between the two fricatives in 10 speakers' pronunciation. Sp10, however, produced /z/ realizations with higher vertical tongue position in the posterior tongue contour region and without further differences at the other regions of the tongue. The estimated difference of the tongue contours was very low in the case of sp11; however, it was consistent throughout the entire tongue contour.

3.2.2. Tongue contours at the start point of the fricatives' duration

The smooth of the models of polar GAMMs of the tongue contours at the start point (0% duration) of the fricatives are shown in Figure 6, and their estimated differences are shown in Figure 7. The directionality of the tongue contours is identical to those in Figure 4 and 5, i.e., the posterior part is on the right of the panels.

The t- and p-values of the polar GAMM for the difference between the tongue contours at the start point of the consonants is shown in Table 2. In sp03 and

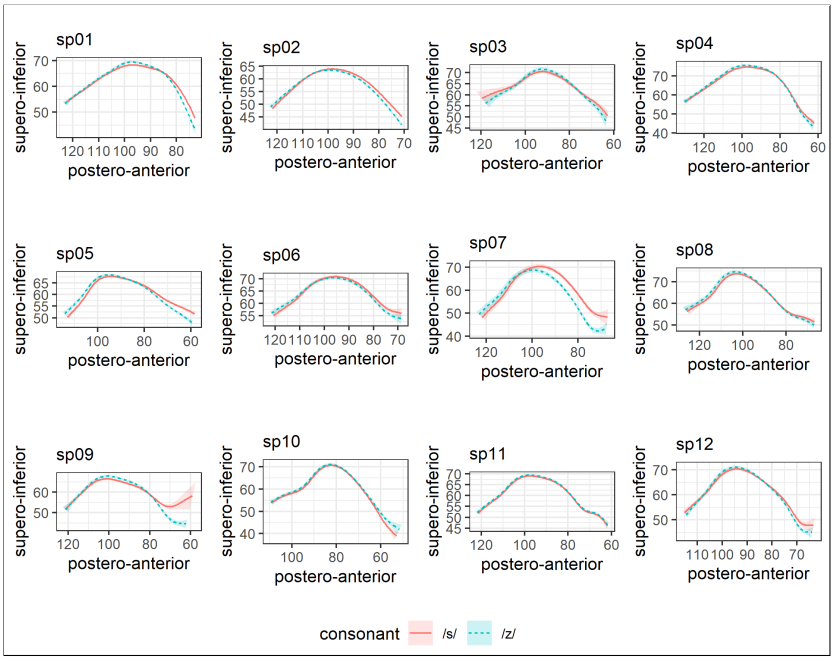


Figure 4: The smooth (mean, 95% CI) of tongue contours at the midpoint of the fricatives. The posterior part of the tongue contour is on the right side of the figure, the anterior part is on the left side.

sp12 the first, basic model yielded the lowest AIC values that did not include the consonant as factor. In the case of the other speakers, either the model including the consonant as factor with random slope for the consonants, or the one without random slopes (sp04) yielded the lowest AIC value. The best fitting models explain at least the 88.8% of the deviance.

The results for the regions of the midsagittal tongue contours are the followings. As shown in Table 2, the tongue contours at the start point of the fricative duration did not show any differences between the two consonants in five subjects' pronunciation (sp03, sp04, sp08, sp11, sp12). And according to Figure 6 and 7, in line with midpoint data, the posterior part of the tongue contour shows a difference between the consonant pairs at the start of the pronunciation in the seven further subjects' samples. This difference was relatively small in the case of sp06. In this speaker's case this small difference was the

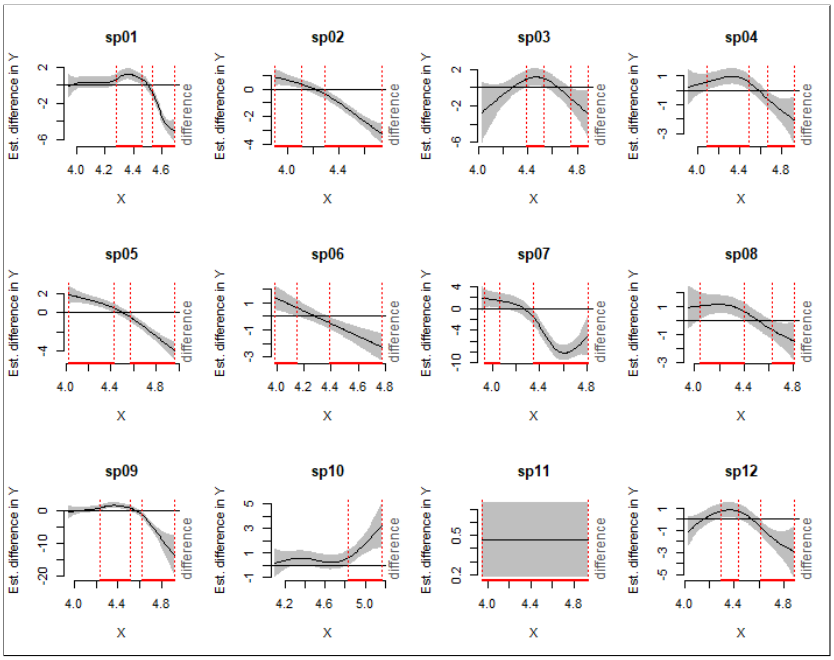


Figure 5: The estimated difference of the smoothed tongue contours at the midpoint of the fricatives. The red dashed lines indicate the intervals in which the mean and CI of the estimated differences of the contours are not equal to zero. The anterior part of the tongue contour is on the left side of the figure, the posterior part is on the right side.

only difference, while in the other seven subjects the difference was larger and appeared in the anterior and/or mid region of the tongue contour, as well. In one subject this difference did not appear at the backmost part of the tongue (sp01), but slightly anterior to that. The mid part of the tongue contour was different in five subjects' production between the two consonants at the start point of the total duration (sp01, sp02, sp07, sp09, sp10). In the case of sp02 and sp07 the entire mid-posterior part of the tongue appeared distinct at this time point. The anterior part of the tongue showed difference in four speakers' pronunciation at the start time point of the fricatives (sp01, sp02, sp05, sp07). Although in two subjects' production (sp01, sp07) only a small anterior region showed difference between the two fricatives, but a large posterior-mid region was distinct.

Table 2: The t- and p-values of the polar GAMM models for the tongue contours at the start point of the fricatives. ('-' denotes cases where that the first, basic model had the lowest AIC score.)

	sp01	sp02	sp03	sp04	sp05	sp06	sp07	sp08	sp09	sp10	sp11	sp12
t	-0.137	-6.901	-	0.560	-3.963	-0.201	-5.74	0.511	0.685	-1.303	1.493	-
p	*0.910*	*<0.001*	-	0.120	*<0.001*	0.841	*<0.001*	0.610	0.494	0.194	0.137	-

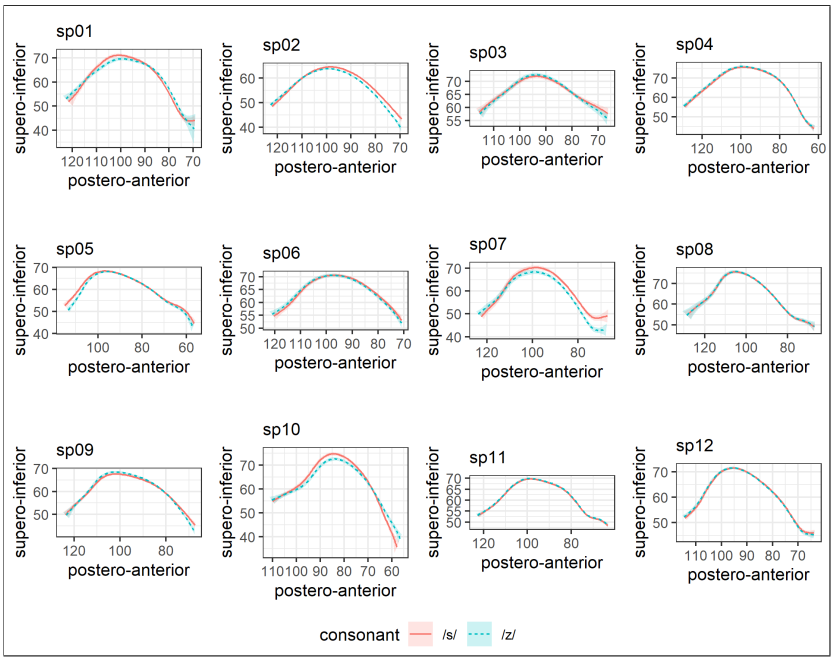


Figure 6: The smooth (mean, 95% CI) of tongue contours at the start point of the fricatives. The posterior part of the tongue contour is on the right side of the figure, the anterior part is on the left side.

3.2.3. Tongue contours at the endpoint of the fricatives' duration

The smooth of the tongue contours at the endpoint of the consonant duration are shown in Figure 8, their estimated differences are shown in Figure 9. The directionality of the tongue contours is again identical to those in Figures 4, 5, 6, 7.

The global difference between the two fricatives showed the following results. In the case of most speakers (sp01, sp02, sp05, sp06, sp07, sp09, sp10, sp11, sp12)

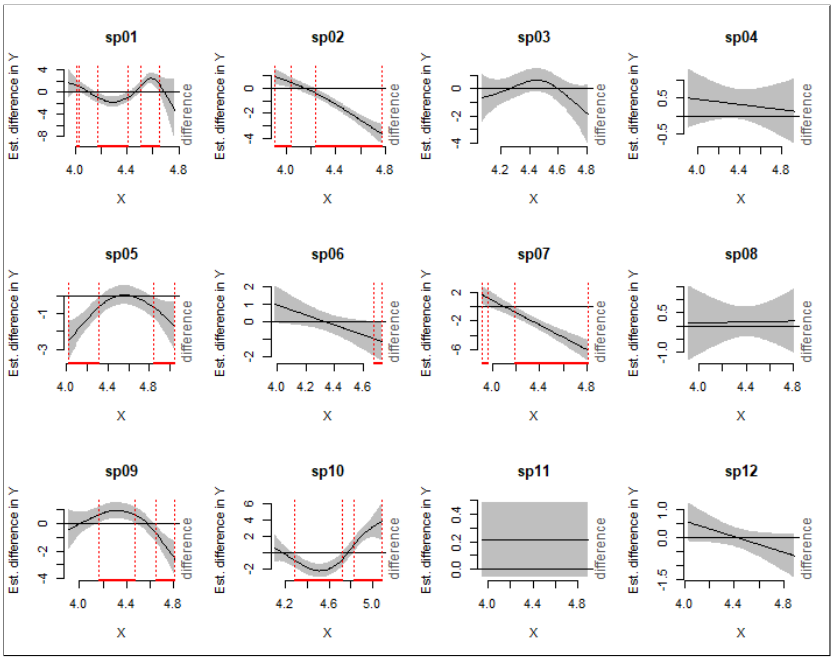


Figure 7: The estimated difference of the smoothed tongue contours at the start point of the fricatives. The red dashed lines indicate the intervals in that the mean and CI of the estimated differences of the contours are not equal to zero. The anterior part of the tongue contour is on the left side of the figure, the posterior part is on the right side.

the third polar GAMM model yielded the lowest AIC score, i.e., the model that included the consonant as factor and random slope on the consonants. In case of sp03 and sp04 the second model had the lowest AIC value in which the consonant was included as factor but no random slopes were added. In the case of sp08 the first, basic model yielded the lowest AIC score, that did not include the consonant as factor. The t- and p-values of the polar GAMMs are shown in Table 3. The results showed that the difference of the tongue contours between the two fricatives at the endpoint of the consonant was significant in six out of the twelve subjects (sp02, sp03, sp05, sp06, sp07, sp09). The best fitting models explain at least the 92.1% of the deviance.

The various regions of the tongue contours showed, however, further differences. Only three speakers (sp04, sp08, sp11) did not have any difference

between the two consonants. Difference was present in the posterior part of the tongue contours in eight subjects (sp01, sp02, sp05, sp06, sp07, sp09, sp10, sp12). However, in the case of sp01 and sp10 the difference was the opposite of that found in the other subjects', since in sp01's and sp10's production, the posterior tongue region was higher in the voiced fricative than in the voiceless one. While in the case of sp12 this was the only tongue contour region that showed any difference between the two fricatives, in sp03 it was only the anterior part that showed any difference. In the pronunciation of sp01, sp02, sp07, and sp09 the anterior part and in some cases the mid part of the tongue contour also showed a difference, and in the case of sp05 and sp10, the mid part (or a nearby region) of the tongue contour showed further differences.

Table 3: The *t*- and *p*-values of the polar GAMM models for the tongue contours at the start point of the fricatives. ('-' denotes cases where means that the first, basic model yielded had the lowest AIC score.)

	sp01	sp02	sp03	sp04	sp05	sp06	sp07	sp08	sp09	sp10	sp11	sp12
t	-1.192	-2.882	-2.220	0.542	-4.861	-2.417	-5.457	-	-4.796	-0.543	1.719	-1.385
p	0.234	*0.004*	*0.029*	0.124	*<0.001*	*0.017*	*<0.001*	-	*<0.001*	0.587	0.087	0.168

3.2.4. Comparison of the estimated differences of the tongue contours at the start, mid and endpoint of the fricatives

Comparing the tongue contours observed within the group of either /z/ or /s/ realizations at the start, mid and endpoint of the fricative duration, the following four distinct tendencies were found (compare Figure 5, 7 and 9).

a) There was no difference in the tongue contours at the start and the end of the consonants, but there was a difference at the midpoint (sp04, sp08, sp11).

b) There was no difference in the tongue contour at the start point of the consonants, but the midpoint showed a greater difference, and the endpoint still showed some, relatively smaller difference (sp12).

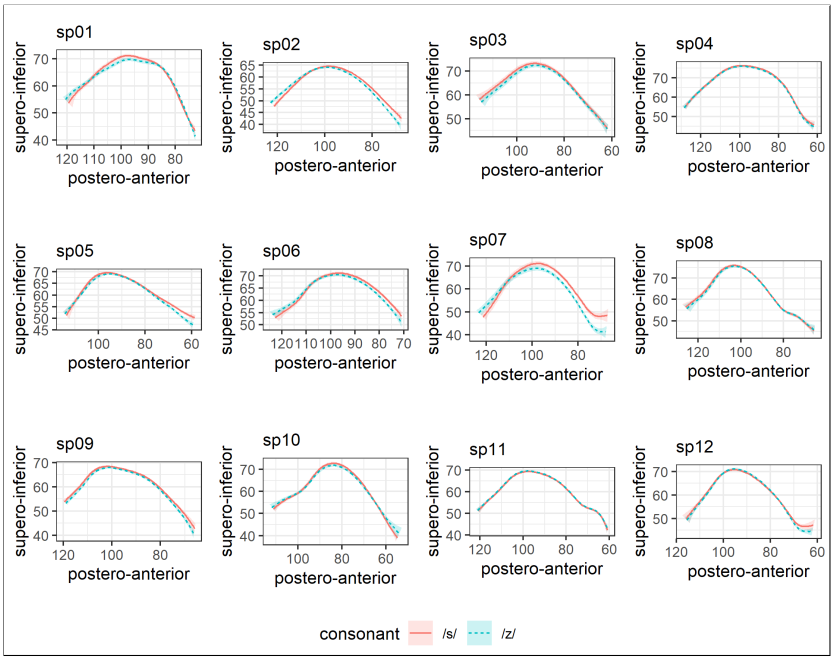


Figure 8: The smooth (mean, 95% CI) of tongue contours at the endpoint of the fricatives. The posterior part of the tongue contour is on the right side of the figure, the anterior part is on the left side.

c) There was some difference in the tongue contours at the start of the fricatives, and the difference was large at both the mid and endpoint of the consonant (sp03, sp06).

d)) Six speakers produced a great distinction in tongue contours throughout the entire fricative duration (sp01, sp02, sp05, sp07, sp09, sp10).

There were some evident tendencies in the differences also with regard to which region of the tongue was concerned.

a) In the cases, in which there was any difference between the two fricatives, the posterior part of the tongue showed a difference (with the only exception of sp03, who showed a difference elsewhere on the tongue contour at the end of her fricatives' duration. The posterior part of the tongue contour was higher in /z/ than in /s/ realizations, with the exception sp10

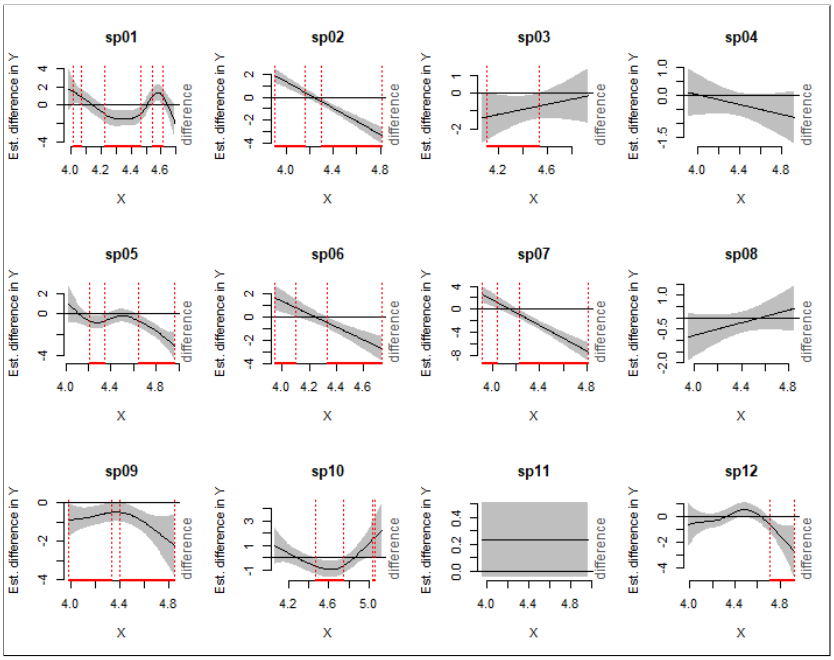


Figure 9: The estimated difference of the smoothed tongue contours at the endpoint of the fricatives. The red dashed lines indicate the intervals in that the mean and CI of the estimated differences of the contours are not equal to zero. The anterior part of the tongue contour is on the left side of the figure, the posterior part is on the right side.

who showed elevation in /z/ realizations in the entire tongue contour, and sp01 who showed elevation in /z/ realizations only at the endpoint.

b) In most cases, there was a difference also at the anterior and/or mid part of the tongue as well. If there was a difference in the position of the anterior part of the tongue, it was always raised higher in /z/ than in /s/ realizions. The difference in the mid region varied across subjects at the start- and endpoints, while it was higher in /z/s at the mid time point of the fricatives.

The subjects who showed devoicing or who did not could not be separated based on the differences in the tongue contours between the two fricatives.

3.3. Spectral measures

Realization of voiced fricatives can be diverse. For instance, in order to maintain voicing, /z/ may be realized approximant-like, but in this case there is no frication. If, however, the speaker favors to maintain the frication noise, voicing may cease during the consonant duration. Also, there are various possibilities between these two ends of the scale, in that the amount of the turbulence and the voicing can vary. Figure 10 shows two /izi/ realizations. The one on the left was produced by sp01, whose voiced fricatives appeared with voicing throughout their total duration in general. This particular realization on the left of the figure had lower intensity friction than the realization on the right side which was pronounced by sp05. This latter /z/ token was realised with partial devoicing. Their CoG at the midpoint of the token on the left of the figure was 658 Hz, while it was 6840 Hz for the token in the right.

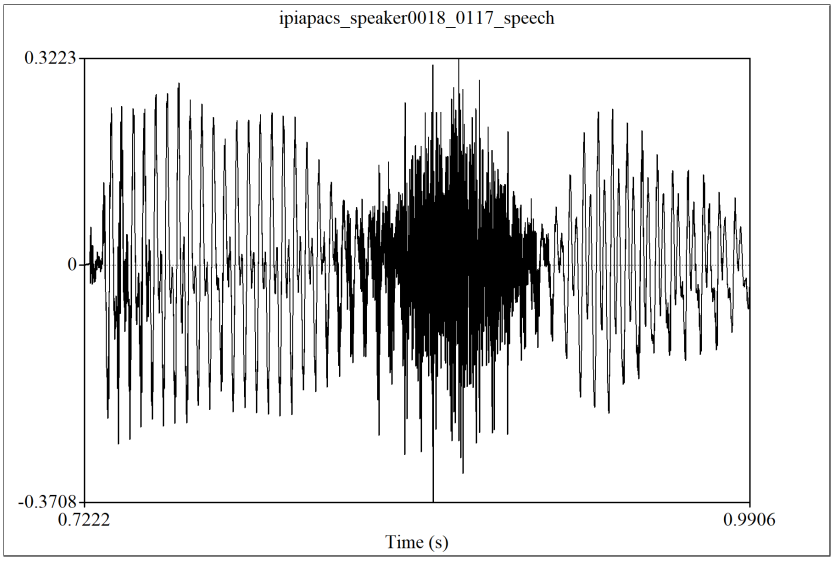


Figure 10: Oscillograms of the acoustic signal in two realizations of /izi/ (left: sp01, right: sp05).

The measurement points throughout the duration of the consonant will be short addressed as "time". The model of best fit was the one which included consonant as a factor, and a random slope for time in the case of all four acoustic

variables. Autocorrelation was detected in CoG, SD and skewness, but not in kurtosis and was involved therefore in the model. We found that in general, only CoG was significantly different between /s/ and /z/ ($t = -10.054$, $p < 0.001$), while the other three measures did not show significant differences in the global comparison of data as handled as two distinct time series. The best fitting models explain the 86.5% of the deviance in CoG, 72.3% of the deviance in SD, 82.5% of the deviance in skewness and 66.9% of the deviance in kurtosis.

3.3.1. CoG

The CoG curve of /s/ realizations was fairly similar across the speakers (Fig. 11), an abrupt increase appeared in the first 10-20% of the duration and an abrupt decrease in the last 10-20% of the consonant, while the middle part showed a slow change or plateau. This was expected as the CoG in voiceless fricatives is largely affected by the reach of the target constriction and then by the release which leads to the target of the following vowel. The realizations of /z/, however, showed various patterns (Figure 11). In eight speakers , the CoG slowly increased until the 50% of the consonant duration, then it showed the same pattern as /s/ realizations (slow increase in CoG: sp03, sp05, sp07, sp08, sp09, faster increase in CoG: sp01, sp06, sp10). The distinction between the two fricatives disappeared or was very low from the 50-60% of the consonants durations (sp01, sp03, sp06, sp07, sp09), and from approx. 30% in sp10. As observed in data on the temporal organization of voicing, some of these speakers frequently devoiced their /z/ realizations (e.g., sp06), or varied in their devoicing pattern (e.g. sp07), while others had voicing throughout the entire duration of /z/ realizations (sp01, sp02, sp03, sp04, sp11, sp12). Four speakers (sp02, sp04, sp11, sp12) /z/ tokens were often realized with a very low CoG in its entire duration, that either could be a result of approximant-like, or that of a very low intensity turbulent noise with voicing throughout the fricative's duration (as observable in the left panel of Figure 10).

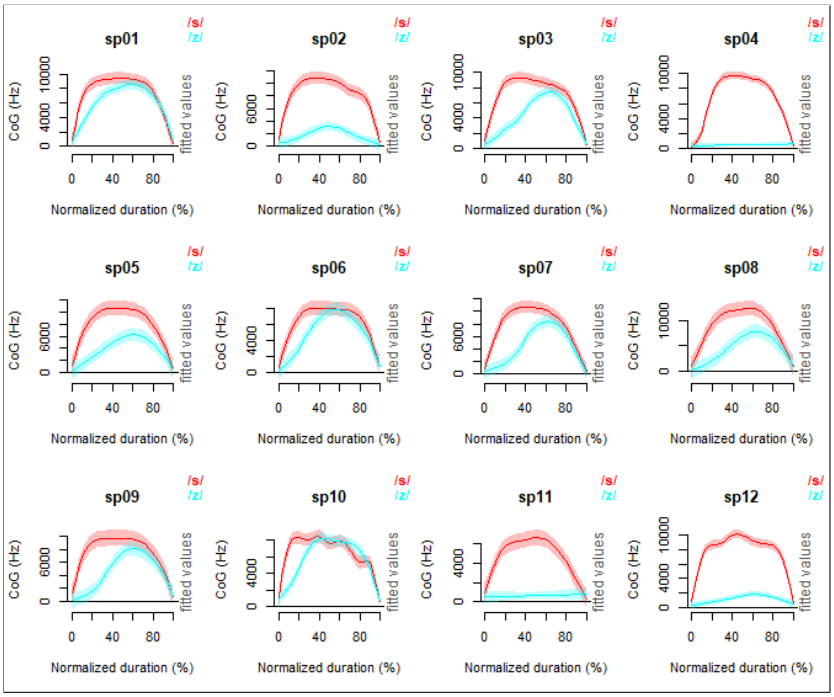


Figure 11: Fitted smooth curves of CoG by speaker (mean, 95% confidence interval)

3.3.2. Spectral variability (SD)

Spectral SD is shown in Figure 12. In the case of /s/ the curve was rather similar across speakers, which is expected and explicable as follows. While voicing ceases abruptly and the turbulent noise gets more intense, the CoG increases, and the higher frequency regions become more dominant in the signal than the lower regions. This leads to a fast increase in SD at the start of /s/. The constriction release, and thus the loss of dominance of the higher frequency regions leads to a fast decrease not only in CoG, but also in SD at the end of the fricative, and the interval between the 10-to-90% of the duration showed a valley in most speakers in /s/. After the cessation of voicing and the reach of the target/maximum CoG, the low-frequency region of the spectrum does not add to the variability of the spectrum. As a result of the above, in general, the shape of the spectral SD in /s/ showed the following trend: an abrupt increase followed by a somewhat variable valley, and then by an abrupt increase again.

In the case of sp10 and sp11 the first, increasing part of the above outlined tendency was slow, and the first SD maximum was reached at the point where the second decrease was expected based on the other speakers' results. The valley did not appear in their case, but the abrupt decrease following the slow increase did.

The realizations of /z/ showed a dome shape in the production of six speakers (sp02, sp04, sp05, sp09, sp11, sp12), while in all the further speakers', a valley appeared in these tokens as well, similarly to /s/ realizations. The shape of the spectral SD in /z/ can be explained by voicing and CoG results taken together: the cessation of voicing and the appearance of lower and higher frequency components.

As we found the spectral SD time series curves to be variable, we can conclude on no systematic tendencies with regard to the distinction of the voicing contrast in spectral SD.

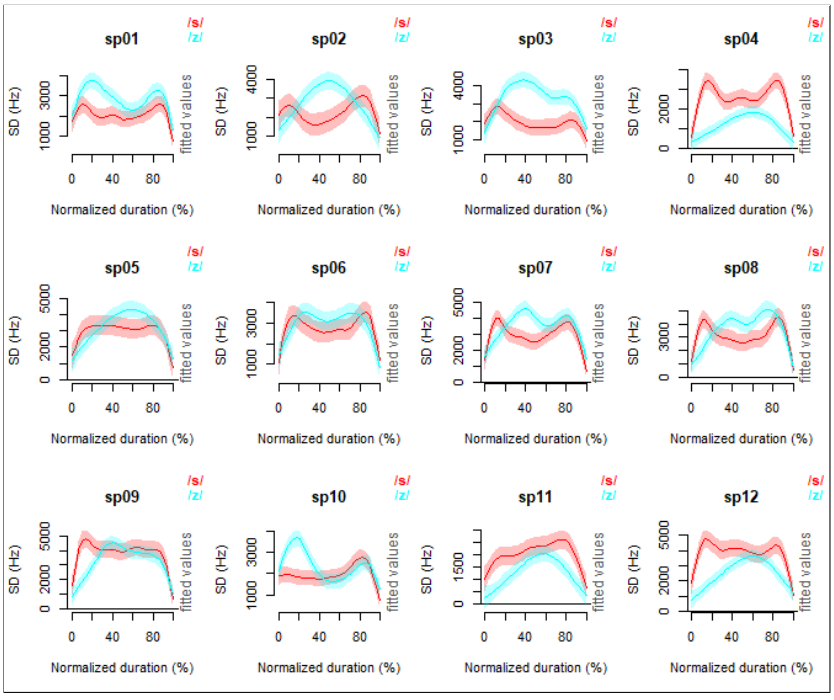


Figure 12: Fitted smooth curves of SD by speaker (mean, 95% confidence interval)

3.3.3. Skewness of the spectral components

The skewness decreased abruptly at the start of the consonants and increased abruptly at the end (Fig. 13). This value was less affected by voicing even in the voiced fricatives: /z/ realizations showed less abrupt decrease at their starting phase; however, the overlap between the members of the consonant pair was reached in the first 30% to 40% of the fricatives regardless of the speakers' devoicing tendencies. The two consonants were distinct throughout their entire duration only in four speakers (sp02, sp04, sp11, sp12). Each of these four speakers had voicing throughout the entire duration of /z/ realizations, and showed low CoG, dome shaped spectral SD, and low ratio of overlap between the two fricatives in spectral SD.

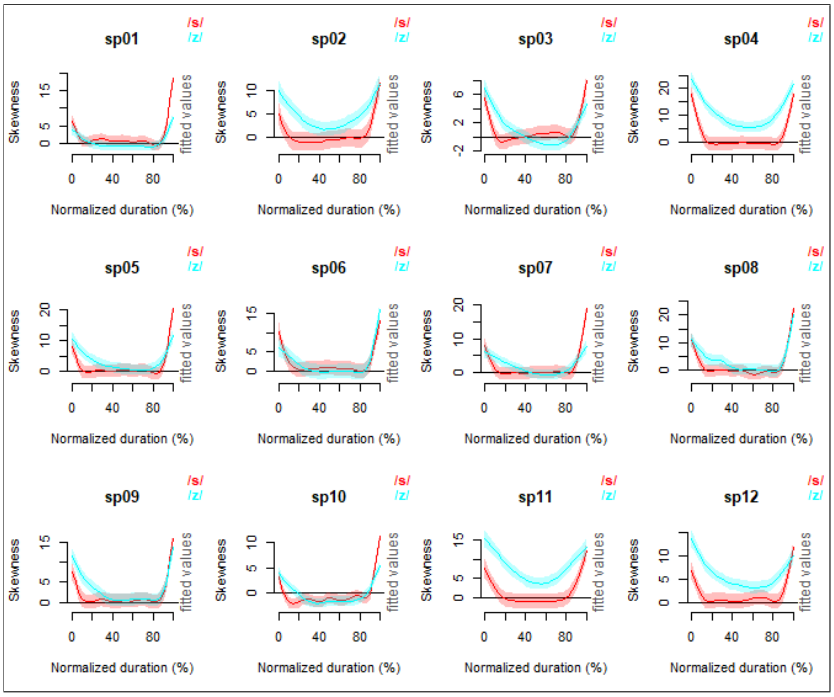


Figure 13: Fitted smooth curves of skewness (mean, 95% confidence interval)

3.3.4. Kurtosis of the spectral components

Compared to the other parameters, kurtosis stayed fairly constant throughout the entire consonant duration (Fig. 14). With respect to kurtosis values in first 10-20% of the fricative duration three distinct tendencies were found: i) it showed either minor or no decrease in the two fricatives (sp01, sp05, sp06, sp07, sp08, sp09), ii) it showed great decrease in both fricatives (sp03, sp04), or iii) it showed smaller decrease in /s/ realizations and larger decrease in /z/ realizations (sp02, sp11, sp12). In five speakers (sp02, sp03, sp04, sp11, sp12), kurtosis of the members of the fricative pairs was different but only in a smaller portion of the total consonant duration, typically at the start of the fricatives , while in other speakers /s/ and /z/ realizations did not show any difference in the kurtosis.

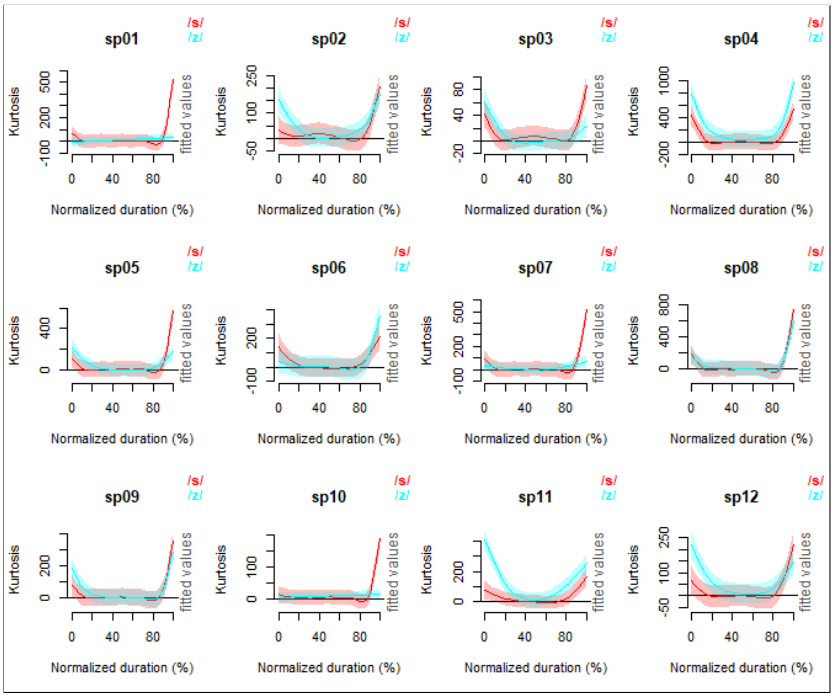


Figure 14: Fitted smooth curves of kurtosis (mean, 95% confidence interval)

4. Conclusions

In the present paper we aimed to (i) describe the articulatory and acoustic distinction of voiced and voiceless alveolar fricatives in Hungarian, and (ii) to describe the timing relations of the articulatory and acoustic features of the voicing contrast in these fricatives.

Since the fricatives in question were analyzed in intervocalic position, the voicing contrast was hypothesized to be apparent in vocal fold vibration throughout the entire consonant duration in most cases in /z/ realizations, i.e., the voiceless part ratio was expected to stay low, close to 0% in /z/. Earlier studies showed that the voicing characteristics of voiced obstruents are diverse across languages and speakers (e.g., Shih et al., 1999), and that fricative realizations in Hungarian were also diverse across speakers both in read and spontaneous speech (e.g., Bárkányi & Kiss, 2009; Gráczi, 2012). In line with these results, interspeaker variation was also demonstrated in the present study revealing speakers not devoicing their phonologically voiced fricatives, or partially devoicing some of their phonologically voiced fricatives, and one speaker who devoiced all of his/her phonologically voiced fricatives. Nevertheless, the devoiced part ratio was lower in /z/ realizations than in /s/ realizations in all 12 subjects, and voicing ceased later and restarted earlier in voiced fricatives than in voiceless ones. This latter finding is the result of the fact that the target in /z/ is voicing, and the reason of the cessation of voicing in /z/ is the intraoral pressure build up (Bickley & Stevens, 1986; Stevens, 1997), while in /s/ the target is voiceless. Therefore, the vocal fold settings may be different in the two consonants even in those fragments where no vocal fold vibration is found, and the voiced phoneme realizes as devoiced. While the restart of voicing in /z/ depends on the decrease of intraoral pressure, as the vocal folds are already set for voicing, in the case of /s/, the restart of voicing is primarily controlled by the voicing gesture of the following vowel.

Articulatory results of previous MRI-, EPG- and ultrasound-studies (e.g., Rothenberg, 1967; Kent & Moll, 1969; Perkell, 1969; Westbury, 1983; Ahn, 2018;

Coretta, in press) provided evidence of the articulatory maneuver of advancing tongue root in voiced fricatives to avoid cessation of voicing. This articulatory maneuver broadens the posterior region of the oral/pharyngeal cavity which results in slower intraoral pressure build up and thus in slower/less frequent devoicing in phonologically voiced fricatives. In our study, this maneuver was also demonstrated, as we found lower position of the posterior region of the midsagittal tongue contour in most speakers in /z/, and this region approximately corresponds the root of the tongue. However, while Coretta (in press) found that the advanced tongue root was already present in the preceding vowel in the case of voiced stops, in the present study we did not find this difference at the starting point of fricatives in seven out of twelve (i.e., in 58.3% of the) speakers. The difference in these results might be a consequence of a difference in the phonetic realisation of the voicing contrast across languages. However, this assumption is to be analyzed in future studies which extend their scope beyond the one vowel context of /i_i/. In the present data we also found that the difference in the vertical position of the posterior region of the tongue contours (i.e., the alleged advanced tongue root) was present in most subjects also at the end of the fricatives. This may be due to the fact that the volume expansion of the vocal tract at the end of the fricative may also help the earlier restart of voicing in the phonologically voiced fricatives, which were realised as devoiced.

At the midpoint of the fricatives, the anterior or mid region of the tongue contour was higher in /z/ than in /s/ in most speakers who showed lower posterior regions in /z/ than is /s/. This tendency was also observable at the boundaries of the fricative, although the position difference of the mid region was the opposite (and showed higher position in /s/) in some speakers'. Although we cannot exactly tell if these 'anterior' points always reflect the position of the tongue tip or the tongue blade (as the ultrasound is not always able to show the entirety of the midsagittal view of the tongue surface), we may claim that our results are in line with those of Narayanan and colleagues' (1995). The MRI-study of Narayanan and colleagues' (1995) revealed that the alveolar fricatives in English may be articulated with either the blade or the tip. Stevens

and colleagues (1992) found smaller cross-sectional area at the oral constriction than at the glottis for voiced fricatives. These results are also in line with our results, as the higher position of the anterior regions we found here does also suggest a similarly narrow constriction, which may provide support to the fricative to reach its target frication intensity through the lower pressure build up (Fuchs et al., 2007). The higher position in the mid region of the tongue contour in /z/? partly contradict previous results, as we expected a larger oral volume, that is, a lower tongue position behind the obstacle, which would help to maintain voicing (Docherty, 1992; Fuchs & Perrier, 2003).

Three further questions regarding the articulatory maneuvers speakers may utilize to avoid devoicing in phonologically voiced fricatives cannot be addressed in the present study, in which we analyzed 2D midsagittal ultrasound data with start, mid and end point measurements. We cannot determine i) the groove size, ii) the area of the tongue-palate contact, nor iii) the timing of the oral articulation of the fricatives at hand. EPG-studies of English and Croatian showed that /s/ realizations needed longer time to achieve the largest contact surface at the constriction in both languages than /z/ realizations (Liker & Gibbon, 2013, 2018). The tongue-palate contact was found to correlate with the occurrence of devoicing (Fuchs et al., 2007).

The GAMM analysis of acoustic parameters in /z/ and /s/ as two sets of time series data taken at eleven measurement points showed, that only CoG was significantly different between the /z/ and /s/ realizations in general. The detailed analysis of the time series data, however, showed that there are differences in the time course of each analyzed spectral measure with a considerable variability across speakers in /z/ realizations, and much less variability in /s/ realizations. /s/ realizations had an abrupt increase in CoG, and in spectral variability and an abrupt decrease in skewness in the first 10-20% of their duration, while the change was opposite in its direction, but similarly abrupt at the final 10-20% of the fricatives. Kurtosis varied with regard the abruptness in the first 10-20%. In the middle portion of the fricative we found a dome like shape in CoG and most often a valley in the other spectral measures.

The CoG shapes of /z/ realizations varied across the subjects. Taken together, spectral results suggest that voiced and partially voiced fricative realizations both occurred, as well as approximant-like realizations. However, devoiced part ratio data and the midsagittal tongue contour shape together are not enough to describe the articulatory-acoustic relationships. Most importantly, articulation and acoustics show aquantal relationship (see Stevens, 1968), but two further reasons are also important to mention here. One of these is that the measurement of the distance of tongue and the palate is difficult and unreliable in ultrasound images as the palate cannot be traced in detail (despite the use of wet swallowing or other tricks that may partially reveal the palate contour), but neither can we be sure if an ultrasound image includes the entirety of the midsagittal tongue surface. The second main reason is that the earlier EPG-studies not only revealed differences of the grooves, tongue palate contact and its timing between the voicing counterparts (e.g. Liker & Gibbon, 2011, 2013, 2018; Fuchs et al., 2007), but also among the subjects Liker & Gibbon (2011). This variability cannot be analyzed in midsagittal tongue contours.

In this study, we conducted a pioneering work on investigating the articulation and acoustics of Hungarian alveolar fricatives. We demonstrated that the phonetic realisation of the voicing contrast in Hungarian /s/ and /z/ requires an appropriate laryngeal-oral coordination, as shown by previous studies, and is not merely the result of differences at the laryngeal level (e.g. Narayanan et al., 1995; Fuchs et al., 2007; Coretta, in press). We showed half of our twelve speakers showed some devoicing of phonologically voiced fricatives, while half of them showed no devoicing at all. We also replicated previous findings revealing an important role of tongue root displacement in the maintenance of voicing in phonologically voiced fricatives, but we also described further distinct articulatory strategies that may aid this articulatory/acoustic goal. Our results may contribute greatly to our knowledge on speaker-dependent phonetic variability of fricative voicing.

References

Ahn, S. (2018). The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics*, *71*, 451–467. doi:10.1016/j.602wocn.2018.10.003.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Bárkányi, Zs., & Kiss, Z. (2009). Word-final fricative contrasts in Hungarian. A phonetic approach. URL: http://budling.nytud.hu/~cash/papers/buphoc09-slide.pdf Előadás a BuPhoC 2009. nov. 5-i ülésén.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bickley, C. A., & Stevens, K. N. (1986). Effects of a vocal tract constriction on the glottal source: experimental and modeling studies. *Journal of Phonetics*, *14*, 373–382.

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer*. URL: http://www.fon.hum.uva.nl/praat/download_win.html. Downloaded: 2019. october 1.

Coretta, S. (2019a). *Assessing mid-sagittal tongue contours in polar coordinates using generalised additive (mixed) models*. doi:10.31219/osf.io/q6vzb. pre-print page: https://www.researchgate.net/publication/335475997_Assessing_mid-sagittal_tongue_contours_in_polar_coordinates_using_generalised_additive_mixed_models.

Coretta, S. (2019b). *rticulate: Ultrasound Tongue Imaging in R. R package version 1.5.0.9000*. URL: https://github.com/stefanocoretta/rticulate.

Coretta, S. (in press). Longer vowel duration correlates with greater tongue root advancement at vowel offset: Acoustic and articulatory data from Italian and

Polish. *Journal of Acoustic Society of America*, . Preprint downloaded from https://osf.io/zrqyx.

Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents*. Berlin – New York: Foris Publications.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fletcher, S. G. (1989). Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech and Hearing Research*, *32*, 736–748.

Fuchs, S., Brunner, J., & Busler, A. (2007). Temporal and spatial aspects concerning the realizations of the voicing contrast in German alveolar and postalveolar fricatives. *Advances in Speech–Language Pathology*, *9*, 90–100.

Fuchs, S., & Perrier, P. (2003). An EMMA/EPG study of voicing contrast correlates in German. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (p. 1057–1060). Barcelona.

Gráczi, T. E. (2012). *Zörejhangok akusztikai fonetikai vizsgálata a zöngésségi oppozíció függvényében. [Acoustic characteristics of obstruents with regard to voicing opposition]. PhD-thesis*. Budapest: Eötvös Loránd University. URL: http://doktori.btk.elte.hu/lingv/graczitekla/diss.pdf.

Kent, R. D., & Moll, K. L. (1969). Vocal–tract characteristics of the stop cognates. *Journal of the Acoustical Society of America*, *46*, 1549–1555. doi:10.1121/1.1911902.

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.

Liker, M., & Gibbon, F. (2011). Groove width in Croatian voiced and voiceless postalveolar fricatives. In *Proceedings of the 18th International Congress of Phonetic Sciences. Hong Kong* (p. 1238–1241).

Liker, M., & Gibbon, F. (2013). Differences in EPG contact dynamics between voiced and voiceless lingual fricatives. *Journal of the International Phonetic Association*, *43*, 49–64. doi:10.1017/S0025100312000436.

Liker, M., & Gibbon, F. (2018). Tongue-Palate Contact Timing during /s/ and /z/ in English. *Phonetica*, *75*, 110–131.

Markó, A., Bartók, M., Csapó, T. G., Deme, A., & Gráczi, T. E. (2019). The effect of focal accent on vowels in Hungarian: articulatory and acoustic data. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (p. 2715–2719). Melbourne, Canberra: Australasian Speech Science and Technology Association Inc. URL: http://intro2psycholing.net/ICPhS/papers/ICPhS_2764.pdf.

McLeod, S., Roberts, A., & Sita, J. (2006). Tongue/palate contact for the production of /s/ and /z/. *Clinical Linguistics and Phonetics*, *20*, 51–66.

Mihajlik, P., Tüske, Z., Tarján, B., Németh, B., & Fegyó, T. (2010). Improved recognition of spontaneous Hungarian speech: Morphological and acoustic modeling techniques for a less resourced task. *IEEE Transactions on Audio, Speech and Language Processing*, *18*, 1588–1600.

Müller, E. M., & Brown Jr., W. S. (1980). Variations in the supraglottal air pressure waveform and their articulatory interpretation. *Speech and Language: Advances in Basic Research and Practice*, *4*, 317–389.

Narayanan, S. S., Alwan, A. A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America*, *98*, 1325–1347.

Ohala, J. J., & Solé, M. J. (2010). Turbulence and phonology. In S. Fuchs, M. Toda, & M. Zygis (Eds.), *Turbulent sounds: An interdisciplinary guide* (p. 37–102). Berlin & New York: De Gruyter Mouton.

Perkell, J. S. (1969). Physiology of speech production: Results and implication of quantitative cineradiographic study. *M.I.T. research monograph*, *53*.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/.

Rothenberg, M. (1967). The breath-stream dynamics of simple-released-plosive production. 7186 (Basel: Biblioteca phonetica).

Shadle, C. H. (1991). The effect of geometry on source mechanisms in fricative consonants. *Journal of Phonetics*, *19*, 409–424.

Shih, C., Möbius, B., & Narasimhan, B. (1999). Contextual effects on consonant voicing profiles: A cross-linguistic study. In *Proceedings of the 14th International Congress of Phonetic Sciences* (p. 989–992). San Francisco, CA.

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics*, *25*, 471–500.

Stevens, K. N. (1968). *The Quantal Nature of Speech: Evidence from Articulatory-acoustic Data*. Northwestern University.

Stevens, K. N. (1997). Articulatory–acoustic–auditory relationships. In W. J. Hardcastle, & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (p. 462–506). Oxford: Blackwell.

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of Acoustic Society of America*, *91*, 2979–3000.

Svirsky, M., Stevens, K. N., Matthies, M., Manzella, J., Perkell, J., & Wilhelms-Tricarico, R. (1997). Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America*, *102*, 562–571.

Tabain, M. (2001). Variability in fricative production and spectra: Implications for the hyper- and hypo- and quantal theories of speech production. *Language and Speech*, *44*, 57–94.

van Rij, J., Wieling, M., Baaye, R., & Rijn, H. (2017). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.3.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *The Journal of the Acoustical Society of America*, *73*, 1322–1336.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between l1 and l2 speakers of english. *Journal of Phonetics*, *70*, 86–116.

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

Ajakvideó alapú beszédszintézis konvolúciós és rekurrens mély neurális hálózatokkal

Rácz Bianka[1], Csapó Tamás Gábor[1,2]

[1] *Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék*
[2] *MTA-ELTE „Lendület" Lingvális Artikuláció Kutatócsoport*

Abstract

Articulatory-to-acoustic mapping methods have the aim to convert articulatory movement to acoustic speech signal. For articulatory acquisition, complex techniques (e.g. ultrasound, MRI) are suitable – but also, the lip movement contains relevant information about the speech sounds. There have been several studies applying deep neural networks for the lip-to-speech problem, and also for automatic lipreading. Inspired by the earlier studies, in this paper we designed and implemented models that can generate spectral parameters of speech from lip videos. Later, from the predicted spectral parameters, we synthesized the speech using a vocoder. For the experiments, we used 1000 sentences from a male English speaker of the GRID audiovisual database, which contains video from the face of speakers, and synchronous speech. Based on the literature, we extended the baseline deep neural network model and identified two models that use convolutional and recurrent layers. The convolutional network has single images as input, whereas the recurrent network can take into account the sequential nature of the input data: it has eight consecutive face images as input. We compared these two new models to the original baseline model in a multi-step experiment. In an objective test, we generated speech by the vocoder and by the DNN models. We calculated the Mel Cepstral Distortion between synthesized and reference sentences, and found that the recurrent model has significantly lower error than the baseline FC-DNN, while the output of the convolutional model was not better. After this we collected several subjects' opinions during an online subjective test. They had to evaluate how natural the speech utterances they heard sounded. Similarly to the objective experiment, in the subjective test the recurrent neural network (which takes eight consecutive images as input) was preferred. The results might be useful for application in Silent Speech Interfaces or for lipreading systems.

1. Bevezetés

Az artikuláció (a beszédképző szervek koordinált mozgása) és az akusztikum (a keletkező beszédjel) kapcsolata az 1700-as évek óta foglalkoztatja a beszédkutatókat (Kempelen, 1791). Ahhoz, hogy a beszédképző szervek (pl. hang-

Email addresses: rczbianka1@gmail.com (Rácz Bianka), csapot@tmit.bme.hu (Csapó Tamás Gábor)

Beszédtudomány – Speech Science 2020. 57–72. DOI: 10.15775/Besztud.2020.57-72

szalagok, nyelv, lágy szájpad) mozgását vizsgálni tudjuk, speciális eszközökre van szükségünk, mivel a legtöbb ilyen szerv nem látható folyamatosan beszéd közben. Az artikulációs szervek közül a szükséges eszközök tekintetében a legegyszerűbb az ajakmozgás vizsgálata, hiszen ehhez egy egyszerű videokamera is elegendő, amely a beszéd közbeni arcmozgást rögzíti.

Magyarországon többek között Bolla kísérletezett az ajkak (fotolabiogram) vizsgálatával, viszont a vizsgálatok csak statikus állóképeken alapultak (Bolla, 1995). Az MTA-ELTE „Lendület" Lingvális Artikuláció Kutatócsoport eszközeivel 2016 óta több magyar beszélőtől is rögzítettünk dinamikus nyelvultrahang és ajakvideó felvételeket (Csapó et al., 2017a). Emellett nemzetközi szinten egyre több adatbázis áll rendelkezésre, melyek alapján az ajakmozgás vizsgálható – pl. a GRID korpuszban 34 beszélőtől rögzítettek 1000–1000 rövid mondatot angol nyelven (Cooke et al., 2006).

1.1. Artikuláció-akusztikum átalakítás ajakvideó alapján

Az artikuláció-akusztikum konverziós módszerek célja, hogy artikulációs mozgás alapján szintetizáljanak beszédet. Az artikulációs információ lehet például a nyelv mozgása ultrahanggal rögzítve (Csapó et al., 2017c,b). A konverzió egy másik lehetséges megoldása a beszédszintézis kizárólag egy arcról vagy ajakról készült videó képkockáiból. A megoldás kétféle megközelítéssel lehetséges: 1) közvetlen 'lip-to-speech', 2) közvetett 'lip-to-text', majd 'text-to-speech' lépésekben. A közvetlen módszerek gyorsabbak, hiszen nincs szükség külön szövegfelolvasó modulra. Ezekre mutat példát Le Cornu & Milner (2015), Ephrat & Peleg (2017), és Akbari et al. (2018). A közvetett módszerek tulajdonképpen automatikus szájról olvasást végeznek, például Wand et al. (2016) és Sun et al. (2018).

A némabeszéd-interfész (Silent Speech Interface) az artikuláció-akusztikum konverziós módszerek egy olyan távlati alkalmazása, amelynek használatával némán beszélve, „tátogva" adhatunk ki hangot (Denby et al., 2010; Csapó et al., 2017c; Kimura et al., 2019). A némabeszéd-interfésszel segíthetünk olyan embereknek kommunikálni, akik egy betegség vagy baleset következtében elvesztették

a hangalkotási képességüket, viszont még tudnak artikulálni. De nem csak az egészségügyben használhatjuk ezt az eszközt. A mindennapi életben is hasznos lehet, ha egy megbeszélésen ülve hang nélkül tudunk válaszolni egy telefonhívásra anélkül, hogy megzavarnánk a társainkat.

A jelen kutatás célja egy olyan rendszer létrehozása, amely az ajakról készült videofelvételekből beszédet tud szintetizálni. Ehhez mély neurális hálón alapuló gépi tanulást alkalmaztunk, melynek bemenete az arcról készült videó volt, kimenete pedig a beszéd spektrális paraméterei. A gépi tanulás által becsült spektrális paraméterekkel egy vokóder használatával mondatokat szintetizáltunk. Az így szintetizált beszéd ugyan nem érthető teljesen, de sok esetben szótöredékek vagy szavak is érthetőek lettek, így a kezdeti eredményeket biztatónak tartjuk.

2. Módszerek

A jelen cikkben bemutatjuk Rácz (2019) szakdolgozata során készült kutatás módszereit és eredményeit.

2.1. Felvételek és adatok

A videókat a GRID adatbázisból töltöttük le (Cooke et al., 2006), amely nyilvánosan elérhető: `http://spandh.dcs.shef.ac.uk/gridcorpus/`. Minden videón egy embert látunk szemből, ahogy egy hat szavas mondatot mond el angol nyelven. A videókat 25 képkocka/sebességgel rögzítették. Minden videó 3 másodperc hosszú, és 75 képkockából áll. A GRID adatbázis 34 beszélőjéből csak egyet választottunk ki a jelen cikk demonstrációs kísérleteihez: az S3-as beszélővel készített felvételeket használtuk fel az egy-beszélős modellek tanításához. Az S3 beszélő választásának oka, hogy Ephrat & Peleg (2017) is ezen beszélő felvételeivel végzett hasonló kísérletet. Az 1. ábra bal oldalán látható egy képkocka az S3 beszélő eredeti videójából. Az 1000 videóból az utolsó 10-et választottuk a teszteléshez és a maradék 990 videó 80%-át használtuk a tanításhoz, 20%-át pedig a validáláshoz.

1. ábra. Bal oldal: egy képkocka a GRID adatbázis S3 beszélő eredeti videójából; jobb oldal: egy képkocka az előfeldolgozás után.

2.1.1. Az ajakvideó előfeldolgozása

A bemeneti képek előfeldolgozása során automatikus módszerekkel (az OpenCV modullal, 'Haar Cascades' alapon) kivágtuk az arcot a képből, és az így keletkezett képkockákat szürkeárnyalatossá, és 128×128-as méretűvé konvertáltuk. A felvételek egységességét (pl. fényerő változások) nem vizsgáltuk. Az eredményre az 1. ábra jobb oldalán látható példa. A szürkeárnyalatos képek pixelei képezték a neurális hálózatok bemenetét.

2.1.2. A beszédjel előfeldolgozása

A beszédjel paraméterekre bontására és a későbbi visszaállításra egy egyszerű vokódert választottunk, a korábbi ultrahangos kísérleteinkhez hasonlóan (Csapó et al., 2017c,b). Először spektrális elemzést végeztünk mel-általánosított kepsztrum (Mel-Generalized Cepstrum, Line Spectral Pair, MGC-LSP, Tokuda et al., 1994) módszerrel, melyet statisztikai parametrikus beszédszintézisben széles körben használnak. Az elemzéshez 12-ed rendű MGC-t számítottunk $\alpha = 0{,}42$ és $\gamma = -1/3$ értékekkel – ezen paraméterek széles körben használtak statisztikai parametrikus beszédszintézishez (Csapó & Németh, 2014; Drugman et al., 2009). Ahhoz, hogy a beszédjel analízise során kapott paraméterek szink-

ronban legyenek az arcképekkel, a kereteltolást 1 / FPS értékre választottuk (40 ms). A viszonylag nagyméretű kereteltolás (szemben a statisztikai parametrikus beszédszintézisben tipikusan használt 5 ms-os nagyságrenddel, Csapó & Németh, 2014) önmagában is ronthatja az újraszintetizált beszéd minőségét, azonban az audiovizuális adatok szinkronitásához ez volt a célszerű választás. A gépi tanulás kimenete tehát a fenti vokóder 13-ad rendű spektrális paraméterei voltak.

A beszéd visszaállításához fehérzaj gerjesztést generáltunk, majd a gerjesztést és az MGC-LSP paramétereket felhasználva MGLSADF szűrővel (Imai et al., 1983) visszaállítottuk a szintetizált beszédet (suttogás jellegű beszédet generálva). A fenti vokóder az SSI témakörében tehát úgy használható, hogy a beszéd visszaállításához a zaj gerjesztés mellett nem az eredeti spektrális paramétereket használjuk fel, hanem az arcképek alapján gépi tanulással becsülteket.

2.2. Gépi tanulás

A modellek feladata, hogy a bemenetükön kapott videókból vagy képkockákból előállítsák a beszédszintetizáláshoz szükséges együtthatókat. Az MGC-paraméterek a beszéd spektrális burkolóját írják le, a neuronháló feladata ezeknek a paramétereknek a minél pontosabb becslése volt az arckép / ajak alapján. Mivel ezek a paraméterek folytonos értékűek, ezért regressziós módban használtuk a mély hálókat. Tekintve, hogy az MGC paraméterek különböző skálán mozogtak, tanítás előtt standardizáltuk őket, hogy várható értékük 0, szórásuk pedig 1 legyen. A standardizálás egy fontos lépés, hiszen amennyiben ezt nem tesszük meg, úgy a regressziós tanulás során a nagyobb értékekkel rendelkező MGC jellemzőt tanulja meg a háló nagy pontossággal, míg a kisebb értéktartományon mozgót kevésbé az MSE hibafüggvény miatt. A modellek az átlagos négyzetes hiba (Mean Squared Error, MSE) hibafüggvényt használták. Annak érdekében, hogy elkerüljük a túltanulást, fontos, hogy a megfelelő időben állítsuk meg a tanítási folyamatot. Ezt 'early stopping' módszerrel oldottuk meg: ha a validációs hiba nem javul 5 epochon keresztül, akkor a tanítás leáll és

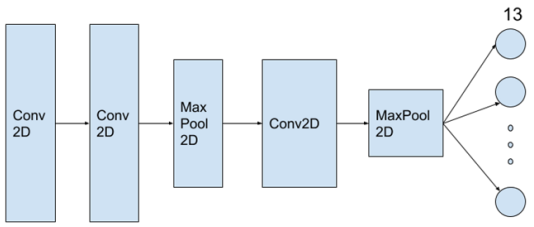
az utolsó legjobb eredményt menti el a hálózat. Automatikus hiperparaméter optimalizálást nem végeztünk.

2.2.1. Előrecsatolt mély neurális hálózat (FC-DNN alaprendszer)

Alaprendszernek egy 5 rejtett réteges, rétegenként 1000 neuront tartalmazó neuronháló struktúrát használtunk lineáris kimeneti réteggel. Az alaprendszer bemeneteként nem az arcvideó pixeleit használtuk, hanem az ezekből főkomponens-analízissel 100 dimenziósra tömörített adatokat, melyhez az EigenFaces módszert alkalmaztuk (Hueber et al., 2007). Enélkül a neuronháló nem tudta megtanulni a bemenet és kimenet közti összefüggést.

2.2.2. Konvolúciós neurális hálózat (CNN)

A konvolúciós neurális hálózatot (Convolutional Neural Network, CNN) gyakran alkalmazzák a képek osztályozására, feldolgozására. A tanulás folyamán képes megtanulni a képek különböző tulajdonságait, mint például a különböző élek, görbék kinézetét. Több egymás utáni konvolúciós rétegből és a hozzájuk tartozó aktivációs függvényekből áll. A hálózat egy képet kap a bemenetén, amelyen egy $n \times n$ méretű szűrővel csúszóablakszerűen végig haladva tömöríti a pixelekből kinyert információt. Így tanulja meg a hálózat a kis $n \times n$-es képekből az eredeti kép különböző tulajdonságait.



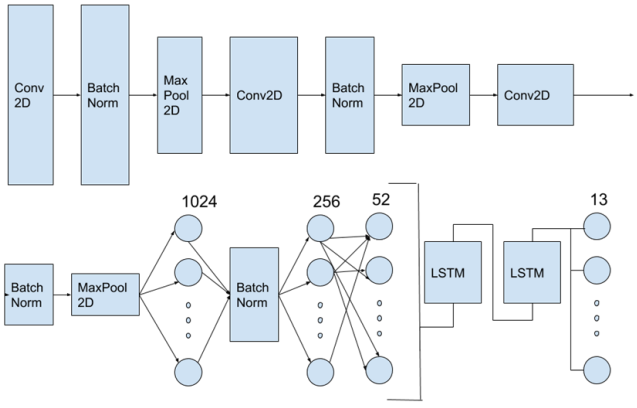
2. ábra. A CNN modell architektúrája.

Az ajakvideó-beszéd átalakításhoz második modellként egy CNN hálózatot használtunk. A hálózat a videókat képkockánként kapta meg. Így a bemeneti adatok nem tartalmazták azt az információt, hogy ezek a képek időben összefüggenek. A 2. ábrán látható a modell felépítése. Ez egy egyszerű hálózat, két

konvolúciós réteg után egy maxpool réteg, majd egy konvolúciós és egy maxpool réteg következik, végezetül pedig tizenhárom neuron adja a kimenetet, mivel a beszédszintetizáláshoz szükséges tizenhárom együtthatót kell meghatároznia a modellnek. Optimizációs algoritmusnak Adam-ot használtunk. Aktivációs függvénynek LeakyReLU-t (Leaky Rectified Linear Unit) alkalmaztunk, amelyet gyakran használnak a konvolúciós hálóknál.

2.2.3. Konvolúciós és rekurrens neurális hálózat (CNN-LSTM)

Bizonyos adatok szekvenciális formában állnak rendelkezésre, azaz az adatsorban az egymás utániság is hasznos információval bír. Például ha egy részvény jövőbeli értékét szeretnénk megjósolni, ahhoz nem elég, ha csak az utolsó nap értékét látjuk. Egy értékből nem tudjuk megállapítani, hogy a részvény ára javuló vagy romló tendenciát mutat vagy éppen stagnál. Az ilyen és ehhez hasonló esetekben hasznos lehet egy olyan mély neurális hálózat, amely rendelkezik valamilyen memóriával, amiben képes tárolni az előzőleg kapott adatokat és ezek alapján meg tudja tanulni az összefüggéseket. Az előrecsatolt hálózatokkal ellentétben a rekurrens neurális hálózatok (Recurrent Neural Networks, RNN) rendelkeznek ilyen memóriával. A Long Short-Term Memory (LSTM) a rekurrens hálózatok egy változata, amely hatékonyabban tanítható, mint a hagyományos RNN.



3. ábra. A CNN-LSTM modell architektúrája.

A harmadik modellünk a konvolúciós hálózat végére csatolt LSTM-mel egészült ki. Ettől a modelltől vártuk a legjobb eredményt, mivel a konvolúciós hálózat képes megtanulni a videó különböző jellegzetességeit, tulajdonságait, az utána kötött LSTM viszont az időbeli összefüggéseket. A hálózat egyszerre több 128×128-as képkockát is megkapott a bemenetén, ez lehetővé tette az időbeliség megtanulását. A bemeneti képek optimális darabszámát egyrészt kísérletezéssel, másrészt a korábbi hasonló kutatások tapasztalatai alapján állítottuk be (Ephrat & Peleg, 2017; Akbari et al., 2018). A vektorokból nyolc képet összefűztünk, majd ezekhez a képkockacsoportokhoz a csoport első képéhez tartozó kimenetet rendeltük. A második modell konvolúciós hálózatát kiegészítettük néhány fully-connected réteggel és két Long Short-Term Memory réteggel is. A 3. ábrán látható a modell felépítése. A fully-connected rétegekre azért volt szükség, mert a hálózat enélkül nem volt képes tanulni a rétegek közti nagy paraméterszám különbség miatt. A hálózat működését kipróbáltuk az Adam és az SGD optimizációs algoritmusokkal is és végül az SGD-t használtuk, mivel így a tanítások jobb eredményeket mutattak.

2.3. Kiértékelési módszerek

A tesztelés folyamán mindhárom modellnek el kellett végeznie egy predikciót egy-egy olyan videón, amit sem a tanító és sem a validációs halmazban nem látott még. A tesztek előkészítéseként a kódoló függvénnyel elvégeztük ugyanazt a kódolást a bemeneten, mint a tanító adatok esetén, majd a megfelelő bemeneti struktúrába rendeztük őket. Itt a hálózat már nem kapja meg az elvárt kimenetet, mivel neki kell egy predikciót készítenie.

A következő lépésekben úgy ellenőriztük le a modellek helyességét, hogy a tanítás során keletkezett tanító és validációs hibákat kísértük figyelemmel, majd a kimenetként kapott együtthatókból hangot generáltunk és meghallgattuk azokat. Hasonlóképpen vetettük össze a modelleket egymással is: megkerestük, hogy melyik hálózatnál volt a legkisebb a validációs hiba értéke, és hogy melyik hálózat hány epochig volt képes tanulni, mielőtt az 'early stopping' leállítot-

ta volna. A generált audió fájlokat meghallgatva összevetettük, hogy melyik modell érte el a legjobb eredményt.

A tesztelés előtt össze kellett állítani egy teszthalmazt, amely segítségével össze tudjuk hasonlítani a modelleket. Kiválasztottunk tíz olyan videót, amelyet nem használtunk fel előtte sem a tanításhoz, sem a validációhoz, így a hálózatok számára teljesen ismeretlenek voltak. Ezeket a videókat megkapták a hálózatok és a kimenetként kapott együtthatókból hangot generáltunk. Három modellt használtunk fel a tesztelés során: az előrecsatolt hálózatot (EigenFaces bemenettel), a konvolúciósat (amely képkockákból tanult), és a Long Short-Term Memory rétegeket használót (amelynek nyolc egymás utáni képkocka volt a bemenete). A tesztelést kétfelé bontottuk: objektív és szubjektív tesztelésre.

2.3.1. Objektív kiértékelés

Az objektív kiértékelés során olyan mérőszámot kerestünk, amellyel pontosabban meg lehet határozni a modellek egymáshoz viszonyított eredményességét, mint a validációs hiba értékével. Az objektív teszt során két-két hangfájl közötti spektrális távolságot (Mel Cepstral Distortion, MCD) számoltunk, Kubichek (1993) alapján. Az MCD-t a beszédszintetizáló rendszerek minőségének felmérésére használják. Minél kisebb az MCD értéke a szintetizált és a természetes beszéd között, annál jobban sikerült a szintetizált beszédnek reprodukálnia a természetes beszédet.

2.3.2. Szubjektív meghallgatásos teszt

A szubjektív tesztelés során az erre a célra készített internetes teszt (`http://leszped.tmit.bme.hu/rb2019/`) kitöltői is meghallgatták a generált hangokat. Az ajakvideó-beszédszintézis kutatásának célja, hogy a jövőben az emberek könnyedén használhassák a vele alkotott szolgáltatásokat, termékeket. Így a tesztelés folyamán a felhasználói élmény felmérése kiemelten fontos, hisz lehet bármilyen hasznos egy alkalmazás, ha a felhasználóknak kényelmetlen, nehézséget okoz a használata, akkor nem fogják használni. Éppen ezért egy szubjektív hallgatásos tesztet végeztettünk el, hogy felmérjük melyik modell hogyan tel-

jesít. A teszt során a teszthalmazban szereplő videókból szintetizált hangokat kellett meghallgatniuk a kitöltőknek. Egy-egy oldalon mindegyik hangminta ugyanahhoz a videóhoz tartozott; MUSHRA-jellegű tesztként (ITU-R Recommendation BS.1534). Referenciaként szerepelt az eredeti hangfájl is. A kitöltők feladata az volt, hogy minden egyes mintát osztályozzanak egy skálán aszerint, hogy mennyire hangzik természetesnek az adott beszéd (0: teljesen természetellenes, 100: teljesen természetes). A tesztben a fent említett három hálózat által szintetizált hangok szerepeltek (az eredeti mondatok F0-ját megtartva a szintézis során), valamint az eredeti beszéd és a vokóderrel szintetizált is; mindegyik rendszerből 10–10 mondat. A kitöltők nem tudták, melyik minta melyik modellhez tartozott és a sorrendjük tesztesetenként meg is volt keverve. A szubjektív teszt egy másik felületén figyelemmel követhettük az eredményeket. A tesztet összesen hét kísérleti alany végezte el (hat férfi és egy nő; 23–45 évesek, átlagos életkor: 31 év; egyikük sem volt beszédtechnológiai szakértő). Itt láthattuk, hogy melyik modell átlagosan milyen értékelést kapott az egyes tesztesetekben vagy a teszt egészén.

3. Eredmények és diszkusszió

3.1. Objektív kiértékelés

Összehasonlítási alapnak nem a természetes beszédet választottuk, mivel a tanítóminták előkészítése során az audió fájlból egy kódoló segítségével nyerjük ki az együtthatókat, amikből majd egy dekódolóval újra hangot generálunk. Ez a vokóder torzítja az eredeti hangot, így a tanítás során is jelen van ez a torzítás. Ezért a referenciának a teszthalmazban szereplő eredeti hangfájlokból vokóderrel szintetizált beszédet választottuk. Ezt hasonlítottuk össze a már említett három modellel. Modellenként mind a tíz tesztmintára lefuttattuk az MCD számítást. Az 1. táblázatban láthatóak a teszt eredményei. Az MCD-t használva annál jobb eredményről beszélhetünk, minél kisebb a kapott érték. Az egész teszt legjobb értékét a CNN-LSTM érte el, átlagosan 4,05-öt. Ezután következett az alaprendszer (MCD: 5,20), majd végül a konvolúciós hálózat

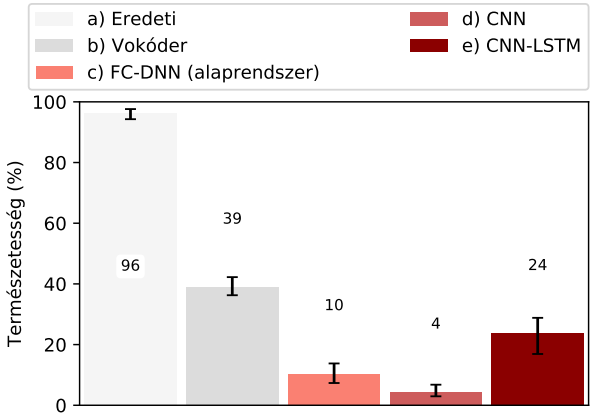
1. táblázat. A különböző modellek MCD értékei az egész tesztet nézve

mondat	Előrecsatolt neurális hálózat (FC-DNN, alaprendszer)	Konvolúciós neurális hálózat (CNN)	Konvolúciós és rekurrens neurális hálózat (CNN-LSTM)
1	5,01	6,15	3,74
2	5,45	6,31	3,88
3	5,31	6,29	4,28
4	5,51	7,05	3,81
5	4,72	5,15	3,61
6	4,27	6,08	3,73
7	5,57	6,47	4,12
8	5,94	6,54	3,70
9	5,23	6,80	4,84
10	5,26	5,63	4,81
átlag	5,20	6,25	4,05

(MCD: 6,25). A konvolúciós hálózat valószínűleg azért teljesített rosszul, mert a bemeneti képek túl nagy mennyiségű információt tartalmaznak, amiből nem tudta hatékonyan megtalálni a spektrális paraméterekkel való összefüggést. Az alaprendszer EigenFaces dimenziócsökkentő eljárást használt, így ott ez nem fordult elő. A CNN-LSTM pedig azért eredményezhetett kisebb hibát, mert ott a nyolc egymás utáni kép összefűzéséből származó információ kompenzálni tudta a konvolúciós rétegeket.

3.2. Szubjektív meghallgatásos teszt

A 4. ábra megmutatja, hogy a szubjektív meghallgatásos tesztben átlagosan milyen értékelést kaptak a modellek az egész tesztre vetítve (a 95%-os konfidenciaintervallumokat is feltüntetve). Az 'eredeti' címkéjű hangot használtuk referenciának. A teszt során a maximális 100 ponthoz nagyon közeli értékeket ért el, tehát a kitöltők is azt a mintát tartották a legtermészetesebbnek, ami a valóságban is az. A tesztből kiderült, hogy a vokóder jelentősen rontja a beszéd minőségét, a kitöltőktől átlagosan 39 pontot kapott a 'vokóder' típus.



4. ábra. A különböző modellek átlagos szubjektív értékei az egész tesztet nézve (a 95%-os konfidenciaintervallumokat is feltüntetve).

Mivel a tanítás során ennek a segítségével szintetizáljuk a beszédet, így várható volt, hogy egyik modell sem fog magasabb pontszámot elérni. Az általunk alkotott hálózatok közül, ebben a tesztben is a CNN-LSTM modell érte el a legjobb eredményeket. A teljes tesztre kiszámolt átlagból látszik, hogy a hálózatnak van még hova fejlődnie, a vokódertől nagyjából 15 ponttal van lemaradva. A legrosszabbul a csak konvolúciót használó hálózat teljesített. Ez a modell nagyon kicsi, 5-höz közeli pontszámot kapott.

Összegezve megállapíthatjuk, hogy a CNN-LSTM hálózat sokkal tisztább, kevésbé zajos hangokat képes szintetizálni, mint a képkockasorokat használó konvolúciós hálózat.

A szakirodalmi áttekintés során találtunk néhány hasonló kutatást, melyek a 'lip-to-speech' témakörrel foglalkoztak. Le Cornu & Milner (2015) az arcról készült képek előfeldolgozásával próbált jobb eredményeket elérni, míg mi ezt a feldolgozást a neurális hálózatokra bíztuk. Ephrat & Peleg (2017) csak konvolúciós hálózatot használt, míg mi rekurrens módszereket is teszteltünk. Akbari et al. (2018) egy komplex beszédkódolót használt a spektrális paraméterekből beszéd szintéziséhez; a saját fenti kísérletekben pedig egy egyszerű vokódert alkalmaztunk.

4. Következtetések

A kutatásban az ajakvideó alapú beszédszintézisre mutattunk be egy kísérletet. Konvolúciós és rekurrens neurális hálózat architektúrákat teszteltünk. Megtapasztaltuk, hogy milyen fontos az adatok megfelelő ismerete és előfeldolgozása, hogy milyen problémák adódhatnak. A hálózatok teljesítményét több lépcsős teszteléssel össze is hasonlítottuk. Először egy objektív teszten összehasonlítottuk a Mel Cepstral Distortion érték szerint a különböző modelleket. Majd egy szubjektív hallgatásos tesztet töltettünk ki néhány emberrel, ahol a hallott mintákat kellett értékelniük aszerint, hogy mennyire érzik természetesnek őket. A tesztekből egyértelműen kiderült, hogy a CNN-LSTM hálózat érte el a legjobb eredményt. Bár a modelljeink folyamatosan fejlődtek, még a CNN-LSTM hálózat által szintetizált beszéd sem érthető teljesen, de szótöredékek felismerhetőek. Az eredmények alkalmazhatóak lehetnek némabeszéd-interfészekben vagy automatikus szájról olvasó rendszer kidolgozásához (Sun et al., 2018).

Egy továbbfejlesztési lehetőség egy teljesen más architektúra, mint például a Generative Adversarial Network (GAN) (Goodfellow et al., 2014) típusú hálózat használata lehetne. CNN hálózat esetében lehetséges 8 keretet felhasználni a bemenethez, akár 2D konvolució esetén 8 csatornával, vagy 3D konvolució felhasználásával (Tóth & Shandiz, 2020). További lehetőség a rekurrens hálózat seq2seq módon történő tanítása (encoder-decoder architektúrával), amely az eredményeket javíthatja, mivel a hosszú távú információ is a hálózat rendelkezésére áll (Sutskever et al., 2014).

Köszönetnyilvánítás

A kutatást részben a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal támogatta (FK 124584 és PD 127915 projektek). Köszönjük a meghallgatásos teszt résztvevőinek a teszt kitöltését.

Hivatkozások

Akbari, H., Arora, H., Cao, L., & Mesgarani, N. (2018). LIP2AUDSPEC : Speech reconstruction from silent lip movements video. In *Proc. ICASSP* (pp. 2516–2520). Calgary, Canada.

Bolla, K. (1995). *Magyar fonetikai atlasz. A szegmentális hangszerkezet elemei*. Budapest: Nemzeti Tankönyvkiadó.

Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, *120*, 2421–2424. URL: http://asa.scitation.org/doi/10.1121/1.2229005. doi:10.1121/1.2229005.

Csapó, T. G., Deme, A., Gráczi, T. E., Markó, A., & Varjasi, G. (2017a). Synchronized speech, tongue ultrasound and lip movement video recordings with the "Micro" system. In *Challenges in analysis and processing of spontaneous speech*.

Csapó, T. G., Grósz, T., Gosztolya, G., Tóth, L., & Markó, A. (2017b). DNN-Based Ultrasound-to-Speech Conversion for a Silent Speech Interface. In *Proc. Interspeech* (pp. 3672–3676). Stockholm, Sweden. URL: http://dx.doi.org/10.21437/Interspeech.2017-939. doi:10.21437/Interspeech.2017-939.

Csapó, T. G., Grósz, T., Tóth, L., & Markó, A. (2017c). Beszédszintézis ultrahangos artikulációs felvételekből mély neuronhálók segítségével. In *MSZNY 2017* (pp. 181–192).

Csapó, T. G., & Németh, G. (2014). Statistical parametric speech synthesis with a novel codebook-based excitation model. *Intelligent Decision Technologies*, *8*, 289–299.

Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., & Brumberg, J. S. (2010). Silent speech interfaces. *Speech Communication*, *52*, 270–287. URL: http://dx.doi.org/10.1016/j.specom.2009.08.002. doi:10.1016/j.specom.2009.08.002.

Drugman, T., Wilfart, G., Moinet, A., & Dutoit, T. (2009). Using a Pitch-Synchronous Residual Codebook for Hybrid HMM/frame Selection Speech Synthesis. In *Proc. ICASSP* (pp. 3793 – 3796). Taipei, Taiwan.

Ephrat, A., & Peleg, S. (2017). Vid2speech: Speech Reconstruction from Silent Video. In *Proc. ICASSP* (pp. 5095–5099). New Orleans, LA, USA. URL: http://arxiv.org/abs/1701.00495. arXiv:1701.00495.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)* (pp. 2672–2680). URL: https://papers.nips.cc/paper/5423-generative-adversarial-nets.

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., & Stone, M. (2007). Eigentongue feature extraction for an ultrasound-based silent speech interface. In *Proc. ICASSP* (pp. 1245–1248). Honolulu, HI, USA.

Imai, S., Sumita, K., & Furuichi, C. (1983). Mel Log Spectrum Approximation (MLSA) filter for speech synthesis. *Electronics and Communications in Japan (Part I: Communications)*, *66*, 10–18. URL: http://doi.wiley.com/10.1002/ecja.4400660203. doi:10.1002/ecja.4400660203.

Kempelen, F. (1791). *Az emberi beszéd mechanizmusa, valamint a szerző beszélőgépének leírása [Eredeti cím: Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine]*. Bécs, Ausztria: Degen.

Kimura, N., Kono, M. C., & Rekimoto, J. (2019). Sottovoce: An ultrasound imaging-based silent speech interaction using deep neural networks. In *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–11). Glasgow, UK. doi:10.1145/3290605.3300376.

Kubichek, R. F. (1993). Mel-cepstral distance measure for objective speech quality assessment. In *Proc. ICASSP* (pp. 125–128). Victoria, Canada. doi:10.1109/pacrim.1993.407206.

Le Cornu, T., & Milner, B. (2015). Reconstructing intelligible audio speech from visual speech features. In *Proc. Interspeech* (pp. 3355–3359). Dresden, Germany.

Rácz, B. (2019). *VID2SPEECH: beszédszintézis ajak videóból konvolúciós és rekurrens mély neurális hálózatokkal*. Technical Report BME TMIT. URL: https://diplomaterv.vik.bme.hu/hu/Theses/VID2SPEECH-beszedszintezis-ajak-videobol.

Sun, K., Yu, C., Shi, W., Liu, L., & Shi, Y. (2018). Lip-Interact: Improving Mobile Device Interaction with Silent Speech Commands. In *UIST 2018 - Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (pp. 581–593). Berlin, Germany. URL: http://dl.acm.org/citation.cfm?doid=3242587.3242599. doi:10.1145/3242587.3242599.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems* (pp. 3104–3112). arXiv:1409.3215.

Tokuda, K., Kobayashi, T., Masuko, T., & Imai, S. (1994). Mel-generalized cepstral analysis - a unified approach to speech spectral estimation. In *Proc. ICSLP* (pp. 1043–1046). Yokohama, Japan.

Tóth, L., & Shandiz, A. H. (2020). 3D Convolutional Neural Networks for Ultrasound-Based Silent Speech Interfaces. In *Proc. ICAISC*. Zakopane, Poland.

Wand, M., Koutník, J., & Schmidhuber, J. (2016). Lipreading with long short-term memory. In *Proc. ICASSP* (pp. 6115–6119). Shanghai, China.

Explozívák realizációja fiatal és öregedő felnőttek beszédében

Száraz Bettina[1], Gráczi Tekla Etelka[1,2]

[1] *Nyelvtudományi Intézet*
[2] *MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport*

Abstract

Possible gender and age related differences have been hypothesized and suggested by previous work on voice onset time of stops. Most studies investigate not truly voicing languages, but those that were carried out on truly voicing languages, like Hungarian, explore only voiceless stops. However, the simultaneous production of voicing and closure are contradictory due to the intraoral pressure build up. Therefore the question arises if gender or age related articulatory differences/changes may result in different realisations. The present study aimed to introduce preliminary results on possible gender and age related realisations of stops in Hungarian that contrast prevoicing and short lag VOT. 10 intervocalic, word-medial realizations of /d t g k/ per subject were analyzed in read speech of 30 young and 30 aging subjects (15 men and 15 women in both groups) from the BEA-database. The Scheirer–Ray–Hare-test was used to compare the results between age and gender groups. The VOT of voiced stops was longer in aging subjects' pronunciation than that in the younger ones, and it was longer in female subjects' speech than in men's read speech. No clear tendencies were found for the voiceless stops. The approximalised/fricated or no-burst realisations of the velar stops were more frequent in men's pronunciation than in that of women. The voicing ratio did not show clear tendencies. The longer VOT of voiced stops in the aging group and in female subjects' speech may be the result of longer closure phases that might be resulted by slower articulation. This assumption needs further investigation. Results of further parameters did not show consequent differences between the analysed speaker groups that, together with previous results for Hungarian, suggest that aging and gender have low or no direct impact on these features. The results of the present study serve as preliminary data on the gender and age related differences in stops of prevoicing languages, therefore, further studies are necessary.

1. Bevezetés

1.1. Szakirodalmi háttér

Beszédképzés során az egyes artikulációs gesztusok egymásra hatva működnek. A szupraglottális artikuláció nemcsak a hangszalagok felől érkező légáramot módosítja, hanem a hangszalagok feletti beszédszervek működésének aerodinamikai jellemzői visszahatnak a hangszalagok működésére is.

Email addresses: szaraz.bettina@nytud.hu (Száraz Bettina), graczi.tekla.etelka@nytud.hu (Gráczi Tekla Etelka)

Beszédtudomány – Speech Science 2020. 73–98. DOI: 10.15775/Besztud.2020.73-98

A magyarban, ahogyan sok nyelvben, az obstruensek csoportjában kettős oppozíció áll fenn (Ladefoged, 2005). A magyar nyelvben az explozívák esetében is a zöngésség tekintetében elsődleges akusztikai kulcsnak – olyan akusztikai információ, amely a fonológiai kategóriák valamely megkülönböztető jegyét az észlelés számára kódolja – a mássalhangzó zárszakaszának időtartama alatt a hangszalagrezgés meglétét vagy hiányát tekinthetjük (Lisker, 1986). Modális fonáció esetén a zöngés hangok képzésekor a hangszalagok zöngeállásban állnak. A zöngeállás esetében a kannaporcok érintkeznek, a hangszalagok pedig zárat alkotnak, ennek következtében a tüdőből kiáramló levegő felgyülemlik alattuk. Amikor a felgyülemlő levegő nyomása elég nagy, a levegőnyomás felnyitja a hangszalagokat. A hátulsó terület nyílik először, a nyitódás folyamatosan előrefelé terjed (Riper & Irwin, 1961). Ez csak akkor következhet be, ha a szupraglottális nyomás alacsonyabb, mint a hangszalagok felnyitásához szükséges nyomás. A felgyülemlett levegő kiáramlásával a nyomás csökken, így a hangszalagok újra összezárulnak. A záródást a csökkenő nyomás, azaz a Bernoulli-hatás idézi elő. A zöngeképzéskor ez a folyamat azonos időközönként ismétlődik. Az obstruensek képzésekor a szájüregben akadályt képzünk, amely mögött felgyülemlik a nyomás, így a hangszalagok alatti és feletti nyomás különbsége a zöngeképzéshez szükséges érték alá csökkenhet, ezért a zöngésség és a mássalhangzó képzési módjának együttes fenntartása nehézséget okozhat (Stevens, 1998). Ezt nevezzük a zönge- és az obstruensképzés „ellentmondásának". Az ellentmondás következtében a hangszalagok rezgése megnehezedik, a nyitott szakasz időtartama növekszik, valamint bekövetkezhet a zöngeképzés leállása (Bickley & Stevens, 1986).

A hangszalagok beállítása és a nyomásviszonyok miatt a zöngét könnyebb fenntartani intervokális helyzetben (főként két azonos magánhangzó-minőség között), mint szünet utáni vagy előtti helyzetben (Westbury & Keating, 1986). A mássalhangzó hátsóbb képzés helye esetében a kisebb térfogat nehezíti a zöngeképzés fenntartását (Shadle, 1997; Stevens, 1998).

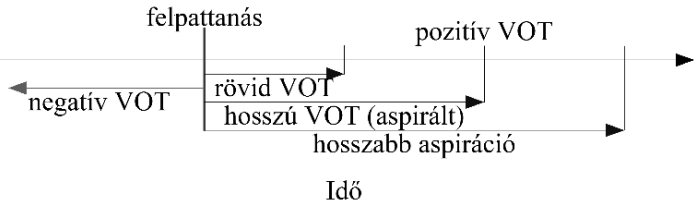
A zöngésség és az akadály egyidejű fenntartásának aerodinamikai nehézsége ahhoz vezet, hogy minél hosszabb a zöngés hang, annál nagyobb arányban zön-

gétlenedhet, azaz a hosszabb konszonánsok gyakrabban zöngétlenednek (Jesus & Shadle, 2003; Gráczi, 2012).

Általánosan elfogadott, hogy az akusztikai jelben több paraméter kulcsolja ugyanazt a jelenséget, a hallgató ezeket a kulcsokat nem egyenlően érzékeli, bizonyos paramétereknek nagyobb súlyt tulajdonít, mint másoknak (Ohde & Haley, 1997). A zöngésséget magyar nyelven elsősorban, ahogyan korábban már említettük, a hangszalagrezgés megléte vagy hiánya kulcsolja (Lisker, 1986; Crowther & Mann, 1992), ugyanakkor vele járó egyéb akusztikai különbségek is szolgálhatnak kulcsként. Ilyen például a mássalhangzók időtartama (pl.: Lisker, 1957, 1986; Pisoni & Remez, 2005; Gráczi, 2012; Száraz, 2019), a mássalhangzót megelőző magánhangzók időtartama (pl.: Chen, 1970; Krause, 1982; Lisker, 1986; Crowther & Mann, 1992; Kovács, 2000; Pisoni & Remez, 2005; Gráczi, 2012; Száraz, 2019), ezek egymáshoz viszonyított aránya (pl.: Denes, 1955; Cole & Cooper, 1975; Port & Dalby, 1982; G. Kiss & Bárkányi, 2018), a megelőző magánhangzó offsetjének első formánsának értéke (pl.: Lisker, 1986; Summers, 1988; Crowther & Mann, 1992; Tuomainen & Lely, 2007) valamint a zöngekezdési idő (= voice onset time, VOT) (pl.: Lisker, 1986; Pisoni & Remez, 2005).

A zöngekezdési idő az az időtartam, amely a zár feloldásától a zönge megindulásáig telik el (Lieberman & Blumstein, 1988). A zöngekezdési időtartam nyelvenként változik (Lisker & Abramson, 1964). Egyes nyelvekben, köztük az angolban a zárszakasz alatt nem várt zönge, az a zár felnyílása után rövid idővel vagy hosszabb idővel indul meg. Azaz a zöngekezdési idő pozitív a zönge szempontjából oppozíciót alkotó párok mindkét tagjában, így a rövidebb és a hosszabb érték, és a hosszabb értékkel járó hehezetesség/aspiráció áll szemben egymással. A zárszakasz alatt megjelenő zöngét előzöngének is szokás hívni, ez – mivel a zár feloldásától számítjuk a zöngekezdési időt – negatív VOT-t jelent. A magyar nyelvben tehát előzönge/negatív VOT és rövid pozitív VOT állnak szemben egymással (1. ábra). Vannak olyan nyelvek, például a thai, ahol háromféle kategóriát különböztetnek meg, az előzöngés, a rövid zöngekezdési idővel és a hosszú zöngekezdési idővel jellemzett explozívákat (Gandour & Dardarananda, 1984). Más nyelvekben akár négyes (például a hindi) vagy többes

szembenállás is fennállhat. Koreai nyelvben három zöngétlen kategóriát különböztetnek meg, a nem aspirált, egy rövidebb és egy hosszabb hehezettel képzett explozívát (Lisker & Abramson, 1964).



1. ábra. A zöngekezdési idő (Gráczi, 2016, 64.)

A VOT egy nyelven belül is variábilis lehet, ugyanis több tényező hatással van az értékére. Henton és munkatársai (1992) azt mutatták ki, hogy zöngés explozíva esetében minél előrébb van a felpattanó explozíva képzési helye a szájüregben, annál hosszabb zöngekezdési idővel realizálódik, a zöngétlen esetében pedig minél előrébb van a képzési hely a szájüregben, annál rövidebb zöngekezdési idővel jelenik meg.

Magyar nyelven először Gósy (2000) vizsgálta a zöngétlen explozívák zöngekezdési idejét. Eredményei azt mutatták, hogy a spontán beszédben a zöngétlen felpattanók VOT-je annál rövidebb, minél előrébb található a képzési hely a szájüregben – tehát eredményei alátámasztották Henton és munkatársai (1992) feltételezéseit –, azonban a felolvasás esetében ez a tendencia nem volt kimutatható. Gósy (2000) zöngétlen felpattanók vizsgálatában kimutatta, hogy a követő magánhangzó tulajdonságai is hatással vannak a zöngekezdési idő hosszára. Gósy (2000) a bemutatott paramétereken kívül vizsgálta még a beszédtípus befolyását a zöngétlen explozívák zöngekezdési időre. Eredményei azt mutatták, hogy a spontán beszéd esetében a VOT szórása nagyobb, valamint a követő magánhangzó hatása kisebb a megjelenő zöngekezdési időre, mint a szólista-felolvasás esetében, azonban a főbb tendenciák azonosak.

Gósy és Ringen (2009) a /b p d t g k/ mássalhangzók VOT-értékét elemezte. A vizsgálatban olvasott szólista alapján szókezdő, intervokális és szóvégi helyzetben elemezték a zöngésségi párokat. Az eredmények alapján szókezdő

és intervokális helyzetben a zöngétlen explozívák esetében minél hátrébb volt a szájüregben az akadály, annál hosszabb zöngekezdési idővel realizálódtak a mássalhangzók. Zöngés explozívák esetében szókezdő pozícióban nem volt kimutatható tendencia, intervokális és szóvégi helyzetben azonban minél előbb volt a szájüregben az akadály, annál hosszabb zöngekezdési idővel realizálódtak az explozívák. Gósy és Ringen (2009) a zöngétlenedést is vizsgálta. Intervokális helyzetben a bilabiális explozíva zöngétlenedett legkevésbé (zöngésrész-arány: 98,1%), a veláris explozíva pedig a legjobban (zöngésrész-arány: 89,4%). Szóvégi pozícióban mindhárom képzéshelyű zöngés explozíva nagyobb mértékben zöngétlenedett. Legkevésbé ebben a pozícióban is a bilabiális explozíva (zöngésrész-arány: 73,6%), legjobban pedig a veláris felpattanó (zöngésrész-arány: 69,5%).

Gráczi és munkatársai (2009) spontán beszédben elemezték a bilabiális, alveoláris és veláris explozívákat intervokális helyzetben. Az ő eredményeik is azt mutatták, hogy a zöngés rész aránya a képzési hellyel hátrafelé csökkent, tehát a legkevésbé a bilabiális, a legjobban pedig a veláris felpattanó zöngétlenedett.

Bóna (2011) vizsgálta a beszédtípus befolyását a zöngétlen felpattanók zöngekezdési időre. Kutatásában spontán narratívát és szövegfelolvasást vetett össze. Eredményei azt mutatták, hogy a két beszédtípus között nem volt kimutatható eltérés.

Gráczi (2011) a /b p d t ɟ c g k/ hangokat elemezte a képzésmódváltás, illetve a zöngétlenedés szempontjából. Amennyiben az explozívában képzésmódváltás következik be, a zár és annak feloldásának hiánya miatt nem lehet zöngekezdési időt mérni az adott beszédhangon. A mássalhangzók megvalósulásai azt mutatták, hogy 93,6%-ban felpattanó zárhangként valósultak meg. Legnagyobb arányban a zöngés palatális és a veláris explozívák jelentek meg eltérő képzéssel: /g/ 8,3%-ban valósult meg közelítőhangként. Elenyésző arányban a bilabiális és az alveoláris mássalhangzók is realizálódtak approximánsként (mindkettő esetében 2,1%-ban). A zöngétlenedés tekintetében elmondható, hogy a szájüregben hátrafelé haladva nagyobb a mértéke, azaz a legkevésbé a bilabiális zöngés mássalhangzó zöngétlenedett (zöngésrész-arány: 97,9%), legjobban pedig a veláris mássalhangzó (zöngésrész-arány: 82,3%). A zöngétlenek esetében

mind a négy vizsgált képzési helyen történt zöngésedés, azonban eltérő arányú. A legnagyobb mértékben a bilabiális zöngétlen mássalhangzó zöngésedett (zöngétlenrész-arány: 79,6%), a legkisebb mértékben pedig a veláris explozíva (zöngétlenrész-arány: 93,4%). Ezek alapján tehát az volt elmondható, hogy a szájüregben hátrafelé haladva kisebb mértékben zöngésedtek a zöngétlen explozívák.

Gráczi (2012) intervokális helyzetben vizsgálta a bilabiális, alveoláris, palatális és veláris zöngés és zöngétlen explozívákat. A zöngés explozívák esetében a szájüregben hátrafelé haladva nagyobb mértékben zöngétlenedtek a mássalhangzók (bilabiálistól a veláris képzéshelyig haladva a zöngésrész-arány: $98,7 \pm 4,4\%$, $87,8 \pm 18,0\%$, $81,2 \pm 23,9\%$, $78,0 \pm 17,5\%$). A zöngétlen explozívák esetében azonban fordított tendencia volt kimutatható, a szájüregben hátrafelé haladva kisebb mértékben zöngésedtek a mássalhangzók (bilabiálistól a veláris képzéshelyig haladva a zöngétlenrész-arány: $71,1 \pm 22,0\%$, $71,4 \pm 15,8\%$, $77,1 \pm 19,2\%$, $85,4 \pm 15,0\%$).

A nem hatását is vizsgálták a zöngekezdési időre. Swartz (1992) amerikai angol nyelvre végzett kutatása alapján a férfiak rövidebb zöngekezdési időt produkáltak, mint a nők. Mindezek mellett egyéni beszélői sajátosságok és további hatások (szájüreg térfogata, artikulációs jellemzők stb., Alphen & Smits, 2004) is befolyásolhatják a realizálódó beszédhang szerkezetét.

Magyar nyelven Gósy és Ringen (2009) vizsgálták a nem hatását a zöngekezdési időre. Az eredményeik azt mutatták, hogy a zöngés explozívák esetében a nők beszédében jelentek meg hosszabb zöngekezdési idővel a felpattanók, míg a zöngétlenek esetében fordítva, azaz a férfiak beszédében realizálódtak hosszabb zöngekezdési idővel az explozívák.

Bóna (2016) kutatásában több életkori csoportban vizsgálta, hogy a nem milyen hatással van a megvalósuló zöngekezdési időre zöngétlen explozívákban. A fiatal felnőttek esetében a zöngétlen alveoláris explozíva a férfiak beszédében jelent meg hosszabb zöngekezdési idővel, míg a veláris explozíva a nők beszédében realizálódott hosszabb zöngekezdési idővel. Az idősek esetében a bilabiális

és az alveoláris képzéshelyű explozívák a nők beszédében realizálódtak hosszabb zöngekezdési idővel, míg a veláris képzéshelyű explozívák a férfiakéban.

Az előző eredmények is mutatják, hogy a beszélő életkora is hatással lehet a zöngeképzésre. Az öregedés következtében időskorban a tüdőkapacitás csökken, valamint a renyhébben dolgozó hangszalagzáró izomzat miatt a hangerő csökken, a hangtartás megrövidül (Levitzky, 1984; Huber, 2008). A hangszalagok rugalmatlanabbá válnak, a gégeizomzat leépül, a porcok meszesedése fokozottabb lesz, a hangképző izmok tónusa csökken, ami a hang gyengüléséhez, szaggatottságához vezet (Balázs, 1993). Petrosino és munkatársai (1993), valamint Ryalls és munkatársai (1997) a beszélő életkorának hatását vizsgálták a zöngekezdési időre. Eredményeik azt mutatták, hogy az idősek beszédében a zöngekezdési idő értékei szélesebb tartományban szóródnak, mint a fiatalokéi. Benjamin (1982) kutatása azt mutatta, hogy az angol nyelvben az idősek beszédében az explozívák rövidebb zöngekezdési idővel realizálódtak, mint a fiatalokéban. Más kutatások azonban nem találtak különbséget a két életkori csoport átlagos zöngekezdési idejében (Petrosino et al., 1993).

Bóna (2011) azt feltételezte, hogy az életkor előrehaladtával az artikulációs tempó és az artikuláció pontosságának változása különbséget okoz a fiatalok és az idősek zöngekezdési idejében. A BEA adatbázisból öt 70–80 év közötti és öt 22–32 év közötti, magyar anyanyelvű nő szövegfelolvasását és spontán narratíváját választotta ki az elemzéshez. A /p, t, k/ mássalhangzók zöngekezdési idejét #CV, VCV és CCV kapcsolatokban elemezte, a hangkörnyezet azonban nem volt kiegyenlített. A vizsgálat eredményei azt mutatták, hogy a bilabiális és az alveoláris zöngétlen felpattanók rövidebb zöngekezdési idővel jelentek meg a fiatalok beszédében, mint az idősekében. A veláris felpattanók esetében ennek a tendenciának a fordítottja történt. Az idősek beszédében rövidebb zöngekezdési idővel realizálódtak a veláris explozívák, mint a fiatalokéban. Továbbá elmondható, hogy mindhárom képzési helyen a zöngekezdési idő nagyobb tartományban szórt az idősek esetében.

A magyar explozívákra kapott eredményeket összefoglalva az alábbi összefüggések vonhatóak le az eddigi tanulmányokból. A hátsóbb képzési hely felé

haladva nagyobb arányú a zöngés fonémák részben zöngétlen realizációja, a VOT pedig hosszabb mind a zöngések, mind a zöngétlenek esetében (vö. Gósy & Ringen, 2009; Gráczi, 2012) – hasonlóan a más nyelvekben megfigyelt tendenciákhoz. A zöngekezdési idő szignifikánsan eltérőnek mutatkozik a női és a férfi adatközlők között, de az egyes képzési helyek mentén (vö. Gósy & Ringen, 2009), illetve az egyes életkori csoportokban (vö. Bóna, 2011) eltérő volt, hogy mely nemi csoport VOT-értékei hoszabbak. Az életkor mentén eddig még csak a zöngétlen felpattanókat elemezték, magyar nyelvben sem vizsgálták a zöngés explozívák életkor mentén mutatott esetleges eltéréseit Mindezek alapján felmerülnek az alábbi kérdések: (1) Mennyiben áll fenn valós nemek közötti eltérés a zöngekezdési időben? (2) Igazolható-e szisztematikus eltérés felnőtt életkori csoportok között? (3) Milyen eltérés várható zöngés explozívák realizációjában felnőtt életkori csoportok között zöngétlenedés és zöngekezdési idő tekintetében?

1.2. Kutatási kérdések, hipotézis

Jelen kutatásunkban arra keressük a választ, hogy a magyar alveoláris és veláris zöngés és zöngétlen explozívák zöngekezdési ideje, a zöngeleállás gyakorisága és a képzésmód váltása különbözik-e a beszélők életkora és neme szerint.

A nemzetközi és a magyar szakirodalom alapján a következő hipotéziseket állítottuk fel:

1. A zöngétlen explozívák esetében a zöngekezdési idő hosszabb az öregedők felolvasásában.
2. A zöngés explozívák esetében a zöngekezdési idő hosszabb a fiatal felnőttek felolvasásában.
3. A zöngés felpattanók esetében a nőknél hosszabb a zöngekezdési idő.
4. A zöngétlen felpattanók esetében a férfiaknál hosszabb a zöngekezdési idő.
5. Az öregedők beszédében kisebb a vizsgált explozívák zöngésrész-aránya, mint a fiatal felnőttekében.
6. Az öregedők esetében a felpattanók gyakrabban jellenek meg alternatív, azaz approximalizált vagy spirantikus képzésmóddal, mint a fiatal felnőttekében.

2. Kísérleti személyek, nyelvi anyag, módszer

A vizsgálathoz 60 magyar anyanyelvű beszélőt választottunk ki a BEA adatbázisból (vö. Gósy, 2008). Az adatközlőket a fiatal felnőttek és az öregedők csoportjára osztottuk. Mindkét korcsoportban 30–30 adatközlő volt, azonos arányban férfiak és nők. A fiatal felnőttek csoportját 23–35 (átlag: 29,7; szórás: 3,72) év közötti, az öregedők csoportját pedig 65–77 (átlag: 69,6; szórás: 3,46) év közötti kísérleti személyek alkották. A beszélők nem rendelkeztek beszédzavarral, és hallásuk az életkoruknak megfelelően ép volt.

A nyelvi anyaghoz a BEA-felvételek közül a mondatfelolvasás, illetve a szövegfelolvasás részt használtuk fel. A /d/, /t/, /k/, /g/ fonémák realizációit Praat szoftverben (Boersma & Weenink, 2017) kézzel annotáltuk, intervokális helyzetben. A /b/ és /p/ explozívák kimaradtak a jelen vizsgálatból, ugyanis a BEA-felolvasás anyagában nem volt elég előfordulás.

A célmássalhangzókat tartalmazó szavak a következők voltak: *speciali**t**ása, kétsé**g**et, ü**t**ötte, i**d**őt, hétvé**g**én, gyere**k**ek, Ba**k**onyba, be**t**egség, elő**a****d**ás, i**d**eig, szi**g**orúan, menetjegye**k**et, i**g**azolványo**k**at, e**g**ész, lehe**t**ett, me**g**ítélni, festménye**k**et, vi**d**é**k**i, me**g**éri, biztosí**t**ást, u**t**azás, ala**k**ultak, i**g**aza, hé**t**en, önma**g**át, éne**k**esnek, világon, kö**t**ött, zöldsége**k**en, növényvé**d**ő, megbetege**d**ést, salá**t**ának, re**t**eknek, felszívó**d**ott, százalé**k**ában, megbetege**d**éseket.* Egyes szavak ismétlődtek. A fonetikai kontextus és a célmássalhangzó helye a szóban változó volt, azonban minden vizsgált explozíva szó belseji intervokális helyzetben állt, hangsúlytalan szótag kezdetén. A kontextusok eltérése miatt a felpattanók egymáshoz viszonyított zöngekezdési idejéből nem vonhatóak le következtetések, azonban az életkori csoportokon belül az egyes célhangok azonos pozíciókban és hangkörnyezetekben szerepeltek.

Az explozívák zárfeloldásakor a felnyíláskori hirtelen nyomáscsökkenés eredményezhet visszazáródást, így többszöri felpattanást (Gráczi & Kohári, 2012; Gráczi, 2012, 2013). A zöngétlen explozívák esetében a mássalhangzó zöngekezdési idejének megállapításához minden esetben a legintenzívebb felpattanást, valamint a zönge (a követő magánhangzó ejtéséhez való) megindulásának idő-

pontját vettük figyelembe. A zöngés explozívák zöngekezdési idejét pedig a megelőző magánhangzó második formánsának lecsengése, valamint a mássalhangzó legintenzívebb felpattanása között határoztuk meg, abban az esetben is, ha a zönge a zárszakasz képzése közben leállt. Minden adatközlő hanganyagában minden vizsgált explozíva tízszer szerepelt, így egy korcsoportban egy mássalhangzónak összesen 300 előfordulása volt található. A zöngekezdési időtartamok, a zöngésrész-arány kiszámításához a annotáltuk a VC- és a CV-határokat, a zönge lecsengésének és megindulásának idejét (ha volt ilyen), a realizáció képzésmódját az annotáció során jelöltük a címkében. A VOT-t és a zöngésrészarányt egy szkript segítségével a címkékből automatikusan mértük. Alternatív képzésmódnak a VOT számítása miatt a detektálható felpattanás nélkül vagy a zöngések esetében a zöngés réshang-/approximánsszerű, a zöngétlenek esetében a frikatívaszerű realizációkat tekintettük. Ezek esetében zárszakasz nem detektálható, hanem a képzés teljes időtartama alatt réses/szűkületes képzésre utaló lenyomat található.

A statisztikai elemzéseket az R programban (R Core Team, 2018) végeztük.

A VOT, a zöngésrész-arány és a realizációtípusok nemek és korok közötti összevetéséhez az egyes mássalhangzókra külön a Scheirer–Ray–Hare-próbát alkalmaztuk (rcompanion, Mangiafico, 2020 és FSA, Ogle et al., 2020 csomagok használatával), amely egy kétfaktoros nonparametrikus próba. A modellekben a zöngésrész-arány, a zöngekezdési idő és a felpattanóként megvalósult realizációk aránya szerepelt függő változóként, míg a nem és a kor szerepelt független változóként ezek interakcióját megengedve. Mivel az adatbázis felolvasott anyagában a hangkörnyezet nem kiegyenlített, a mássalhangzókra külön-külön végeztük el a statisztikai számításokat.

3. Eredmények

3.1. Zöngekezdési idő

A zöngés explozívák esetében mind a nem, mind a kor, mint főhatás hatása szignifikáns volt, a /d/ esetében pedig a kettő interakciójáé is. A zöngétlen fel-

1. táblázat. A zöngés explozívák zöngekezdési ideje nem és kor szerint (i.k.t. = interkvartilis tartomány)

			VOT(ms)			
mássalhangzó	**nem**	**kor**	**átlag**	**szórás**	**medián**	**i.k.t.**
/d/	férfi	fiatal	$-45,00$	$10,87$	$-44,65$	$13,03$
		öregedő	$-51,70$	$13,55$	$-51,40$	$17,30$
	nő	fiatal	$-49,88$	$13,08$	$-50,70$	$13,90$
		öregedő	$-65,67$	$14,91$	$-65,85$	$19,50$
/g/	férfi	fiatal	$-37,14$	$13,93$	$-38,00$	$21,60$
		öregedő	$-47,78$	$13,43$	$-47,90$	$18,83$
	nő	fiatal	$-47,24$	$15,57$	$-46,10$	$18,95$
		öregedő	$-55,51$	$14,88$	$-55,10$	$19,30$

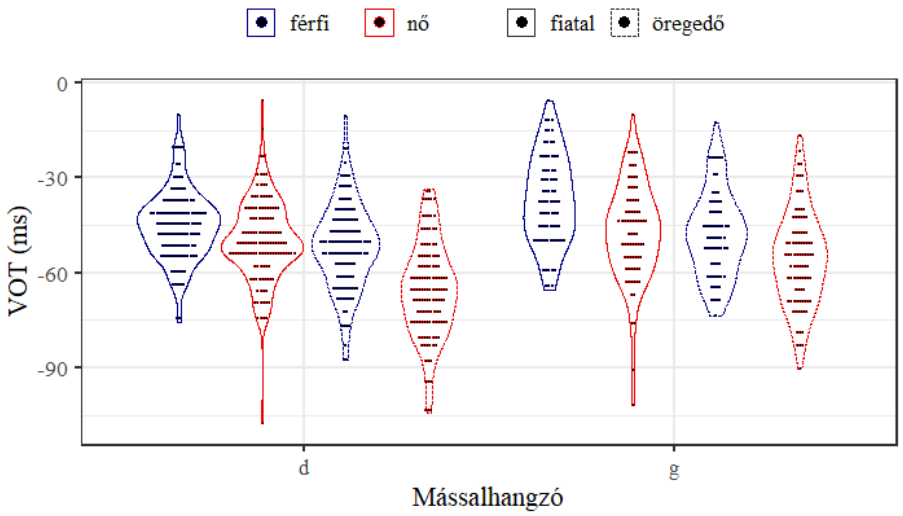
pattanó zárhangok esetében a főhatások nem bizonyultak szignifikáns hatásúnak a zöngekezdési időre, az interakciójuk mentén is csak a /t/ VOT-je mutatott szignifikáns eltérést.

A zöngés explozívákat tekintve elmondható, hogy a /d/ [$H(1,502) = 70,515$, $p < 0,001$] és a /g/ [$H(1,306) = 28,921$, $p < 0,001$] az öregedők beszédében realizálódott szignifikánsan hosszabb zöngekezdési idővel, mind a férfiaknál, mind a nőknél (1. táblázat, 2. ábra).

Az adatokból továbbá kiolvasható, hogy a /d/ [$H(1,502) = 44,937, p < 0,001$] és a /g/ [$H(1,306) = 25,189$, $p < 0,001$] esetében is a nők beszédében realizálódott szignifikánsan hosszabb időtartamban a zöngekezdési idő a fiatal felnőttek és az öregedők korcsoporjátban is (1. táblázat, 2. ábra).

A nem és a kor interakciója a /d/ esetében szignifikáns [$H(1,502) = 7,37$, $p = 0,007$] volt, ami arra vezethető vissza, hogy az öregedő nők beszédében a /d/ hosszabb zöngekezdési idővel realizálódott, mint az összes többi csoportban.

A zöngétlen explozívák esetében a zöngekezdési idők a mássalhangzó szerint különböző módon alakultak. A /t/ zöngekezdési ideje a férfiak és a nők beszédében is az öregedők beszédében realizálódott szignifikánsan hosszabb időtar-



2. ábra. A zöngés explozívák zöngekezdési ideje fiatal felnőttek és öregedők beszédében

tamban [$H(1,574) = 19,834$, $p < 0,001$], a /k/ esetében nem volt eltérés (2. táblázat, 3. ábra).

A nemek tekintetében az adatok alapján elmondható, hogy a /t/ és a /k/ esetében sem volt jelentős eltérés a nemek között (2. táblázat, 3. ábra).

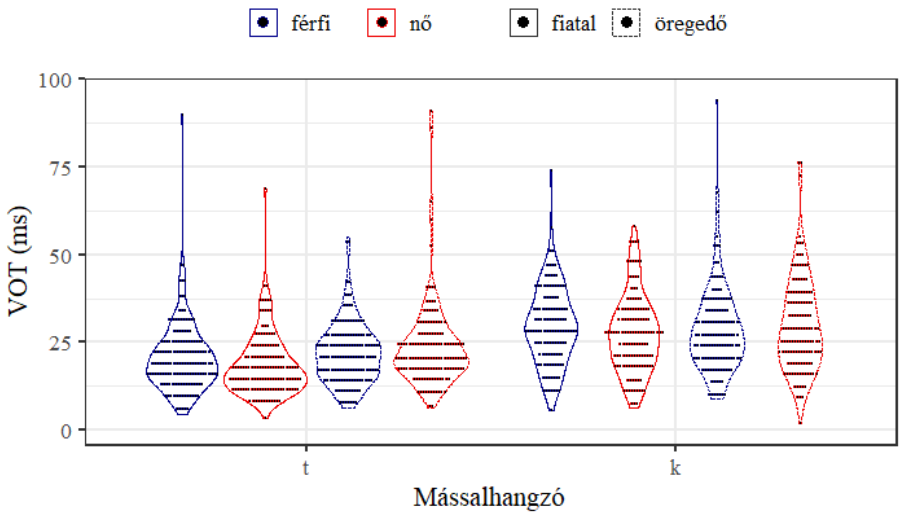
A nem és a kor interakciója a /t/ esetében szignifikáns volt [$H(1,574) = 5,129$, $p = 0,024$]. Ez arra vezethető vissza, hogy a fiatal nők és az öregedő férfiak beszédében ritkábban valósul meg kiugróan hosszabb zöngekezdési idővel a /t/.

3.2. Zöngésrész-arány

Bár a vizsgált beszélői csoportokban ritka és kis mértékű volt a zöngétlenedés a zöngés fonémák realizációiban, a csoportok között mutatkozott statisztikailag szignifikáns eltérés. A zöngés felpattanók zöngésrész-arányára a főhatások közül a nem gyakorolt szignifikáns hatást, illetve a nem és a kor interakciója is meghatározónak bizonyult. A zöngétlen explozívák esetében a főhatások változatos eredményt mutattak, míg a nem és a kor interakciója mindkét konszonáns esetében szignifikáns hatást gyakorolt.

2. táblázat. A zöngétlen explozívák zöngekezdési ideje nem és kor szerint (i.k.t. = interkvartilis tartomány)

			VOT(ms)			
mássalhangzó	**nem**	**kor**	**átlag**	**szórás**	**medián**	**i.k.t.**
/t/	férfi	fiatal	20, 79	10, 17	19, 25	10, 30
		öregedő	21, 80	8, 44	21, 50	10, 45
	nő	fiatal	18, 69	8, 89	16, 50	9, 85
		öregedő	23, 72	11, 91	21, 10	9, 58
/k/	férfi	fiatal	29, 64	11, 09	28, 75	13, 90
		öregedő	28, 56	12, 29	27, 15	13, 60
	nő	fiatal	27, 50	11, 02	27, 10	14, 10
		öregedő	29, 96	13, 29	27, 50	17, 60



3. ábra. A zöngétlen explozívák zöngekezdési ideje fiatal felnőttek és öregedők beszédében

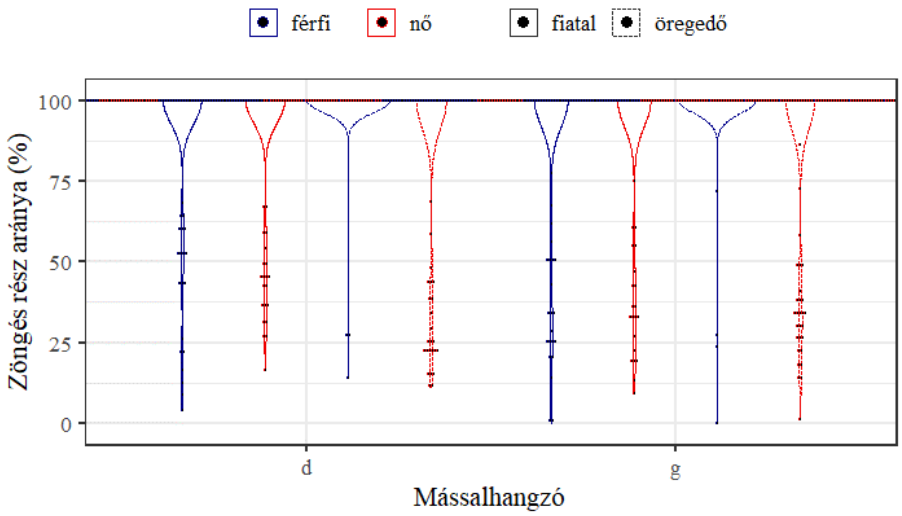
3. táblázat. A zöngés explozívák zöngés rész aránya nem és kor szerint (i.k.t. = interkvartilis tartomány)

			VOT(ms)			
mássalhangzó	**nem**	**kor**	**átlag**	**szórás**	**medián**	**i.k.t.**
/d/	férfi	fiatal	92, 13	21, 39	100	0
		öregedő	98, 39	11, 09		
	nő	fiatal	91, 75	20, 65		
		öregedő	88, 71	26, 42		
/g/	férfi	fiatal	90, 5	24, 07		
		öregedő	98, 09	12, 3		
	nő	fiatal	90, 29	23, 86		
		öregedő	87, 55	26, 85		

A zöngés explozívákat tekintve elmondható, hogy az átlag és szórás figyelembevételével a /d/ és a /g/ esetében is a férfiak beszédében az öregedők korcsoportjában volt valamivel nagyobb a zöngésrész-arány, míg a nők beszédében ennek az ellenkezője érvényesült, azonban a tendencia sem a /d/, sem a /g/ esetében nem volt szignifikáns (3. táblázat, 4. ábra). A viszonylag kis arányú zöngétlenítés miatt azonban a medián és az interkvartilis tartomány azonos (100% és 0%) az összes vizsgálati csoportban.

Az eredmények a nem tekintetében mind a /d/ [$H(1, 581) = 8, 303$, $p = 0, 004$], mind a /g/ [$H(1, 576) = 9, 139$, $p = 0, 003$] esetében szignifikáns különbséget mutattak (átlag és szórás életkori bontás nélkül: /d/ nők: $90, 2 \pm 23, 8\%$, férfiak: $95, 2 \pm 17, 4\%$; /g/ nők: $88, 8 \pm 25, 4\%$, férfiak: $94, 3 \pm 19, 4\%$). A /d/ és a /g/ is a férfiak beszédében jelent meg szignifikánsan nagyobb zöngésrészaránnyal mind a fiatal felnőttek, mind az öregedők korcsoportjában (3. táblázat, 4. ábra).

A nem és a kor interakciója a /d/ [$H(1, 581) = 6, 462$, $p = 0, 011$] és a /g/ [$H(1, 576) = 8, 068$, $p = 0, 005$] esetében is szignifikáns volt. Ez mindkét explo-



4. ábra. A zöngés explozívák zöngés rész aránya a fiatal felnőttek és az öregedők beszédében

zíva esetében arra vezethető vissza, hogy az öregedő férfiak csoportja ritkábban zöngétlenített, mint az összes többi csoport beszélői.

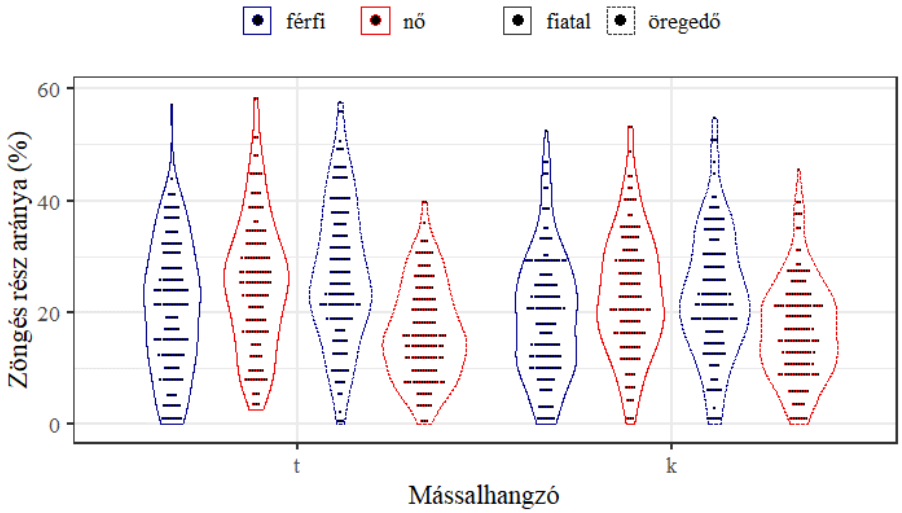
A zöngétlen explozívákat tekintve a /t/ és a /k/ esetében is a férfiak csoportjában az öregedők beszédében jelentek meg az explozívák nagyobb zöngésrész-aránnyal, a nők csoportjában pedig ellentétesen, azaz a fiatal felnőttek beszédében jelentek meg az explozívák nagyobb zöngésrész-aránnyal (4. táblázat, 5. ábra). Azonban ezek a tendenciák sem a /t/, sem a /k/ esetében nem voltak szignifikánsak.

A zöngésrész-arány tekintetében a /t/ esetében a nem szignifikáns különbséget mutatott $[H(1,584) = 8,195,\ p = 0,004]$, a /k/ esetében azonban nem. A fiatal felnőtt korcsoportban a nőknél jelent meg a /t/ és a /k/ is nagyobb zöngésrész-aránnyal, míg az öregedők csoportjában ezzel ellentétesen, a férfiaknál (4. táblázat, 5. ábra).

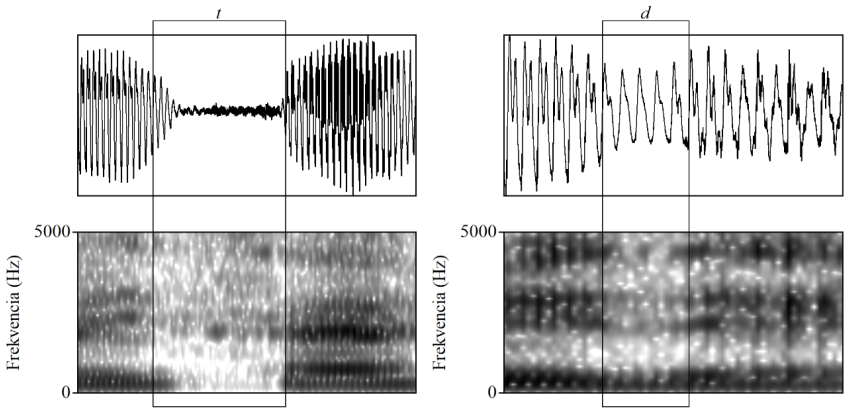
A nem és a kor interakciója a /t/ $[H(1,584) = 49,705,\ p < 0,01]$ és a /k/ $[H(1,582) = 32,720,\ p < 0,01]$ esetében is szignifikáns volt. Ez arra vezethető vissza, hogy mindkét explozíva esetében az öregedő nők ejtésében csengett le a zönge leggyorsabban.

4. táblázat. A zöngétlen explozívák zöngés rész aránya (átlag ± szórás) nem és kor szerint)

			Zöngésrész-arány(%)			
mássalhangzó	**nem**	**kor**	**átlag**	**szórás**	**medián**	**i.k.t.**
/t/	férfi	fiatal	21, 11	13, 37	21, 11	17, 87
		öregedő	28, 43	16, 74	25, 48	18, 00
	nő	fiatal	26, 58	16, 05	25, 83	16, 46
		öregedő	17, 81	12, 67	15, 56	12, 95
/k/	férfi	fiatal	20, 89	15, 84	19, 96	16, 21
		öregedő	24, 34	15, 42	22, 12	15, 45
	nő	fiatal	24, 72	14, 57	22, 42	14, 86
		öregedő	16, 83	9, 09	16, 06	11, 95



5. ábra. A zöngétlen explozívák zöngés rész aránya a fiatal felnőttek és az öregedők beszédében



6. ábra. A /t/ és a /d/ alternatív megvalósulása

3.3. Alternatív képzésmód

A képzésmódváltás során a zöngétlen zárhangokból réshangok, a zöngés zárhangokból pedig közelítőhangok keletkeztek a vizsgált anyagban (6. ábra). A realizációtípusok között nem mutatkozott sem a főhatások, sem azok interakciója mentén egységes tendencia a zöngés, illetve a zöngétlen explozívák esetében sem.

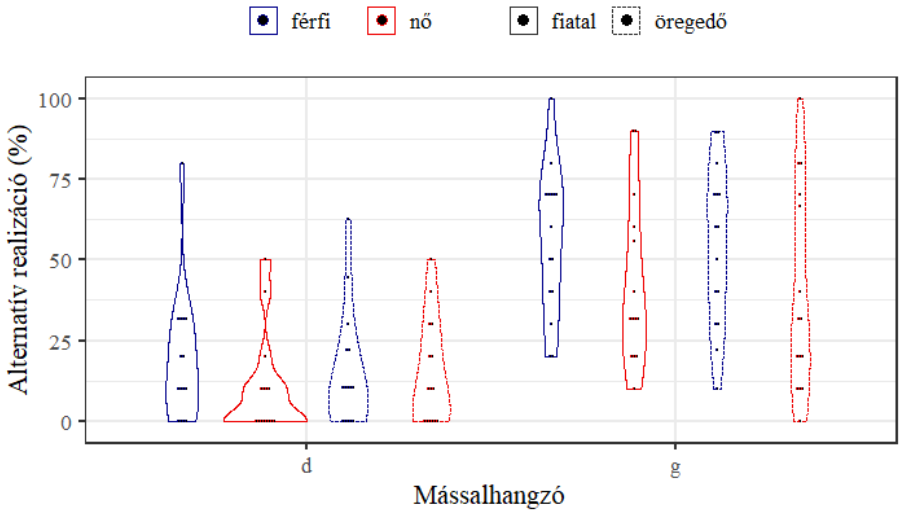
A képzésmód tekintetében elmondható, hogy a férfiak esetében a /d/ és a /g/ is a fiatalok beszédében jelent meg gyakrabban alternatív képzésmóddal. A nők esetében a /d/ az öregedők beszédében jelent meg gyakrabban alternatív képzésmóddal, a /g/ pedig éppen ellenkezőleg, a fiatalokéban. Csak az életkorok mentén elemezve ezeket atendenciákat sem a /d/, sem a /g/ esetében nem voltak szignifikánsak (5. táblázat, 7. ábra).

A nemek tekintetében elmondható, hogy a /d/ és a /g/ is mind a fiatal felnőtt, mind az öregedő korcsoportban a férfiak beszédében realizálódott gyakrabban alternatív képzésmóddal (5. táblázat, 7. ábra). Az eltérés a /d/ esetében nem, de a /g/ $[H(1, 56) = 4, 458, p = 0, 035]$ esetében szignifikáns volt.

A /t/ a férfiak és a nők beszédében is az öregedő korcsoportban jelent meg gyakrabban alternatív képzésmóddal, azonban az eltérés nem volt szignifikáns.

5. táblázat. A zöngétlen explozívák zöngés rész aránya (átlag ± szórás) nem és kor szerint)

			Alternatív realizáció (%)			
mássalhangzó	**nem**	**kor**	**átlag**	**szórás**	**medián**	**i.k.t.**
/d/	férfi	fiatal	18, 89	20, 8	10, 00	25, 00
		öregedő	15, 5	18, 3	10, 00	22, 00
	nő	fiatal	210	15, 58	0, 00	10, 00
		öregedő	14, 67	16, 42	10, 00	25, 00
/g/	férfi	fiatal	56	22, 93	60, 00	30, 00
		öregedő	54, 07	24, 69	60, 00	20, 00
	nő	fiatal	43, 02	26, 02	33, 33	36, 39
		öregedő	39, 33	31, 68	30, 00	53, 33



7. ábra. Alternatív képzésmóddal realizálódott zöngés explozívák aránya a fiatal felnőttek és az öregedők beszédében

6. táblázat. A zöngétlen explozívák alternatív realizációja nem és kor szerint (i.k.t. = interkvartilis tartomány)

			Alternatív realizáció (%)			
mássalhangzó	**nem**	**kor**	**átlag**	**szórás**	**medián**	**i.k.t.**
/t/	férfi	fiatal	0,67	2,58	0,00	0,00
		öregedő	2	4,14		
	nő	fiatal	0,67	2,58		
		öregedő	2,15	6,12		
/k/	férfi	fiatal	20,73	18,55	0,00	10,00
		öregedő	13,79	15,98	10,00	20,00
	nő	fiatal	7,67	10,83	0,00	15,00
		öregedő	13,41	23,79	11,11	15,56

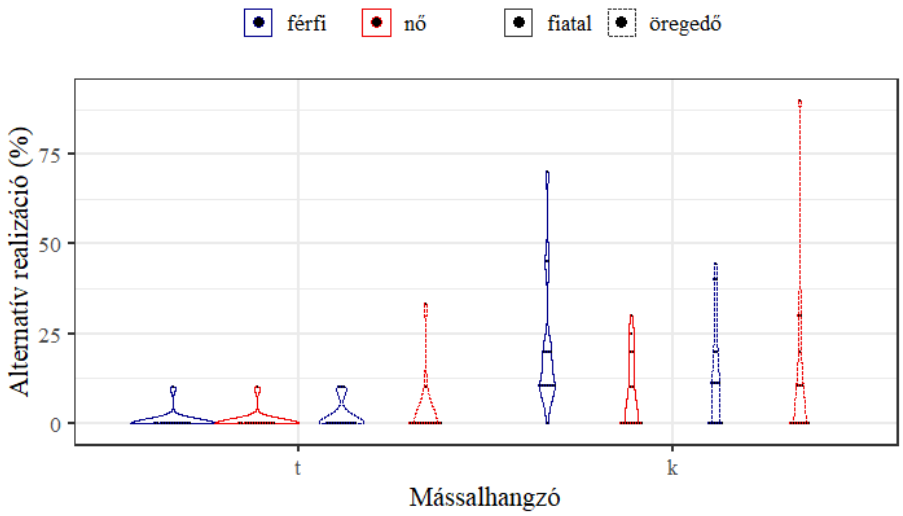
A /k/ a férfiak beszédében a fiatal felnőtteknél jelent meg gyakrabban alternatív képzésmóddal, míg a nőknél ezzel ellentétes tendencia mutatható ki, tehát az öregedőknél jelent meg gyakrabban alternatív képzésmóddal, de az eltérés nem volt szignifikáns (6. táblázat, 8. ábra).

A nemek tekintetében elmondható, hogy a /t/ alternatív realizációja nem mutatott eltérést egyik korcsoportban sem.

A /k/ mind a fiatal felnőttek, mind az öregedők korcsoportjában a férfiak beszédében jelent meg nagyobb arányban alternatív képzésmóddal, amely eltérés szignifikáns [$H(1,56) = 4,578$, $p = 0,032$] (6. táblázat, 8. ábra).

4. Következtetések

Kísérletünkben az alveoláris és a veláris zöngés, valamint zöngétlen explozívák zöngekezdési idejét, zöngésrész-arányát, képzésmódváltását vizsgáltuk fiatal felnőttek és öregedők felolvasásában. A kutatás során hat hipotézist fogalmaztunk meg. A korábbi nemzetközi és magyar szakirodalom alapján azt vártuk, hogy a zöngétlen és zöngés explozívák esetében különböző módon alakulnak a



8. ábra. Alternatív képzésmóddal realizálódott zöngétlen explozívák aránya a fiatal felnőttek és az öregedők beszédében

zöngekezdési idők mind a korcsoport, mind a nem viszonylatában. Hipotéziseinket a korábbi vizsgálatok eredményei alapján állítottuk fel. Az alábbiakat feltételeztük: A zöngétlen zárhangok esetében az öregedők beszédében jelenik meg az explozíva hosszabb zöngekezdési idővel, míg a zöngések esetében a fiatal felnőttekében. A nem tekintetében azt vártuk, hogy a zöngés explozívák esetében a nőknél, a zöngétlenek esetében pedig a férfiaknál jelenik meg az explozíva hosszabb zöngekezdési idővel. A zöngésrész-arány tekintetében azt vártuk, hogy az öregedők felolvasásában a vizsgált explozívák kisebb zöngésrész-aránnyal jelennek meg. A képzésmódváltással kapcsolatban pedig azt feltételeztük, hogy az öregedők beszédében gyakrabban jelennek meg az explozívák alternatív képzésmóddal a fokozatosan megjelenhető renyhébb artikuláció miatt.

A zöngekezdési idő a zöngés felpattanók esetében az idősebb beszélők ejtésében volt hosszabb, mint a fiatalokéban, illetve a nők beszédében volt hosszabb, mint a férfiakéban. A zöngétlen explozívák esetében azonban nem találtunk sem az életkorral, sem a nemmel összefüggő tendenciát. Mind a zöngés, mind a zöngétlenek esetében az alveoláris explozívák esetében volt a kor és a nem

interakciójának szignifikáns hatása. Összességében tehát a zöngések esetében találtunk jellemző tendenciát. Ki kell emelni, hogy ezek VOT-jét a zárszakasz kezdetéig mértük minden esetben, azaz zöngétlenedett konszonánsok esetében is. Így a hosszabb zárszakasz lehet az esetlegesen lassabb artikuláció következménye. Az eredmények tehát az életkorral összefüggő hipotéziseinket nem támasztották alá.

A zöngésrész-arány nem mutatott szisztematikus eltérést sem a nem, sem a kor tekintetében, azaz az erre vonatkozó hipotéziseinket nem támasztotta alá. Azt feltételeztük, hogy az idősebbek esetében gyakoribb/nagyobb arányú lesz a részben zöngétlenedett /d/- és /g/-realizációk megjelenése az öregedés okozta nehézségek következtében. Ez nem jelent meg, a két nemi csoportban eltérő tendenciákat találtunk az életkor mentén.

Az alternatív realizációk, tehát a zöngések réses-approximánsos, illetve a zöngétlenek réses képzése az életkorral nem mutatott egyértelmű összefüggést, míg a velárisok (tehát a /g/ és a /k/) esetében a férfiak ejtésében gyakoribb volt ezen realizációk előfordulása. A velárisok gyakoribb alternatív realizációja feltehetően azzal magyarázható, hogy a hátsóbb képzési helyen a zár fenntartása jelentősebb kompenzációs stratégiát igényel a gyorsabban megnövekvő szupraglottális nyomás miatt. Az, hogy ez a férfiak esetében volt jellemzőbb, magyarázható talán lazább artikulációval, ehhez azonban számos további vizsgálat lenne szükséges. Az idősebb beszélők 65 és 77 év közötti beszélők voltak. Az artikulációs szervek működésének renyhülése miatt feltételeztük az ő beszédükben gyakoribbnak az alternatív képzésmód megjelenését. Ez azonban (bizonyos mértékig) kompenzálható feszesebb artikulációval, illetve a beszélők közötti variabilitás is nagy ebben az életkorban.

A képzési hely szerint jelen vizsgálatban nem vetettük össze az explozívákat, mert a BEA-adatbázisból vett olvasott anyagban nem kiegyenlített a hangkörnyezet. A jelen kísérletet felolvasott értelmes szavakon, kontextusba ágyazva végeztük. A magyar zöngekezdési idők, a zöngésrész-arány és a képzésmódváltás teljesebb megismeréséhez a kísérlet kiterjeszthető bilabiális és palatális explozívákra is. Valamint további kutatási irány lehet az explozívák zöngekez-

dési idejének, zöngésrész-arányának és képzésmódváltásának vizsgálata izolált szavakban, logatomokban, a magánhangzó-környezet kontrollálásával, a szóban elfoglalt pozíciója alapján és spontán beszédben is.

A jelen kutatás látszólagos időt vizsgál, így a továbbiakban szükséges olyan longitudinális vizsgálatot végezni, ahol fő kutatási szempont az életkor előrehaladtával megjelenő befolyásoló tényezők, pl. protézis, fogpótlás hatása a mássalhangzó képzésre. A jelen vizsgálatot 60 adatközlő beszédprodukcióján végeztük el, így a statisztikai elemzések megbízhatóan jelzik a magyar zöngekezdési idők változását az életkor előrehaladtával, valamint a nembeli különbségeket a zöngésrész-arányok tekintetében. Kutatásunk eredményei az időskori beszéd, valamint a nemi különbségek pontosabb megismerése mellett olyan gyakorlati alkalmazásokban is felhasználhatók, mint például a kriminalisztikai fonetika.

Köszönetnyilvánítás

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal FK-128814 számú pályázata támogatta.

Hivatkozások

Alphen, P. M. van., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics*, *32*, 455–449.

Balázs, B. (1993). Az időskori hangképzés jellemzői. *Beszédkutatás*, *93*, 156–165.

Benjamin, B. J. (1982). Phonological performance in gerontological speech. *Journal of Psycholinguistic Research*, *11*, 159–167.

Bickley, C. A., & Stevens, K. N. (1986). Effects of a vocal tract constriction on the glottal source: experimental and modeling studies. *Journal of Phonetics*, *14*, 373–382.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer. URL: http://www.fon.hum.uva.nl/praat/download_win.html utolsó letöltés: 2017.08.22.

Bóna, J. (2011). A [p, t, k] mássalhangzók zöngekezdési ideje idősek és fiatalok spontán beszédében. *Beszédkutatás*, (p. 61–72).

Bóna, J. (2016). *Női beszéd – férfi beszéd a fonetikai és a pszicholingvisztikai vizsgálatok tükrében*. Budapest: Akadémiai Kiadó.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*, 129–159.

Cole, R., & Cooper, W. E. (1975). Perception of voicing in English affricates and fricatives. *Journal of the Acoustical Society of America*, *58*, 1280–1287.

Crowther, C. S., & Mann, V. (1992). Native language factors affecting use of vocalic cues to final consonant voicing in English. *Journal of the Acoustical Society of America*, *92*, 711–722.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, *27*, 761–764.

G. Kiss, Z., & Bárkányi, Z. (2018). A fonetikai korrelátumok szerepe a zöngekontraszt fenntartásában – beszédprodukciós és észleléses eredmények. *Általános nyelvészeti tanulmányok*, *30*.

Gandour, J., & Dardarananda, R. (1984). Voice onset time in aphasia: Thai II. *Production. Brain and Language*, *23*, 177–205.

Gósy, M. (2000). A [p t k] mássalhangzók zöngekezdési ideje. *Magyar Nyelvőr*, *124*, 195–204.

Gósy, M. (2008). Magyar spontánbeszéd-adatbázis – BEA. *Beszédkutatás*, (p. 194–207).

Gósy, M., & Ringen, C. O. (2009). Everything you always wanted to know about VOT in Hungarian. In *Előadás. International Conference on the Structure of Hungarian* (p. 1). Budapest.

Gráczi, T. E. (2011). Intervokális explozívák a zöngésségi oppozíció függvényében. *Beszédkutatás*, (p. 46–60).

Gráczi, T. E. (2012). Zörejhangok akusztikai fonetikai vizsgálata a zöngésségi oppozíció függvényében.

Gráczi, T. E. (2013). Explozívák és affrikáták zöngésségének időviszonyai. *Beszédkutatás*, (p. 94–120).

Gráczi, T. E. (2016). A zöngekezdési időről. In J. Bóna (Ed.), *Fonetikai olvasókönyv* (p. 61–73). Budapest: ELTE Fonetikai Tanszék.

Gráczi, T. E., & Kohári, A. (2012). A zöngekezdési idő egy módszertani kérdés függvényében. In A. Markó (Ed.), *Beszédtudomány. Az anyanyelvelsajátítástól a zöngekezdési időig* (p. 228–248). Budapest: ELTE Bölcsészettudományi Kar–MTA Nyelvtudományi Intézet.

Gráczi, T. E., Markó, A., & Beke, A. (2009). *Zöngekezdési idő a spontán beszédben*. Elhangzott: Beszédkutatás.

Henton, C., Ladefoged, P., & Maddieson, I. (1992). Stops in the world's languages. *Phonetica*, *49*, 65–101.

Huber, J. E. (2008). Effects of utterance length and vocal loudness on speech breathing in older adults. *Respiratory physiology & neurobiology*, *164*, 323–330.

Jesus, L. M. T., & Shadle, C. H. (2003). Temporal and devoicing analysis of European Portuguese fricatives. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *5th International Congress of Phonetic Sciences* (p. 779–782).

Kovács, M. (2000). Kontextushatás a beszédhangok időviszonyaiban. *Beszédkutatás*, (p. 15–25).

Krause, S. E. (1982). Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *Journal of the Acoustical Society of America*, *71*, 990–995.

Ladefoged, P. (2005). *Vowels and consonants: An introduction to the sounds of languages*. Oxford: Blackwell Publishing.

Levitzky, M. G. (1984). Effects of aging on the respiratory system. *Physiologist*, *27*, 102–107.

Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics. Cambridge Studies in Speech Science and Communication*. Cambridge: Cambridge University Press.

Lisker, L. (1957). Closure duration and intervocalic voiced-voiceless distinction in English. *Language*, *33*, 42–49.

Lisker, L. (1986). „voicing" in English: A Catalogue of Acoustic Features Signaling /b/ Versus /p/ in Trochees. *Language and Speech*, *29*, 3–11.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops. *Word*, *20*, 384–422.

Mangiafico, S. (2020). rcompanion: Functions to support extension education program evaluation. URL: https://CRAN.R-project.org/package=rcompanion R package version 2.3.25,.

Ogle, D. H., Wheeler, P., & Dinno, A. (2020). FSA: Fisheries Stock Analysis. URL: https://github.com/droglenc/FSA R package version 0.8.30,.

Ohde, R. N., & Haley, K. L. (1997). Stop-consonant and vowel perception in 3- and 4-year-old children. *Journal of the Acoustical Society of America*, *102*, 3711–3722.

Petrosino, L., Colcord, R. D., Kurcz, K. B., & Yonker, R. J. (1993). Voice onset time of velar stop productions in aged speakers. *Perceptual and Motor Skills*, *76*, 83–88.

Pisoni, D., & Remez, R. (Eds.) (2005). *The handbook of speech perception*. Oxford: Blackwell Publishing.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics*, *32*, 141–152.

R Core Team (2018). R: A language and environment for statistical computing. URL: https://www.R-project.org/. utolsó letöltés: 2018.07.03.

Riper, C. van., & Irwin, J. V. (1961). *Voice and Articulation*. Englewood Cliffs. N.J: Prentice-Hall, INC.

Ryalls, J., Cliche, A., Fortier-Blanc, J., Coulombe, I., & Prud'Hommeax, A. (1997). Voice-onset time in younger and older French-speaking Canadians. *Clinical Linguistics and Phonetics*, *11*, 205–212.

Shadle, C. H. (1997). The Aerodynamics of Speech. In W. J. Hardcastle, & J. Laver (Eds.), *Handbook of Phonetics* (p. 33–64). Oxford: Blackwell.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, Massachusetts–London: The MIT Press.

Summers, W. V. (1988). F1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America*, *84*, 485–492.

Swartz, B. L. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills*, *75*, 983–992.

Száraz, B. (2019). Explozívák és az őket megelőző magánhangzók időtartama modális fonációval létrehozott beszédben és suttogásban. *Beszédkutatás*, *27*, 75–86.

Tuomainen, O. T., & Lely, H. (2007). Processing of acoustic cues for voicing in English: a MMN study. In *Presented at: 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany.

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*, 145–166.

The intonation of lengthenings in northern and southern dialects of Spanish

Kata Baditzné Pálvölgyi[1]

[1] *Eötvös Loránd University*

Abstract

This research focuses on the prosodic patterns of lengthenings attested in northern and southern dialects of European Spanish, more precisely, on their intonation. A corpus of 200 spontaneous utterances has been elaborated (including 100 utterances from the northern dialects and 100 utterances from the southern ones, produced by 16 male and 16 female informants, respectively). The analysis has been carried out following the standardization protocol offered by Cantero Serena & Font-Rotchés (2009), Cantero Serena (2019) and Cantero & Font-Rotchés (2020), in which the representative values of intonation (in Hz) are taken for each syllable, and then these values undergo a process of standardization, in order to be comparable objectively and speaker-independently. It is expected that lengthenings show no remarkable inter-dialectal melodic differences. Also, it is predicted that lengthenings do not present prominent prosodic features as compared to their context, because they may serve as a tool for the speaker to maintain the conversational turn, without interrupting the tonal movement of the utterance.
Keywords: dialects, European Spanish, lengthenings, intonation, standardization

1. Introduction

Disfluency phenomena include, for example, noises, repetitions, false starts, silent pauses, repairs, truncations, filled pauses and lengthenings (Eklund, 2004; Gósy, 2002; Lickley, 1994, 2015; Rodríguez et al., 2001; Shriberg, 1994), the last two phenomena being the two most common subtypes of hesitation (Deme & Markó, 2013).

Lengthenings form part of phenomena that are applied to gain time without implying necessarily the interruption of elocution. Their aim is to slow down the velocity of speech without affecting communication (Rebollo Couto, 1997, p. 667). Filled pauses and lengthenings are typically considered to be two

Email address: b.palvolgyi.kata@btk.elte.hu (Kata Baditzné Pálvölgyi)

different acoustic disfluency phenomena (Rodríguez et al., 2001, 2015), but for some, lengthenings are subsumed under filled pauses (Maclay & Osgood, 1959, cited by (Machuca Ayuso, 2018)), as a special, "lexical" type (Blondet, 2001; Villa Villa, 2017). Filled pauses in Spanish are characterized by the Spanish vowel [e] (Machuca Ayuso et al., 2015), or, in less cases, the Spanish vowel [a] or even the consonantal [m] (Garrido Almiñana et al., 2017). Lengthenings affect primordially unstressed syllables in Spanish (Rebollo Couto, 1997). The most common lengthened vowel is the [a], whereas amongst consonants, word-final [l], [n] or [s] (Baditzné Pálvölgyi, 2019). The following examples show how these hesitation phenomena are realized in their context (all the utterances are taken from my corpora):

(1) ECA-2-3 *Vale rodea...* [a:] *eh* [e:] *la...* [a:] 'Ok, then you go around eh the'

EMA-1-2 *Y tengo que ir a...* [a:] *la tienda Nueva Moda* 'And I have to go to the Shop New Fashion'

EGI-2-3 *Mmm* [m:] *pues continúa por allí* 'Mmm then you go on there'

EGI-2-5 *Eh* [e:] *continúas todo de frente y te vas a encontrar el...* [l:] *el ayuntamiento* 'Eh you go straight on and you will find the town council'

EOV-1-1 *Ves un* [n:] e*stablecimien*[n:]*to que se llama* [a:] *Modas Nuria* 'you see a building called Nuria Fashion'

EOV-1-3[a] *pues*[s:] *luego* 'well then'

The present research focuses on the prosody of lengthenings – more precisely, on their melodic characteristics – in two well-defined dialects of European Spanish.

Spanish is spoken by more than 400 million speakers all over the world, and due to this fact, its dialectology presents a considerable variation. We can distinguish two main dialectal areas in Europe that share several common characteristics in their pronunciation: those from the north (including also central varieties) and the dialects from the south (including the varieties spoken in the Canary Islands, cf. Hualde, 2014, 285–288). Taking into account this dichotomy,

my objective has been to compare the strategies of hesitation applied by the dialects of Spanish from north and south, in three aspects related to prosody: (a) intonation, (b) intensity and (c) the duration of prolonged segments and filled pauses (Baditzné Pálvölgyi, 2020). I conducted an investigation that compared 100 sentences provided by speakers of northern Spanish dialects with 100 sentences taken from informants of southern Spain. In both corpora the same methodology was applied, a three-phase prosodic analysis of speech proposed by Cantero Serena (2019), in order to answer two research questions:

1. Do these two dialects show prosodic differences in case of hesitation phenomena?
2. Are hesitation phenomena prosodically salient as compared to their adjacent context?

This study aims at answering part of the questions mentioned above: it describes the intonational aspects of lengthenings in northern and southern dialects of European Spanish.

It is essential to define where exactly dialectal differences take place within the scope of intonation. The dialectology of Spanish intonation traditionally was based on the description of the shape of the whole contour, focusing primordially on the characteristics of the final tonal movement parting from the last accented syllable (cf. Quilis, 1999, 454–483; Sosa, 1999, 177–245; Hualde, 2014, 280–281). Cantero Serena (2002, 86–88), however, claims that it is rather the parsing of speech into blocs indicated by the melodic changes anchoring in stressed syllables that makes dialectal accents recognizable. In this sense, it is the melody (and in a broader sense, prosody) of stressed syllables that is responsible for dialectal intonation. By accepting either of these two points of view, as lengthenings in Spanish generally affect unstressed segments (Rebollo Couto, 1997), it is expected that they are not responsible for considerable differences of dialectal intonation. Regarding Latin American Spanish dialects, Blondet (2001), who considers lengthenings as lexical filled pauses, did not find any considerable differences in the intonation of lengthenings in different Venezuelan

dialects. Apart from this work, no significant studies have been carried out in the field of comparative intonation of lengthenings. The lack of the treatment of hesitation phenomena in intonational studies can partly be explained by the fact that most of the comprehensive works are based on the analysis of read sentences (cf. Sosa, 1999; Quilis, 1999) – or even sentences taken from different literary genres such as novels, as in Navarro Tomás (1966). As read and rehearsed sentences are less characterized by hesitation phenomena than spontaneous conversations, it is not surprising that we do not find abundant papers concerning the intonation of lengthenings. Unfortunately, the extended dialectal study within the autosegmental framework, carried out by Prieto et al. (2010), which collected wide data of induced (but not read) and spontaneous sentences, does not include in its analysed samples spontaneous utterances either.

Based on what has been revealed so far, according to my first hypothesis concerning the intonation of lengthenings in Spanish dialects, (1) there will not be significant differences in the intonational aspect of lengthenings between northern and southern variants of European Spanish.

Regarding research question (2), my prediction is that lengthenings will show no salient tonal movements as compared to their immediate context. This hypothesis can be explained by the observation that hesitation phenomena such as lengthenings in general cannot only be defined as the blockers of fluency; they can also guarantee that the speaker will hold the dialogue turn, so their role in spoken discourse is definitely important. This is also true in case of Spanish, a language known for the so-called 'Mediterranean debate' rules, in which native speakers hold and gain conversational turns with apparent vehemence (Berry, 1994).

According to previous research, the melody of filled pauses is rather plain when the speaker has no specific communicative function with the hesitation, only 'gaining time', but if the vocalization is accompanied by an emotion or is used to check listeners' attention, they show special tonal movements, such as a fall-rise (Garrido Almiñana et al., 2017). As for lengthenings, Blondet (2001) found in Venezuelan Spanish that lengthenings typically showed a linear

melodic fall. This means that lengthenings completely adopted the by default descending melody of the declination observable in Spanish utterances, without interrupting it tonally.

For this reason, according to my assumption, utterance-internal lengthenings must not be prosodically prominent, as their only aim is to be a continuation to their context. If they presented abrupt movements, they would definitely break the prosody of the utterance, and might cause the speaker to lose his/her conversational turn.

Based on what has been revealed so far, in this study I focus on lengthenings from a prosodic point of view, more concretely, from a melodic perspective, by formulating two hypotheses:

1. utterance-internal lengthenings do not present different intonational behavior in the two examined dialects;
2. utterance-internal lengthenings are not characterized by prominent melodic movements compared to their adjacent context.

2. Corpus and informants

The corpus was obtained from the 'Map Task' activities in the interactive Atlas of Romance intonation compiled by Prieto et al. (2010–2014), on the one hand, and on the other, of spontaneous interviews uploaded to YouTube. I have analysed all the utterances that presented lengthenings in case of the Map Tasks, and this corpus was completed by the same number of utterances containing lengthenings in case of the interviews, in a way that male and female speakers were represented equally in both corpora. This way we obtained only spontaneous speech samples with the same speech style (spontaneous conversations, the speakers of which are aware of being recorded). 32 speakers were selected altogether, 16 informants from the north (8 men and 8 women), and 16 from the south of Spain (also 8 men and 8 women), from recordings of 291 minutes and 30 seconds in total. In the northern corpus, 116 lengthenings have

been detected, compared to the 120 cases in the southern corpus. Table 1 and table 2 sum up the data related to the informants.

Only monolingual areas were chosen for the analysis (leaving apart thus, territories such as Catalonia, Valencia or the Balearic Islands (Catalan-speaking zones), Galicia (Galician-speaking zone), the Basque Country and La Rioja (Basque-speaking zones), because these regions could have shown influences by other peninsular languages.

3. Method

The theoretical background used in this work is based on the intonational theory presented by Cantero Serena (2002), implemented later by a protocol for melodic analysis (Cantero Serena & Font-Rotchés, 2009; Cantero & Font-Rotchés, 2020), and a protocol for prosodic analysis (Cantero Serena, 2019).

Within this model, the smallest unit for the melodic analysis is the tonal segment with the relative tonal value of the syllabic nucleus (in Spanish, almost exclusively the vowel). Each vowel constitutes one tonal segment, except for accented vowels, which can constitute tonal inflections, that is, combinations of two or more tonal segments. Consonants occupy a marginal status in the syllable, except for nasals and liquids, which may, in certain cases, hold a tonal contrast alone.

For Cantero, intonation is defined as relevant f0 variations in the utterances (2002, p. 18). Other elements sometimes traditionally considered as part of intonation (such as tempo, intensity, duration, timbre) are out of his scope. Cantero holds that alterations in tempo or intensity are non-melodic changes, often analysed as emphatic features of intonation (2002, p. 178). Intensity is subsumed under the definition of intonation in Quilis (1981, p. 394), for example, but Cantero regards that intensity can add intonational information only in whispered speech, where there is no f0. Di Cristo (1982), Gili Gaya (1924), Hombert (1978) and Mateo (1988) all consider timbre as potentially part of intonation; Cantero, nevertheless, excludes this possibility (2002, p. 17–18).

Table 1: The informants' data

Northern Spanish data					
origin	**speakers**	**sex**	**age**	**no of utterances selected**	**duration**
Gijón (Map Task)	Speaker 1	f	24	5	5:49'
	Speaker 2	f	22	7	
Oviedo (Map Task)	Speaker 1	f	20	9	4:52'
	Speaker 2	f	25	6	
Cabezón de la Sal (Map Task)	Speaker 1	f	31	12	11:50'
	Speaker 2	f	31	2	
Madrid (Map Task)	Speaker 1	f	33	3	14:00'
	Speaker 2	f	37	5	
Salamanca (interviews)	Speaker 1	m	49	7	4:39'
	Speaker 2	m	57	6	21:36'
Burgos (interviews)	Speaker 1	m	59	6	15:53'
	Speaker 2	m	36	6	16:16'
Ávila (interviews)	Speaker 1	m	60	6	22:31'
	Speaker 2	m	59	7	25:23'
León (interviews)	Speaker 1	m	56	6	26:30'
	Speaker 2	m	51	6	27:24'
age (years; mean)				40,63	
utterances (total)				100	
Southern Spanish data					
origin	**speakers**	**sex**	**age**	**no of utterances selected**	**duration**
Canary Islands (Map Task)	Speaker 1	f	38	6	4:41'
	Speaker 2	m	38	8	
Jaén (Map Task)	Speaker 1	f	22	10	4:14'
	Speaker 2	m	21	1	
Constantina (Map Task)	Speaker 1	f	23	8	2:42'
	Speaker 2	f	22	8	
Jerez de la Frontera (Map Task)	Speaker 1	f	41	3	3:15'
	Speaker 2	m	46	6	
Málaga (interviews)	Speaker 1	m	49	6	14:05'
	Speaker 2	m	45	6	18:14'
Sevilla (interviews)	Speaker 1	f	50	7	15:19'
	Speaker 2	m	51	6	
Badajoz (interviews)	Speaker 1	m	47	6	12:18'
	Speaker 2	m	59	6	6:20'
Granada (interviews)	Speaker 1	f	55	6	8:37'
	Speaker 2	f	47	7	5:03'
age (years; mean)				40,88	
utterances (total)				100	

Pitch, duration and intensity are considered to be suprasegmental features, and as such, are relatively difficult to interpret. First, because we must neglect speaker-dependent characteristics that carry no linguistic significance, and second, because prosodic units must be understood as bearing relative prominence with respect to adjacent units, so they have no information alone.

A solution to overcome these difficulties is offered by Cantero's Melodic Analysis of Speech (MAS) (2009) and his latter implementation to the theory, Prosodic Analysis of Speech (PAS) (2019). As for melodic analysis, there is an acoustic phase, assisted by an acoustic analysis software. The second step is the melodic representation: in order to concentrate only on the melodically relevant features, it is necessary to ignore irrelevant micromelodic variations and reduce the intonational contour in case of each syllable to a characteristic frequency value (or in case of syllables with tonal instability, to two or three values, depending on the tonal inflection carried by the syllable). The third step is the melodic standardization: the contours are represented taking into account not the absolute values, but the relative ones, as each syllable is given a percentage based on its melodic rise/fall experienced with respect to the previous syllable (Baditzné Pálvölgyi, 2012). The same algorithm is used in case of intensity in the extended PAS model (the relative values are intensity peaks associated to each syllable with respect to the previous one) and duration (the relative duration of each syllable with respect to the previous one).

This analysis permits us to describe more objectively the prosodic features of a given language, and compare prosodically, for instance, dialects. The melodic process is presented in the next section, using examples from my corpora.

3.1. The standardization of tonal data in the MAS model

It is an essential step in the MAS model that the original f0 curve is reduced to a standardized copy of it without micromelodic variations, ultimately by the help of the analysis and synthesis program Praat (Boersma & Weenink, 2019). Standardization of contours was first done using semitones in the 'Dutch School', also known as the IPO model. The most emblematic work of this

approach is t'Hart et al. (1990), which was followed by various researches in different languages (Adriaens, 1991; Beaugendre, 1994; Odé & Heuven, 1994). In Spanish, Garrido (1991, 1996) and Estruch et al. (1999) worked with similar automatic stylization methods (Baditzné Pálvölgyi, 2012).

The difference between the standardized curves in the MAS model and the ones in the Dutch School is that the MAS model uses percentages for the standard values, which is a system easier to handle than the one with semitones. The standardized contour is represented by a line which starts with an arbitrary value of 100% and anchors in each syllable, which is itself characterized by a percentage based on its tonal position as compared to the previous syllable. If the syllable is located lower, it is a negative percentage, and if it is higher than the previous syllable, it is a positive one. The standardized contour, as in the case of the Dutch school, is submitted to perceptual tests so as to confirm that it is melodically identical to the original curve; if not, it is corrected manually. The percentages can show more than the autosegmental labels would, because they can express illocution (in Spanish, for example, an utterance-final rise of over 80% is perceived as interrogative tone). Still, according to Font-Rotchés, the MAS analysis is compatible with autosegmental labeling, as it is a model that also permits any subsequent type of annotation, including ToBI methodology, cf. (Font-Rotchés & Mateo Ruiz, 2011, p. 1112). Though first applied to Spanish intonation (Cantero Serena et al., 2005; Cantero Serena & Font-Rotchés, 2007; Font-Rotchés & Mateo Ruiz, 2011), it has been extended to the study of intonation in other languages as well, such as Catalan (Font-Rotchés, 2005, 2007, 2009), or Chinese (Kao, 2011). For a partial Spanish application see Patiño (2008). In Hungarian, a similar analysis was carried out in Olaszy & Koutny's investigation, also based on percentages and stylized contours. For them, however, the first value (100%) is not an arbitrary number, but the first abstract f0 value of declarative sentences. Yes-no questions start at 80% as compared to this value (Olaszy & Koutny, 2001, p. 182–183).

This model has also been applied in the description of the intonation of interlanguages, for example the Spanish spoken by Brazilians (Fonseca & Can-

tero Serena, 2011), Italians (Devís, 2011), Swedes (Martorell, 2011) or Hungarians (Baditzné Pálvölgyi, 2011, 2012, 2018, 2019).

3.2. The steps of tonal standardization

The first phase of the analysis guarantees that we get rid of irrelevant micromelodic variations, reducing each syllable to a characteristic tonal value. In case of tonal instability within a syllable, the extreme values of f0 are taken. The following figures exemplify the process.

Figure 1 below shows how the distinct tonal values can be perceived in the utterance *Y viendo el patrimonio monumental de la...* 'And seeing the monumental patrimony of the...' (an utterance from my corpus, with the speaker from Badajoz; the image is produced by the voice analysis software Praat, the text is my addition):

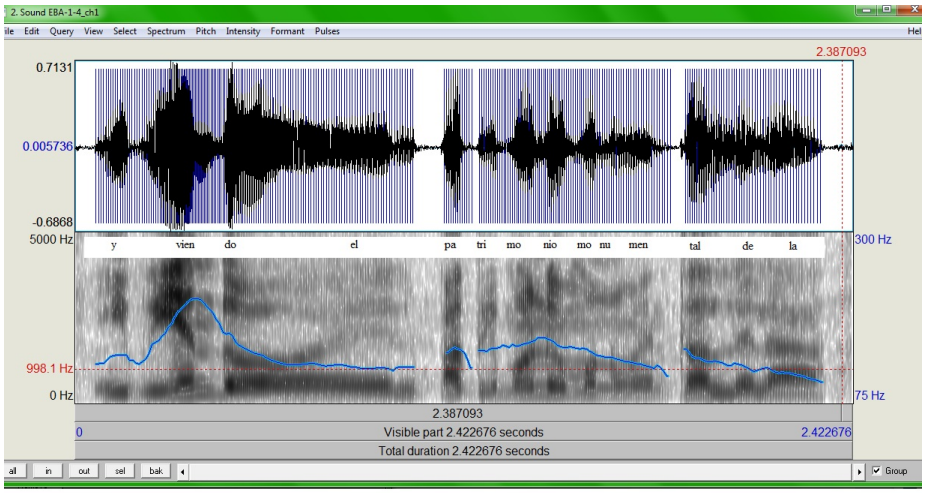


Figure 1: Tonal values of an utterance from Badajoz

The vowels of the utterance which are characterized by tonal stability are measured at their middle point. This is so in case of the first vowel [i] of the word *y*, as, though visually it appears to be tonally instable, in fact the tonal movement that characterizes the syllable is unperceivable. Figure 2 shows that the minimum value is 125,5 Hz, the maximum value is 138,4 Hz (measured in the

middle), so the difference between them does not reach the perception threshold of 10% (Font-Rotchés & Mateo Ruiz, 2011).

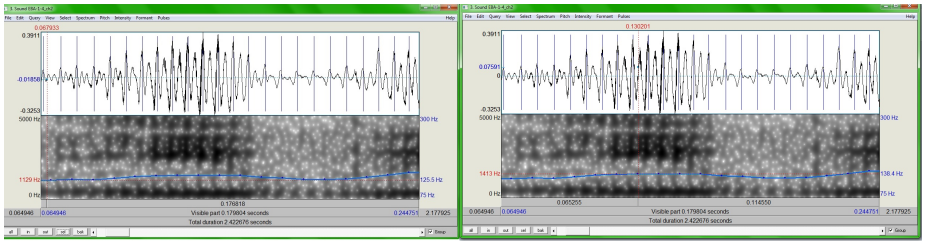


Figure 2: Amplified images of the tonal values for the syllable "y"

In the case of the syllable *vien-*, we cannot take the central value (which would be only 183 Hz), as the syllable is characterized by a tonal inflection and the melody reaches even 212 Hz at its highest point, so we must measure this extreme pitch value instead of the central one (cf. Fig. 3).

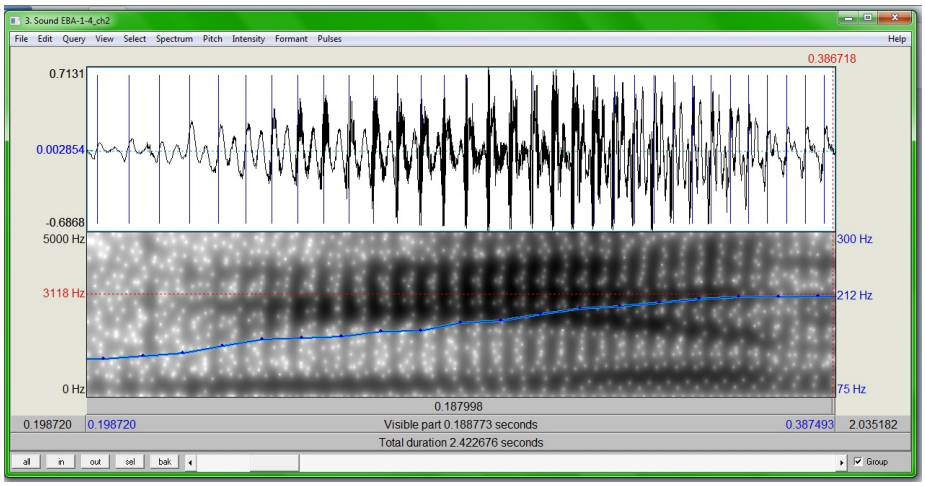


Figure 3: Amplified image of the tonal values for the syllable *vien-*

Bearing in mind this principle, we can display all the absolute f0 values measured for each syllable, and this is how we get the melody of the utterance reduced to only the relevant tonal information (Fig. 4). In case of the syllable *pa-*, as it is characterized by tonal instability superior to 10%, both extreme

values are represented in the curve (in my representation, the point before the vowel a indicates inner inflection in the syllable).

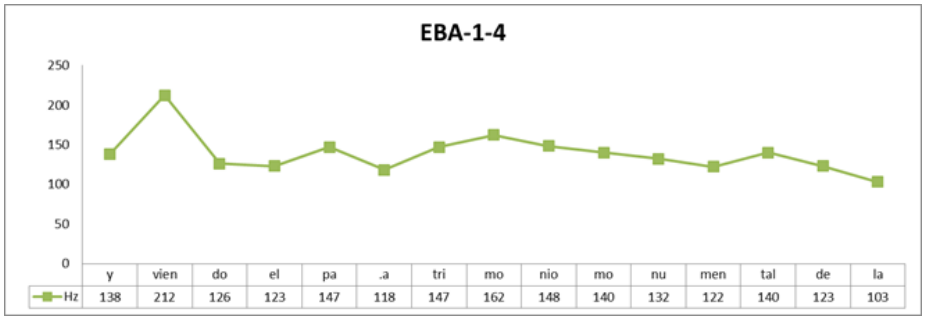


Figure 4: Absolute curve of the utterance "*Y viendo el patrimonio monumental de la...*"

After this phase, we proceed to the standardization. Each absolute value (measured in Hertz) becomes a relative value, depending on the previous value: the first value of the utterance is given an arbitrary value '100', and the following values represent the tonal distance measured in % with respect to the previous syllable. For example, a jump from 138 Hz to 212 Hz in the following syllable would result in the values of 100 and 154 respectively, since between 138 and 212 there is a rise of 52,52%. In Figure 5, we show how the absolute values obtained in Hertz (green line) are converted into relative values (blue line).

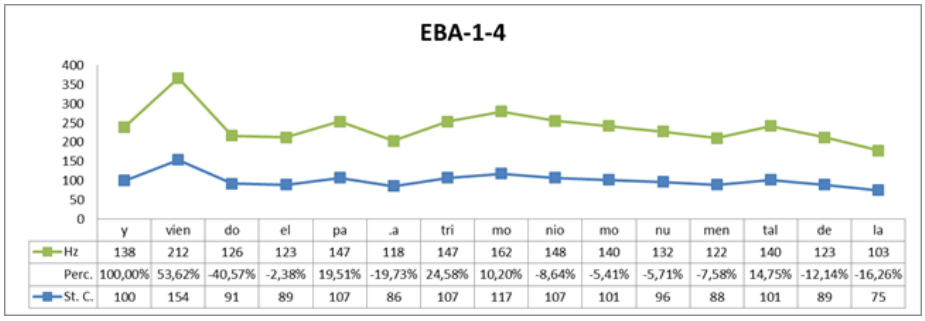


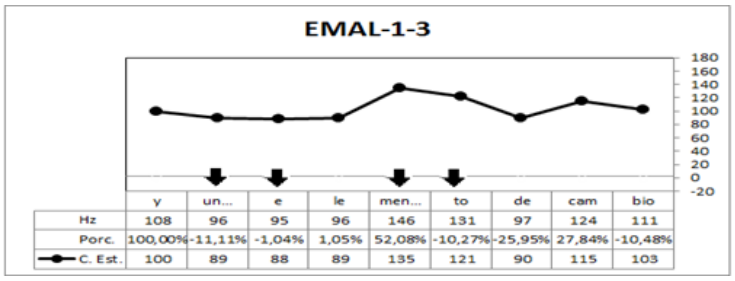
Figure 5: Standardized curve of the utterance "*Y viendo el patrimonio monumental de la...*"

The standardized curve thus ensures that the described melodies are objectively comparable to each other, regardless of the individual tonal characteristics

of the speakers (e.g. if it is a child with a tonal height much higher than in the case of a man; what would matter are the proportions of the tonal movements and not the absolute values of each curve). Both curves (the absolute one and the standardized copy) are melodically identical, though in order to validate whether the standardized copy sounds the same as the original, it can be synthesized in Praat and a series of perceptive tests can be applied. First, all f0 values are deleted and then replaced by the values of the standardized copy by using the function "Manipulate" in Praat. Both samples (the original and the synthesized) are submitted to the listeners' judgment. If correction is needed, it can be realized as a final phase (Cantero & Font-Rotchés, 2020, p. 34–35).

3.3. The use of melodic data in the present research

In accordance with my objectives, segments affected by lengthening will be examined from a tonal point of view, assuming that they are tonally irrelevant in their context (i.e. they present no prominent melodic changes compared to adjacent segments), and also assuming no significant differences between the two dialectal zones. In order to study their tonal behavior, two melodic data will be examined: the percentage of tonal movement **to** the lengthened segment from the previous one, and the percentage of the tonal movement **from** the lengthened one to the following. These data are indicated by arrows in the following plot (Figure 6).



	y	un...	e	le	men...	to	de	cam	bio
Hz	108	96	95	96	146	131	97	124	111
Porc.	100,00%	-11,11%	-1,04%	1,05%	52,08%	-10,27%	-25,95%	27,84%	-10,48%
C. Est.	100	89	88	89	135	121	90	115	103

Figure 6: Graph of an utterance from Málaga "*y un elemento de cambio*" 'and an element of change', with *un...* and *–men...-* as the lengthened segments. Arrows indicate the tonal movements to and from the lengthened segments (standardized melodic representation).

As we are analyzing relative prosodic values, we cannot take into consideration utterance-initial lengthenings when we measure the proportion of tonal movements to the lengthened segment, as these values cannot be contrasted with any previous value, so they cannot even be relativized. Similarly, we cannot analyze utterance-final hesitation phenomena from the point of view of the proportion of the tonal movement from the lengthened segment either, as there is no segment following them and thus no comparison can be made.

4. Results

In the following section we will focus on the analysis of the received data in order to validate our hypotheses, i.e. (1) lengthenings do not present different melodic behavior in the two examined dialectal zones; and (2) lengthenings are not characterized by prominent melodic movements compared to their adjacent context.

As in this study the focus is on lengthenings as disfluency phenomena, first we must define which segments were considered as part of this group.

The first criterion to fulfil was, obviously, that the segment should be prolonged. There are several ways to determine whether a segment should be considered lengthened. Lengthening is easily detectable by listeners by ear, so Deme & Markó (2013) based their judgement on perceptive tests: if a segment was recognized as lengthened by 6 out of 10 listeners (all linguists), it was considered lengthened. My choice was to base this judgement on durational data: as the minimum duration of filled pauses is considered to be 0,2s by Goldman-Eisler (1973) and Guaitella (1996), cited by (Blondet, 2001, p. 8), and filled pauses are generally longer in Spanish than lexical vowel lengthenings (Villa Villa, 2017, p. 167), I took 0,2s as the minimum value for lengthenings. The other criterion was, in case of non-initial segments, that it should be longer than the previous one.

As the study focuses on lengthenings as disfluency phenomena, we must also contemplate two cases in which lengthening is a natural by-product of certain

conditions but does not serve as a tool to gain time, thus, is definitely not a disfluency phenomenon, and exclude those samples from the analysis. First, we must bear in mind that phrase-final lengthening is a natural process in numerous languages, including Spanish (Gósy & Krepsz, 2018). Second, stressed position can also result in longer duration in Spanish (Ortega-Llebaria, 2006), so stressed syllables were also excluded from our analysis (though lengthenings in Spanish, as we have already seen, mostly affect unstressed segments anyway, cf. Rebollo Couto, 1997).

Based on what has been said, if a segment sounded prolonged but its duration was inferior to 0,2s, it was sentence-final or stressed, it was automatically excluded from the analysis. This can be seen in Table 2, which sums up the tonal movements related to the lengthenings attested in both corpora: in the northern corpus 116 cases were detected as compared to the 120 cases in the southern corpus. However, in the analysis only those were taken into consideration which were not utterance-final and were unstressed, so this number was reduced to 65 northern cases and 50 southern ones in case of the analysis of the tonal movement to the lengthened segment, as here utterance-initial segments could not have been analysed (there was not any tonal movement to an utterance-initial segment). In case of the analysis of the tonal movement from the lengthened segment, utterance-initial unstressed lengthenings were already taken into consideration, so we had 81 northern samples and 71 southern ones.

Regarding the melodic aspect, the average value of the proportion of the movement **to** the lengthened syllable, in case of the northern dialects, is $-1,56\%$, while in case of the southern ones, it is 0,85% (cf. Figure 7), with no significant difference between the means.

In the case of the proportion of the tonal movements **from** the lengthened syllable, the mean value is $-1,67\%$ in the case of the northern dialects and $-6,73\%$ in the case of the southern ones (cf. Figure 7), again without a significant difference between the means. This implies that prediction (1) about the same tonal behavior of the two examined dialects as far as lengthenings are concerned has proved to be true according to my corpora.

Table 2: Data related to the tonal movements of lengthenings

	North	South
lengthenings (total)	116	120
no of non-initial non-final lengthenings on unstressed syllables	65	50
lowest tonal movement to the syllable (%)	-54,46	-40,57
highest tonal movement to the syllable (%)	51	132
mean tonal movement to the syllable (%)	-1,56	0,85
rising tonal movement to the syllable (cases)	25 out of 65 (38%)	20 out of 50 (40%)
rising tonal movement superior to 10% to the syllable (cases)	8 out of 65 (12%)	9 out of 50 (18%)
no of non-final lengthenings on unstressed syllables	81	71
lowest tonal movement from the syllable (%)	-40,52	-56,25
highest tonal movement from the syllable (%)	52,58	68,31
mean tonal movement from the syllable (%)	-1,67	-6,73
rising tonal movement from the syllable (cases)	33 out of 81 (41%)	22 out of 71 (31%)
rising tonal movement superior to 10% from the syllable (cases)	12 out of 81 (15%)	11 out of 71 (15%)

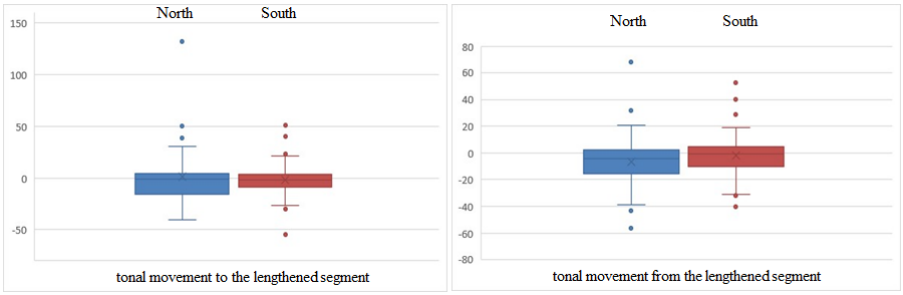


Figure 7: Boxplots of the tonal movement to and from the lengthened syllable in % in the two corpora (boxplots are generated by Excel 365 pro plus).

In order to examine hypothesis (2), we should analyse the average values of tonal movements associated to the prolonged segments. A salient tonal movement in Spanish is perceived if the listener is exposed to at least 10% of melodic variation between segments (Font-Rotchés & Mateo Ruiz, 2011). In this case, the average value of neither the tonal movement to the lengthened segment nor the one from the lengthened segment reaches this threshold in either of the two dialects. Rises – especially over the perception threshold of 10% - occur in less than 20% in both corpora. This means that based on our corpus, we can conclude that the average tonal movements associated to the prolonged segment – the movement from the previous segment to the prolonged one and the movement from the prolonged segment to the next one – cannot be considered as salient melodic differences with respect to their adjacent contexts.

5. Conclusions and discussion

In this study the two main dialectal areas of European Spanish were examined from the point of view of the melodic behavior of lengthenings. A corpus of 200 utterances (100 northern Spanish and 100 southern Spanish ones, taken from Map Task activities and spontaneous interviews downloaded from YouTube videos) were contrasted, in order to verify the following two hypotheses:

1. utterance-internal lengthenings do not present different intonational behavior in the two examined variants
2. utterance-internal lengthenings are not characterized by prominent melodic movements compared to their adjacent context

The first prediction is part of a more complex hypothesis set up by Baditzné Pálvölgyi (2020), assuming that as for prosody, southern and northern variants will only be different in relative duration values, but not in relative intensity or intonation values. This assumption is partly based on the observation that so far we have not discovered radical differences in the intonational characteristics between southern and northern variants of Spanish at segment level (and lengthenings affect segments), especially taking into consideration

unstressed syllables, and lengthenings in Spanish typically affect unstressed syllables (Rebollo Couto, 1997). As for duration, however, there is an argument to suppose that southern dialects present relatively longer hesitation phenomena in utterance-internal segments than their northern counterparts. This could be so because the southern variants are characterized by elision more frequently than the northern ones, thus, even if segments affected by hesitation are of the same absolute duration in both dialects, they are perceived relatively longer in the southern dialects where syllables are realized shorter due to elision than in the northern dialectal zone (Toledo, 2010).

We have seen based on the results that effectively, in case of relative tonal values from the previous segment to the one affected by lengthenings, the average percentage of the tonal movement was not significantly different in case of the two examined corpora. This was also true for the average percentage of tonal movements from the lengthened syllable to the following one. This means that the first hypothesis was validated. As for our second hypothesis, both corpora presented very low means as for intonational relevance of the lengthened segment compared to its context, not reaching 10% of tonal difference, which is considered a perception threshold in Spanish. As has been predicted, speakers did not realize lengthenings accompanied by striking melodic movements. It may be explained by the wish of the speakers not to interrupt melodically the utterance and provide this way tonal continuity, in order to maintain the conversational turn.

As for future research, more data should be collected in order to support our results, and other disfluency phenomena could be analysed. Also, as in this paper we have only seen the melodic aspect of lengthenings in European Spanish dialects, but there are other prosodic components to be analysed, intensity and duration. By the help of Cantero's (2019) extended Prosodic Analysis of Speech (PAS) model we could define prosodic features other than melody in case of hesitation phenomena. The PAS model offers a standardization protocol for intensity and duration similar to the procedure we have seen in the case of intonation; as for intensity, the standardized intensity curve is generated by

reducing every syllable to its prominent intensity peak value, and these values are standardized in terms of proportion always compared to the previous value. Duration standardization is a more complex process, the distance between intensity peaks is calculated for each segment, and thus the standardized curve is generated over these values. The perception threshold is not yet established unanimously for either intensity nor duration in case of Spanish, and without these values we cannot fully interpret the results.

Acknowledgement

Supported by the ÚNKP-19-4 New National Excellence Program of the Ministry for Innovation and Technology and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

Adriaens, L. M. H. (1991). *Ein Modell deutscher Intonation. Eine experimentell-phonetische Untersuchung nach den perzeptiv relevanten Grundfrequenzänderungen in vorgelesenem Text*. Ph.D. thesis Technological University of Eindhoven.

Baditzné Pálvölgyi, K. (2011). The intonational patterns used in Hungarian students' Spanish yes-no questions. *Phonica*, *7*, 80–99. URL: `http://www.publicacions.ub.edu/revistes/phonica7/`. Last accessed 12/02/2020.

Baditzné Pálvölgyi, K. (2012). Spanish intonation of Hungarian learners of Spanish: yes-or no questions. Thesis Doctoral. *Biblioteca Phonica*, *15*. URL: `http://www.publicacions.ub.edu/revistes/phonica-biblioteca/`. Last accessed 12/02/2020.

Baditzné Pálvölgyi, K. (2018). La presencia de rasgos prelingüísticos en la entonación de la interlengua húngaro-española. *Colindancias: Revista de la Red de Hispanistas de Europa*, (p. 237–248).

Baditzné Pálvölgyi, K. (2018). La presencia de rasgos prelingüísticos en la entonación de la interlengua húngaro-española. *Colindancias: Revista de la Red de Hispanistas de Europa*, (p. 237–248).

Baditzné Pálvölgyi, K. (2019). A hezitálás mintázatai küszöbszinten álló magyar ajkú nyelvtanulók spontán beszédprodukciójában. *Alkalmazott Nyelvtudomány*, *19*, 1–14.

Baditzné Pálvölgyi, K. (2020). Hesitation patterns in Northern and Southern dialects of European Spanish. In *Presentation at the 17th Old World Conference on Phonology, 5-7th of February, 2020*. University of Warsaw.

Beaugendre, F. (1994). *Une étude perceptive de l'intonation du français*. Phd thesis, University of Paris XI, Orsay, France.

Berry, A. (1994). Spanish and American turn-taking styles. *Pragmatic and Language Learning. Monograph Series*, *5*, 180–190.

Blondet, M. A. (2001). Las pausas llenas: marcas de duda e identidad lingüística. *Lingua Americana*, *V*, 5–15.

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer. URL: `http://www.praat.org/` accessed 22 March 2019.

Cantero, F. J., & Font-Rotchés, D. (2020). Melodic analysis of speech (mas). phonetics of intonation. In J. Abasolo, I. Pablo, & A. Ensunza (Eds.), *Contributions on education* (p. 20–47). Universidad del País Vasco.20-47.

Cantero Serena, F. J. (2002). *Teoría y análisis de la entonación*. Barcelona: Ed. Universitat de Barcelona.

Cantero Serena, F. J. (2019). Análisis prosódico del habla: más allá de la melodía. In M. Álvarez Silva, A. Alvarado, & L. Miyares (Eds.), *Comunicación Social: Lingüística, Medios Masivos, Arte, Etnología, Folclor y otras ciencias afines. Volumen II* (p. 485–498). Santiago de Cuba: Ediciones Centro de Lingüística Aplicada.

Cantero Serena, F. J., Alfonso, R., Bartolí, M., Corrales, A., & Vidal, M. (2005). Rasgos melódicos de énfasis en español. laboratori de fonètica aplicada – lfa. *Phonica*, *1*. URL: `http://www.publicacions.ub.edu/revistes/phonica1/PDF/articulo_03.pdf`. Accessed 12/02/2020.

Cantero Serena, F. J., & Font-Rotchés, D. (2007). Entonación del español peninsular en habla espontánea: patrones melódicos y márgenes de dispersion. *Moenia*, *13*, 69–92.

Cantero Serena, F. J., & Font-Rotchés, D. (2009). Protocolo para el análisis melódico del habla. *Estudios de Fonética Experimental*, *18*, 17–32.

Deme, A., & Markó, A. (2013). Lengthenings and filled pauses in Hungarian adults' and children's speech. In *Eklund, R. (ed.) Proceedings of DiSS 2013, The 6th Workshop on Disfluency in Spontaneous Speech KTH Royal Institute of Technology* (p. 21–24). Stockholm: Sweden: Department of Speech Communication and Music Acoustics, Royal Institute of Technology.

Devís, E. (2011). La entonación del español hablado por italianos. *Didáctica (Lengua y Literatura)*, *23*, 35–58.

Di Cristo, A. (1982). *Prolegomènes à l'étude de l'intonation. Micromelodie*. Paris: CNRS.

Eklund, R. (2004). *Disfluency in Swedish human–human and human–machine travel booking dialogues*. Phd thesis. Linköping Studies in Science and Technology.

Estruch, M., Garrido, J. M., Gudayol, F., Jiménez, J. M., & Riera, M. (1999). Validación perceptiva de un sistema de estilización automática de contornos melódicos. In U. R. i Virgili (Ed.), *Actas del 'I Congrés de Fonètica Experimental', Universitat Rovira i Virgili, Tarragona, 22-24 febrero 1999* (pp. 217–223).

Fonseca, A., & Cantero Serena, F. J. (2011). Características da entonação do espanhol falado por brasileiros. In *Anais do VII Congresso International*

Abralin. Ed. Abralin (p. 84–98). Curitiba (Brasil: Associação Brasileira de Lingüística.

Font-Rotchés, D. (2005). *L'entonació del català. Patrons melòdics, tonemes i marges de dispersió*. Ph.D. thesis Laboratori de Fonètica Aplicada, Universitat de Barcelona.

Font-Rotchés, D. (2007). *L'entonació del català. Biblioteca Milà i Fontanals 53*. Barcelona: Publicacions de l'Abadia de Montserrat.

Font-Rotchés, D. (2009). Les interrogatives pronominals del català central. anàlisi melòdica i patrons entonatius. *Els Marges. Revista de llengua i literatura*, *87*, 41–64.

Font-Rotchés, D., & Mateo Ruiz, M. (2011). Absolute interrogatives in spanish: a new melodic pattern. In *Actas do VII congresso internacional da ABRALIN* (p. 1111–1125). Curitiba (Brasil.

Garrido, J. M. (1991). *Modelización de patrones melódicos del español para la síntesis y el reconocimiento*. Bellaterra: Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.

Garrido, J. M. (1996). *Modelling Spanish Intonation for Text-to-Speech Applications*. Phd dissertation. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.

Garrido Almiñana, J. M., Laplaza, Y., & García, C. L. (2017). La caracterización pragmática y prosódica de la vocalización "mmm" en español. In V. Marrero Aguiar, & E. Estebas Vilaplana (Eds.), *Eds.): Tendencias actuales en fonética experimental: Cruce de disciplinas en el centenario del Manual de Pronunciación Española (Tomás Navarro Tomás* (p. 125–129). Madrid: Uned.

Gili Gaya, S. (1924). Influencia del acento y de las consonantes en las curvas de entonación. *Revista de Filología Española*, *11*, 154–177.

Goldman-Eisler, F. (1973). *Pschycolinguistics. Experiments in Spontaneus Speech*. New York: Academic Press.

Gósy, M. (2002). A megakadásjelenségek eredete a spontán beszéd tervezési folyamatában. [The origin of disfluency phenomena in the spontaneous speech planning process]. *Magyar Nyelvőr*, *126*, 192–204.

Gósy, M., & Krepsz, V. (2018). Phrase-final Lengthening of Phonemically Short and Long Vowels in Hungarian Speech across Ages. In M. Gósy, & T. E. Gráczi (Eds.), *Challenges in analysis and processing of spontaneous speech* (p. 99–126). Research Institute for Linguistics, Hungarian Academy of Sciences.

Guaitella, I. (1996). Analyse prosodique des hésitations vocales: propositions pour un modèle rythmique. *R. P. A*, *118-119*, 113–145.

Hombert, J. M. (1978). Consonantal types, vowel quality and tone. In V. Fromkin (Ed.), *Tone: a linguistic survey.*. New York: Academic Press.

Hualde, J. I. (2014). *Los sonidos del español*. CUP.

Kao, W. (2011). *La entonación de enunciados declarativos e interrogativos en chino mandarín hablado por taiwaneses. Trabajo de investigación final de Máster*. Laboratori de Fonètica Aplicada de la UB.

Lickley, R. J. (1994). *Detecting disfluency in spontaneous speech*. Phd dissertation, University of Edinburgh.

Lickley, R. J. (2015). Fluency and disfluency. In M. Redford (Ed.), *The handbook of Speech production* (p. 445–469). John Wiley Blackwell.

Machuca Ayuso, M. J. (2018). Pausas sonoras y bilingüismo. *Estudios de Fonética Experimental*, *XXVII*, 75–95.

Machuca Ayuso, M. J., Llisterri, J., & Ríos, A. (2015). Las pausas sonoras y los alargamientos en español. *Un estudio preliminar. Revista Normas*, *5*, 81–96.

Maclay, H., & Osgood, C. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*, 19–44.

Martorell, L. (2011). *Les interrogatives absolutes de l'espanyol parlat pels suecs. Trabajo Final de Máster*. Barcelona: Universitat de Barcelona, Facultad de Formación del Profesorado.

Mateo, A. (1988). Experimento sobre el tono intrínseco de las vocales castellanas. *Estudios de Fonética Experimental*, *3*, 157–179.

Navarro Tomás, T. (1966). *Manual de entonación española*. La Habana: Edición revolucionaria.

Odé, C., & Heuven, V. J. (1994). *Experimental studies of Indonesian prosody*. Dep. of Languages and Cultures of Southeast Asia and Oceania, University of Leiden.

Olaszy, G., & Koutny, I. (2001). Intonation of Hungarian Questions and their prediction from text. In S. Puppel, & G. Demenko (Eds.), *Prosody 2000, Speech recognition and synthesis* (p. 179–196).

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

Patiño, E. (2008). *Prosodic Comparative Study of Mexico City and Madrid Spanish*. Freie Universität Berlin, Germany. Escuela Nacional de Antropología e Historia, Mexico.

Prieto, P., Borràs-Comes, J., & Roseano, P. (2010). Interactive atlas of romance intonation. URL: `http://prosodia.upf.edu/iari/`.

Quilis, A. (1981). *Fonética acústica de la lengua española*. Madrid: Gredos.

Quilis, A. (1999). *Tratado de fonética y fonología españolas*. Madrid: Gredos.

Rebollo Couto, L. (1997). Pausas y ritmo en la lengua oral. Didáctica de la pronunciación. In *ASELE Actas VIII* (p. 667–676). Centro Virtual Cervantes.

Rodríguez et al. (2001). Annotation and analysis of disfluencies in a spontaneous speech corpus in Spanish. In *ITRW on Disfluency in Spontaneous Speech (Diss '01* (p. 1–4). Edinburgh, Scotland, UK. Last accessed 12/05/2020.

Rodríguez et al. (2015). Las pausas en el discurso de individuos con demencia tipo Alzheimer. Estudio de casos. *Revista de Investigación en Logopedia*, *1*, 40–59.

Shriberg, E. (1994). *Preliminaries to the theory of speech disfluencies*. Phd dissertation University of California Berkeley.

Sosa, J. M. (1999). *Entonación española. Su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra.

Toledo, G. (2010). Métricas rítmicas en tres dialectos Amper-España [Rhythmic metrics in three dialects of Amper-Spain]. *Estudios Filológicos*, *45*, 93–110. URL: `https://scielo.conicyt.cl/scielo.php?pid=S0071-17132010000100008&script=sci_arttext`. Last accessed 12/05/2020.

t'Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Villa Villa, J. et. al. (2017). Las vocales de relleno en español: nuevos datos y algunas reflexiones. In R. Minares (Ed.), *Nuevos estudios sobre Comunicación Social* (p. 165–169). Santiago de Cuba: Centro de Linguistica Aplicada volume I.

A prozódiailag jelölt fókusz azonosításának elsajátítása

Surányi Balázs[1,2], Pintér Lilla[2,1]

[1] *Nyelvtudományi Intézet*
[2] *Pázmány Péter Katolikus Egyetem*

Abstract

The goal of the present study is to explore whether and how the development of the comprehension of prosodic focus-marking may be affected by the variation found in the marking of focus across different languages. We investigate focus-identification in Hungarian, a language that not only has prosodic focus-marking, but mandatorily uses syntactic focus-marking as well. In pursuit of comparability, the experiment this paper reports on employed a task that was recently applied by Szendrői et al. (2018) in a study of English, German, and French pre-school children. Our hypothesis was that the systematic syntactic marking of focus in Hungarian diminishes the disambiguating role of prosodic marking for the child. Therefore we expected that in sentences in which syntactic focus-marking fails to unambiguously identify the focus, the comprehension of prosodic focus-marking will be delayed in comparison to the languages investigated in Szendrői et al., in which syntactic focus-marking is at best only an option. This prediction was borne out by our data: at age four Hungarian children give congruent responses to sentences containing non-default narrow (subject) focus roughly half as often as their English, German and French peers, who already exhibit adult-like performance at this age. By contrast, Hungarian children reach the adult-like level only at age seven. Thus, the paper makes a strong case that, similarly to the acquisition of the production of prosodic focus-marking, the developmental trajectory of the comprehension of prosodic focus-marking is also robustly affected by the cross-linguistic variation found in the marking of focus.

1. Bevezetés

A fókusz az egyik legfőbb olyan információszerkezeti szerep, amely rendszerszerűen – és nagyrészt grammatikalizálódott módon – összekapcsolja a kommunikációs kontextust a mondatprozódiával. A prozódiai fókuszjelölés feldolgozásának elsajátítását több szerző is elhúzódó fejlődési folyamatként jellemezte, különösen a produkcióhoz viszonyítva (bővebben lásd Hornby, 1971; Crutten-

Email addresses: suranyi@nytud.hu (Surányi Balázs), pinter.lilla@btk.ppke.hu (Pintér Lilla)

Beszédtudomány – Speech Science 2020. 124–151. DOI: 10.15775/Besztud.2020.124-151

den, 1985; Wells et al., 2004) – annak ellenére is, hogy közismert a gyerekek korai érzékenysége a prozódiára.

A kép ugyanakkor messze nem egységes: a szakirodalom alapos vizsgálata rávilágított arra, hogy a gyerekek teljesítménye a megértést vizsgáló kísérletekben nagyfokú változatosságot mutat a tesztelés során alkalmazott módszertan, illetve feladat függvényében. Néhány, a közelmúltban megjelent tanulmány eredményei arra utalnak, hogy ha a kísérleti feladat kellőképp egyszerű, akkor a fókusz helyes azonosítása a prozódiai jelölés alapján már korai életkorban tetten érhető (Höhle et al., 2009; Speer & Ito, 2009; Sekerina & Trueswell, 2012; Szendrői et al., 2018). A különféle kimenetelek egy másik lehetséges forrását jelenthetik az egyes munkákban vizsgált nyelvek fókuszjelölési rendszerei között megfigyelhető eltérések. Míg a fókusz produkciójának kutatásában hangsúlyos szerepet kapott ez a tényező (egy újabb áttekintést nyújt például Chen, 2018), addig a fókusz megértésének tanulmányozásában ezt jobbára figyelmen kívül hagyták.

Jelen kutatás célja annak a feltérképezése, hogy a prozódiai fókuszjelölés megértésének fejlődését befolyásolja-e – és ha igen, miként – a lehetséges fókuszjelölési módok változatossága a különféle nyelvekben. A fókuszazonosítást éppen ezért a magyarban vizsgáltuk, ahol a prozódiai jelölés mellett kötelező a szintaktikai fókuszjelölés is. A lehető legpontosabb összehasonlíthatóság érdekében az általunk végzett kísérletben ugyanazt a feladatot alkalmaztuk, mint nemrégiben publikált, angol, francia és német óvodásokat tanulmányozó munkájában Szendrői et al. (2018). Hipotézisünk szerint a szisztematikus szintaktikai fókuszjelölés a magyarban csökkentheti a gyermekek számára a prozódiai jelölés egyértelműsítő szerepét, éppen ezért azt jósoljuk, hogy azokban a mondatokban, amelyekben a szintaktikai fókuszjelölés segítségével nem azonosítható egyértelműen a fókusz, a pusztán prozódiai fókuszjelölés megértése késést fog mutatni azokhoz a nyelvekhez képest, amelyeket Szendrői et al. (2018) teszteltek, és amelyekben a szintaktikai fókuszjelölés legfeljebb csak opcionálisan van jelen.

A tanulmány felépítése a következő: a második fejezetben áttekintünk néhány kiemelkedő jelentőségű kísérleti eljárást a prozódiai fókuszjelölés megér-

tésének, illetve produkciójának elsajátítására vonatkozóan, különös tekintettel az előbbire. Ez a rövid szemle elsősorban azokra a megállapításokra összpontosít, amelyek ahhoz a konklúzióhoz vezettek, hogy a prozódiai fókuszjelölés felnőttszerű értelmezése egy jelentősen elhúzódó nyelvelsajátítási folyamat terméke, miközben bemutatjuk azokat a kísérleti eredményeket is, amelyek kétségbe vonták ennek az általánosításnak a létjogosultságát. A harmadik fejezetben részletesen kifejtjük az arra vonatkozó kutatási kérdést, hogy milyen mértékben járulhatnak hozzá a korábbi eredmények esetében megfigyelhető eltérésekhez az eddig vizsgált nyelvek közötti, a fókuszjelölés nyelvtanát érintő különbségek. Ugyanebben a részben röviden ismertetjük az általunk tanulmányozott magyar nyelv fókuszjelölési rendszerének legfőbb sajátosságait, majd bemutatjuk az elvégzett kísérletet és megvitatjuk annak eredményeit. Végül a negyedik fejezetben összefoglaljuk a levonható következtetéseket.

2. Háttér

Az vitán felül áll, hogy a gyerekek már nagyon korai életkorban megtanulnak olyan megnyilatkozásokat tenni, amelyek az elsajátítandó anyanyelvük prozódiai mintázatához alkalmazkodnak, méghozzá általában hamarabb, minthogy az adott nyelv szintaxisának jelentős részét megtanulnák (Lieberman, 1967; Menyuk, 1969; Bloom, 1970; Brown, 1973). Ez összefügg azzal is, hogy a gyerekeknek már a születésüket követő első hónaptól kezdve megvan a képességük arra, hogy észleljenek olyan prozódiai információkat, mint a hangmagasság, a szóhangsúly vagy a prozódiai tagolás (Sansavini et al., 1997; Schmitz et al., 2006; Höhle et al., 2009; Wellmann et al., 2012; Gervain & Werker, 2013).

Ezzel összhangban bizonyos (elsősorban germán és újlatin) nyelvekben, ahol a prozódiai fókuszjelölés elsajátítását kísérletesen vizsgálták, az derült ki, hogy a gyerekek produkciója már igen korán sok szempontból felnőttszerű (Hornby, 1971; Wieman, 1976; Schmitz et al., 2006; Sauermann et al., 2011; Yang & Chen, 2014).

Az ehhez hasonló, produkcióra vonatkozó megállapítások ellentétben állnak az információszerkezet prozódiai jelölésének értelmezésére vonatkozó eredményekkel. Hornby (1971) és Cruttenden (1985) képkiválasztási feladata, amellyel hat- és tízéves kor közötti angol anyanyelvű gyerekeket teszteltek, kimutatta, hogy még a tízévesek sem tudják hasznosítani a hangsúlymintázatokat a kontrasztív információnak, illetve a mondat topik–komment tagolásának azonosítására. Cutler & Swinney (1987) szintén nem talált a hatévesnél fiatalabb angol kisgyerekeknél a felnőttekéhez hasonló előnyt a mondat hangsúlyos szavainak felismerésében (azok hangsúlytalan előfordulásaihoz viszonyítva). Wells et al. (2004) öt-, illetve tizenhárom éves – ugyancsak angol anyanyelvű – gyerekeket vizsgáltak, és azt találták, hogy a korrektív fókuszt is tartalmazó mondatokra épülő produkciós feladatban már az ötévesek is hangsúlyozták a fókuszált összetevőket a megnyilatkozásaik zömében (a tizenhárom évesekétől nem különböző mértékben). Ezzel szemben ugyanezen feladat receptív részében, amely a hangsúlyos elem által jelölt referens azonosítását kívánta meg egy képre való rámutatás útján, az ötévesek teljesítménye nem haladta meg a véletlenszerű szintet (miközben az idősebb korcsoport egyértelműen fejlődést mutatott). A mandarin kínai prozódiai fókuszjelölés megértését vizsgálva Chen et al. (2019) azt találták, hogy a három- és ötéves kor közötti gyerekek a felnőttekhez képest szignifikánsan kevésbé szisztematikusan javították kongruens módon az alanyi fókuszt tartalmazó mondatokat. A szerzők ezt azzal magyarázták, hogy a prozódiai fókuszjelölés alkalmazása a mandarinban meglehetősen korlátozott, amit érzékletesen illusztrál az is, hogy még a felnőtt anyanyelvi beszélők is csupán az esetek 38%-ában adtak kongruens alanyifókusz-javításokat. Különös jelentőséggel bírhat az is, hogy egy tonális nyelvről van szó, amely sajátosság a gyerekek számára kiemelkedően magas funkcionális terhelést róhat a fókusz olyan fonetikai indikátoraira, mint például a hangmagasság. Ugyanakkor a prozódiai fókuszjelölés elsajátítására vonatkozó kísérletek eredményei a mandarinban vitatottak: egy az említetthez nagyon hasonló kísérletben (Chen, 1998) a részt vevő gyerekek körülbelül 65%-os pontossággal javították a fókuszált alanyokat kongruens módon. Végül azt is érdemes kiemelni, hogy a gyerekek tipikusan

iskoláskorig gyenge teljesítményt mutatnak az olyan mondatok értelmezésében, ahol a prozódiai fókuszt a *csak* vagy az *is* fókuszpartikulákkal kellene asszociálniuk (Gualmini et al., 2003; Hüttner et al., 2004; Bergsma, 2006; Costa & Szendrői, 2006; Zhou et al., 2012).

A fent említett kutatások alapján tehát az a kép rajzolódik ki, hogy a fókuszprozódia felnőttekéhez közelítő produkciója megelőzi annak felnőttszerű értelmezését, ami épp a fordítottja a megszokott mintázatnak. Noha az ilyen típusú megértés–produkció aszimmetria váratlannak tűnhet, korántsem példa nélküli (lásd például Chien & Wexler, 1990). A lehetséges magyarázatok között szerepel a megértésbeli elmaradás kísérleti műtermékként vagy a tesztfeladat által kiváltott hatásként való értelmezése, de többen tartják kognitív vagy pragmatikai korlátok eredményének, a nyelvtan optimalitáselméleti keretében az ellentétes irányú optimalizáció következményének, vagy akár ezek különféle kombinációinak (lásd különösképpen Hendriks & Koster, 2010).

Ami kifejezetten a fókusz prozódiáját illeti, több tanulmányban is meggyőzően érveltek amellett, hogy a megértésbeli késés az esetek legalábbis egy részében nem támasztható alá egyértelműen. Berger & Höhle (2012) például kimutatták, hogy a német anyanyelvű három- és négyévesek jelentősen jobban teljesítenek a *csak*-kal és *is*-sel módosított kifejezéseket tartalmazó mondatok esetén, amennyiben a kísérleti módszer – a legtöbb korábbi munkával ellentétben – kiemelkedően fontossá teszi a partikulához társított információt a feladat teljesítésének vonatkozásában. A gyerekek fókuszjelölés-feldolgozásának felmérésére használt igazságérték-megítélési feladatokkal kapcsolatban általában is gyakran felvetődik az a vád, mely szerint nem kellőképpen világos, hogy a gyermek a tesztmondatokat az alapján a pragmatikai komponens alapján ítéli-e meg, amellyel a fókuszjelölés hozzájárul a mondatok jelentéséhez, vagy pedig kizárólag azok szemantikai jelentése alapján. Az sem egyértelmű továbbá, hogy vannak-e különbségek a gyerekek között annak tekintetében, hogy e kettő közül melyiket tartják feladatuknak az adott kísérletben (Gualmini et al., 2001; Papafragou & Musolino, 2003).

Az online kísérletek, különösen a szemmozgáskövetéses vizsgálatok, amelyek egyáltalán nem igényelnek explicit ítéleteket, mentesek a fent említett lehetséges problémáktól. Zhou et al. (2012) vizuális világ paradigmát alkalmazó munkája például valóban azt mutatja, hogy a mandarin kínai gyerekek a *csak* partikulát tartalmazó mondatok értelmezésekor felnőttszerűen viselkednek, miközben az explicit ítéletet váró feladatokban a mintázat továbbra sem hasonlít a felnőttekéhez. Egy másik, ugyancsak szemmozgáskövetéses, implicit feladatot tartalmazó kísérlet hasonló eredményeket hozott: Höhle et al. (2009) kimutatták, hogy a három- és négyéves német gyerekek igenis tudják hasznosítani a hangsúlyjelölést a fókusz azonosításához az *is* fókuszpartikulát tartalmazó mondatok esetében. Sekerina & Trueswell (2012) pedig arra találtak bizonyítékot, hogy az orosz anyanyelvű öt-, illetve hatéves gyerekek képesek arra használni a hangsúlyt, hogy a felnőttekhez hasonlóan kontrasztív értelmezést tulajdonítsanak olyan főnévi kifejezéseknek, amelyekben vagy a melléknév vagy a főnév prominens prozódiailag.

Visszatérve az offline megítélési feladatokhoz, a *csak* és *is* partikulákhoz hasonló fókuszoperátorokat tartalmazó tesztmondatok kiváltképp érintettek a fent említett módszertani kihívások kérdésében. Ezeknek az állításoknak az esetében ugyanis a gyerekeknek nem csupán a fókuszt kell azonosítaniuk annak prozódiai jelölése alapján, hanem hozzá is kell kapcsolniuk a fókuszt a fókuszoperátorhoz (amely akár nagyobb távolságra is lehet tőle), majd pedig kiszámítani azt a szemantikai többletjelentést, amelyet az hordoz (például a *csak* esetében az általa kifejezett kimerítőséget). Éppen ezért – amint arra Szendrői et al. (2018) is kitérnek – az ilyen feladatokban a gyerekek ítéletei nem mutatják közvetlenül a fókusz azonosítására, illetve alapjelentésének feldolgozására vonatkozó kompetenciájukat, hanem sokkal inkább azt tükrözik, hogy mindezen felül képesek-e további szintaktikai és szemantikai műveletek elvégzésére.

A fent tárgyalt okok miatt Szendrői et al. (2018) bevezettek egy olyan kísérleti feladatot, amelyben semmilyen további szemantikai operátor nem lép működésbe, sőt a résztvevőknek explicit ítéleteket sem kell meghozniuk. Ehelyett egy Hornby (1971) és Chen (1998) által kidolgozott teszt leleményes adaptáci-

óját alkalmazták, amelyben a feladat hamis állítások javítása volt, a megfelelő korrekció pedig a tesztmondatban található fókusz helyes azonosításán alapult. A kapott eredmények azt mutatták, hogy az angol, francia és német gyerekek ezt a feladatot már háromévesen felnőttszerűen teljesítették, mi több, egyáltalán nem volt korcsoporti hatás a három-, négy-, öt-, illetve hatévesek között. A szerzők ezt amelletti bizonyítékként értelmezték, hogy a gyerekeknek a prozódiai prominencia fókusszal való társítására vonatkozó kompetenciája már igen korai életkorában teljes.

Amint azt ez a rövid áttekintés is egyértelműen mutatja, a prozódiai fókuszjelölés elsajátítására vonatkozó korábbi empirikus munkák eredményei és következtetései nagyfokú változatosságot mutatnak. E különbségek egyik legfőbb forrása az alkalmazott módszerek sokféleségében keresendő. Egy másik, ám mindeddig sokkal kevesebb figyelemben részesített eredet lehet az, hogy sok esetben a különféle kísérletekkel más-más nyelveket vizsgáltak. Miközben a fókusz produkciójának kutatásában kiemelt szerepet kapott a nyelvek fókuszjelölési rendszerei között megfigyelhető jelentős variabilitás tényezője (ennek áttekintését lásd Chen, 2018), addig ugyanezt nem tanulmányozták szisztematikusan a fókusz megértésének tesztelésekor. (Ez alól kivételt képez Szendrői et al., 2018, valamint Chen et al., 2019.) Jelen kísérletes vizsgálat e hiány betöltéséhez kíván hozzájárulni.

3. A vizsgálat

3.1. Kutatási kérdés

A cél egy tágabb kontextusban annak vizsgálata, hogy milyen hatással van a prozódiai fókuszjelölés megértésének fejlődésére az egyes nyelvekben megfigyelhető, a fókuszjelölési módokra vonatkozó változatosság. Jelen tanulmány szűkebb kutatási kérdése pedig az, hogy egy adott nyelvben befolyásolja-e – és ha igen, milyen módon – a szintaktikai fókuszjelölés szisztematikus jelenléte a prozódiai fókuszjelölés megértésének elsajátítási menetét, illetve ütemét. Éppen

azért a magyarban teszteltük prozódiai fókuszjelölés elsajátítását, mert ebben a nyelvben a fókusz szintaktikai jelölése gyakorlatilag kötelező.

A magyar – a germán és újlatin nyelvekhez hasonlóan – hangsúllyal jelöli a fókuszt. A leggyakoribb szórend az alany–ige–tárgy (SVO) sorrend. Az SVO-nyelvekben az ige előtti alany alapértelmezetten topik szerepű, miközben az ige és a tárgy kommentként értelmeződik. Fontos továbbá, hogy a nukleáris hangsúly – amelyet a cikk hátralevő részében félkövérrel jelölünk – alapesetben a komment bal szélső elemére esik (É. Kiss, 1987, 2002; Kenesei & Vogel, 1989; a nukleáris hangsúllyal kapcsolatban nincs ugyanakkor teljes egyetértés: például Varga, 1983, 2008 szerint a semleges mondatokban nincs jelen nukleáris hangsúly). A fókusz (a korrektív fókusz és az átlagos válasz-fókusz is) kötelezően jelölve van a szórend által: a fókuszált kifejezést előre kell vinni abba a pozícióba, amely balról csatlakozik az igéhez. Gyakori, hogy az igéhez tartozik egy igekötő vagy valamilyen más igemódosító (É. Kiss, 2002), amelynek pozíciója szintaktikailag megkülönbözteti egymástól a pre-verbális topikot (1a) és a pre-verbális fókuszt (1b). Hasonló igemódosító hiányában az ige előtti argumentumot tartalmazó (pl. SVO szórendű) mondatok potenciálisan többértelműek, és számos információszerkezeti értelmezés társítható hozzájuk. Lehetnek például semleges, tág fókuszos mondatok, amelyeket általában az alapértelmezett információszerkezetnek tartanak. Amint említettük, ugyancsak alapértelmezettnek tekinthető, hogy a pre-verbális, határozott, külső argumentum szerepű alanyt topikként, a mondat többi részét pedig kommentként elemezzük, és ekkor a nukleáris hangsúly az igére esik (2a). Egy ilyen szerkezetű mondat igéje ugyanakkor kaphat szűk fókuszos értelmezést is (2b). Annak ellenére, hogy (2b) prozódiája számos fonetikai paraméter tekintetében eltérhet (2a) prozódiájától, a nukleáris hangsúlyt viselő elem mindkettőben az ige. Végül kaphat főhangsúlyt, és ily módon szűkfókusz-értelmezést a mondat alanya is (2c). Fontos megemlíteni, hogy az utóbbi az egyetlen olyan eset, amikor a határozott, külső argumentum szerepű alany pre-verbális pozícióban a komment részeként értelmeződik.

(1) a. ...TOPIK [**IGEMÓDOSÍTÓ** IGE...]$_{\text{KOMMENT}}$
János **meg** főzi a krumplit

b. ...[**FÓKUSZ** IGE IGEMÓDOSÍTÓ...]$_{\text{KOMMENT}}$
János főzi meg a krumplit

(2) a. S_{TOPIK} [**V...**]$_{\text{FÓKUSZ=KOMMENT}}$ (tág fókusz)
János **főzi** a krumplit (és nem alszik).

b. S_{TOPIK} [$\mathbf{V_{FÓKUSZ}}$**...**]$_{\text{KOMMENT}}$ (igefókusz)
János **főzi** a krumplit (és nem süti).

c. [$\mathbf{S}_{\text{FÓKUSZ}}$ V...]$_{\text{KOMMENT}}$ (alanyi fókusz)
János főzi a krumplit (és nem Mari).

Mivel a magyarban a fókuszált kifejezést kötelezően jelöljük a szórend által is, az anyanyelvi beszélőknek ritkán kell pusztán a prozódiai jelölésre hagyatkozniuk a fókusz azonosításakor (azonban a (4) alatti, illetve a kísérletünkben használt (5) alatti mondatok éppen ezt a szóban forgó esetet példázzák majd). Hipotézisünk szerint a fókusz szisztematikus jelölése a felszíni mondatszerkezetben csökkenti a prozódiai jelölés egyértelműsítő szerepét a gyerekek számára, éppen ezért azt várjuk, hogy a prozódiai fókuszjelölés megértésének elsajátítása késést fog mutatni azokhoz a germán és újlatin nyelvekhez képest, amelyekre az eddigi kutatások zöme koncentrált, és amelyekben a szintaktikai fókuszjelölés legjobb esetben is csak opcionális.

3.2. Módszertan

3.2.1. A kísérlet anyagai és menete

Annak érdekében, hogy minél magasabb fokú összehasonlíthatóságot tegyünk lehetővé, néhány feltétlenül szükséges módosítással ugyan, de azt a kísérleti feladatot alkalmaztuk, mint Szendrői et al. (2018). Ebben a mondat–kép összevetési feladatban a résztvevőknek az volt a feladatuk, hogy megítéljék egy bábu állításait a számítógép képernyőjén vetített képekről: elfogadják az igaz,

illetve kijavítsák a hamis kijelentéseket. A kizárólag prozódiai fókuszjelölést tartalmazó kritikus mondatok minden esetben hamisak voltak a hozzájuk tartozó képre vonatkozóan, és az általunk vizsgált függő változó a javítások kongruenciája volt a kritikus mondatok információszerkezeti sajátosságaihoz viszonyítva.

Szendrői et al. (2018) olyan SVO szórendű mondatokat használtak, amelyekben vagy az alany, vagy a tárgy volt prozódiai fókuszként jelölve.

(3) a. **[The BIRDIE]**$_{\text{FÓKUSZ}}$ has the bottle (alanyi fókusz)
a madárka AUX az üveg

'**A MADÁRKÁNÁL** van az üveg.'

b. The birdie has **[the BOTTLE]**$_{\text{FÓKUSZ}}$ (tárgyi fókusz)
a madárka AUX az üveg

'**AZ ÜVEG** van a madárkánál.'

(Szendrői et al., 2018, 221.)

Azonban mivel a magyarban kötelező a fókuszált összetevőt közvetlenül az ige előtti pozícióba mozgatni, az alanyi, illetve tárgyi fókuszos mondatoknak a prozódiáján kívül a szórendje is eltér: az előbbi esetben az alany áll az ige bal oldalán, míg az utóbbi esetben a tárgyi szerepű összetevő tölti be ezt a pozíciót. Éppen ezért ahelyett, hogy tranzitív igéket használtunk volna, amelyeknek az alanyát vagy a tárgyát lehetett volna fókuszálni, a magyarban igemódosító nélküli SV szórendű mondatokat teszteltünk, amelyekben vagy az alanyt, vagy az igét jelöltük prozódiai fókuszként. Az ilyen szórendű mondatok – amint azt (4a) és (4b) esetében is láthatjuk – anélkül térnek el fókusz-értelmezésükben, hogy szórendileg különböznének egymástól. Azért, hogy fenntartsuk a párhuzamot a (3b) típusú angol (illetve német és francia) tárgyi fókuszos mondatokkal, melyekben a tárgyi fókusz mondatvégi pozíciót foglal el, a magyar tesztmondatok esetében az igét mondatzáró pozícióba helyeztük; a tesztmondatainkban tehát kizárólag alany és ige szerepelt:

(4) a. **[JÁNOS]**$_{\text{FÓKUSZ}}$ főz. (alanyi fókusz)

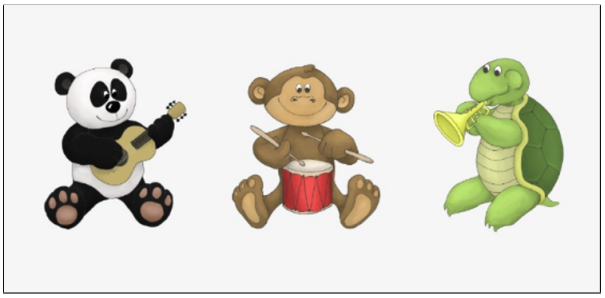
b. János **[FŐZ]**$_{\text{FÓKUSZ}}$. (igefókusz)

További párhuzam a (4a,b) és a (3a,b) típusú stimulusok között, hogy a mondatokban minden esetben pontosan két olyan szó szerepelt, amely lexikális kategóriába tartozván alapértelmezetten dallamhangsúlyt hordoz.

Az alanyi fókusz prozódiai jelölése, mint (2c)-ben láttuk, az alapértelmezetten az igére eső nukleáris hangsúly „áthelyezésével" jár együtt. Ez az áthelyezett nukleáris prominencia szűk alanyifókusz-értelmezést von maga után. Az igefókusz prozódiai jelölése nem kívánja meg a nukleáris hangsúly alapértelmezett helyzetének megváltozatását: a mondat fő hangsúlya mind tág fókusz esetén (2a), mind igefókusz esetén (2b) az igére esik. Mivel a kizárólag alanyból és igéből álló, SV szórendű mondatok kommentje az igén túl nem tartalmaz további összetevőket, a nukleáris hangsúlyt az igére helyező prominencia mintázattal társítható a (2b)-hez hasonló szűk igefókuszú értelmezés is, és a (2a)-hoz hasonló tág (VP) fókuszú értelmezés is. Szendrői et al. (2018) kísérletének tárgyi fókuszos tesztmondatai hasonlóak: a fókusz projekció (Selkirk, 1984) révén az SVO mondatok szűk tárgyi fókuszos változatainak prozódiai megvalósítása szintén lehetővé teszi a tág (VP) fókuszos olvasatot is.

Összegezve: olyan SV szórendű, csak alanyt és igét tartalmazó tesztmondatokat hoztunk létre, amelyekben a szűk fókuszt kizárólag a prozódiai prominenciaviszonyok alapján lehet egyértelműen azonosítani.

Egy-egy tesztmondat lejátszásával egyidejűleg a résztvevőknek egy képet is vetítettünk, amelyen három ábra volt egymás mellett. Annak érdekében, hogy a vizuális stimulusaink a lehető leginkább hasonlítsanak a Szendrői et al. (2018) által használt, három állat–tárgy párból álló rajzokra, azonos szerkezetű képeket készítettünk, azaz mindhárom ábra egy állatot és egy tárgyat tartalmazott, méghozzá úgy, hogy az állat valamilyen tevékenységet végzett a tárgy segítségével (1. ábra). Az a főnév, amelynek jelölete a képen látható tárgy volt, minden kritikus és kontroll mondatban megjelent egy az adott főnévből képzett ige inkorporált névszói töveként, így például az 1. ábrán található trombita az



1. ábra. A (5a)/(5b) kritikus mondathoz tartozó képi stimulus

(5) alatti mondatokban szereplő trombitál ige töveként. Az 1. ábrához tartozó alanyi fókuszos mondat (5a), igefókuszos párja pedig (5b).

(5) a. **A MAJMOCSKA** trombitál. (alanyi fókusz)

b. A majmocska **TROMBITÁL**. (igefókusz)

Egy szűk fókuszt tartalmazó hamis állítás javítása akkor számít kongruensnek, ha ugyanarra a kérdésre válaszol, mint a korrigálni kívánt kijelentés (Rooth, 1992; Roberts, 1996). A kísérletben tehát attól függően, hogy egy résztvevő az alanyt vagy az igét értelmezte-e a stimulus mondat fókuszaként, kétféleképpen javíthatta az elhangzott hamis állításokat: vagy az alany, vagy az ige korrekciójával. Vagyis az (5) alatti példamondat és az 1. ábra esetében vagy azt a választ adta, hogy „nem, mert A TEKNŐS trombitál" (alany-korrekció, mely (5a) esetében kongruens), vagy azt felelte, hogy „Nem, mert a majmocska DOBOL" (ige-korrekció, mely (5b) esetében kongruens). A fókusz típusa (alanyi vagy igefókusz) személyközi faktor volt, azaz minden életkori csoportban a résztvevők egyik fele kizárólag alanyi fókuszt tartalmazó mondatokat hallott tesztmondatként, másik fele pedig csak igefókuszt tartalmazó mondatokat.

A kontroll mondatok minden esetben a kritikus mondatokkal megegyező típusúak voltak (azaz vagy alanyi, vagy igefókuszt tartalmaztak) – a különbség abban állt, hogy ezek az állítások igazak voltak a velük párhuzamosan prezentált képekre, így esetükben nem korrekciót, hanem egyetértést kifejező válaszokat vártunk. A kritikus, illetve kontroll mondatok mellett a kísérlet tartalmazott

fillereket is, amelyek egyik fele igaz, másik fele hamis állítás volt. A fillerek olyan SV alakú mondatok voltak, amelyekben az alanyi funkciót egy univerzálisan kvantifikált főnévi kifejezés töltötte be, mint például a (6) alatti mondatban. A filler mondatokkal párosított képek felépítése megegyezett a kritikus és a kontroll mondatokhoz társított képekével, azonban a hamis fillerek esetén a képen szereplő állatok nem a mondatban szereplő tevékenységet végezték.

(6) Minden állat fűrészel.

Mindegyik kísérleti ülés tizennégy próbából állt: a kísérletvezető egy rövid bemelegítő rész után négy kritikus, négy kontroll és négy filler mondat–kép párt prezentált a résztvevőknek a kétféle kiegyensúlyozott pszeudo-random sorrend valamelyikében. A hangzó stimulust előzetesen az egyik szerző felolvasásában rögzítettük. A képek egy laptop képernyőjén voltak kivetítve, a hanganyagot pedig egy a bábuba rejtett hangszóró játszotta le. Ezt a megoldást két fontos tényező is indokolttá tette. Egyrészt közismert, hogy a gyerekek nagyobb hajlandóságot mutatnak arra, hogy egy bábu állításait kijavítsák, mintsem hogy egy felnőtt kísérletvezető kijelentéseit hamisnak nyilvánítsák. Másrészt pedig a prozódiai jelölés feldolgozásának kísérletes vizsgálatakor különösen nagy jelentőséggel bír a hangzó stimulusok állandóságának megőrzése, kiváltképp mivel a prozódiai fókuszjelölés a fonetikai paraméterek körében többféle rejtett variációt is tartalmazhat, és ez számtalan módon befolyásolhatja a fókusz azonosítását. A két kritikus tesztmondattípust reprezentáló (5a) és (5b) mondatok akusztikai megvalósulását a Melléklet 1. és 2. ábrája szemlélteti. Emellett az – ugyancsak a Mellékletben található – 1. táblázat tartalmazza a két mondattípus néhány releváns fonetikai paraméterét is.

3.2.2. Résztvevők

A négy korcsoportba sorolható magyar anyanyelvű, egynyelvű gyerekeket véletlenszerűen választottuk ki több óvodában, illetve iskolában. A filler próbákon mutatott összteljesítményük alapján végül 14 négyéves (átlagéletkor: 4;5, szórás: 4,02), 22 ötéves (átlagéletkor: 5;5, szórás: 3,11) és 22 hatéves (átlagéletkor:

6;4, szórás: 3,43) óvodás, illetve 22 hétéves iskolás (átlagéletkor: 7;7, szórás: 4,20) adatait vontuk be az elemzésbe. Emellett kontrollcsoportként teszteltünk 20 felnőtt magyar anyanyelvi beszélőt.

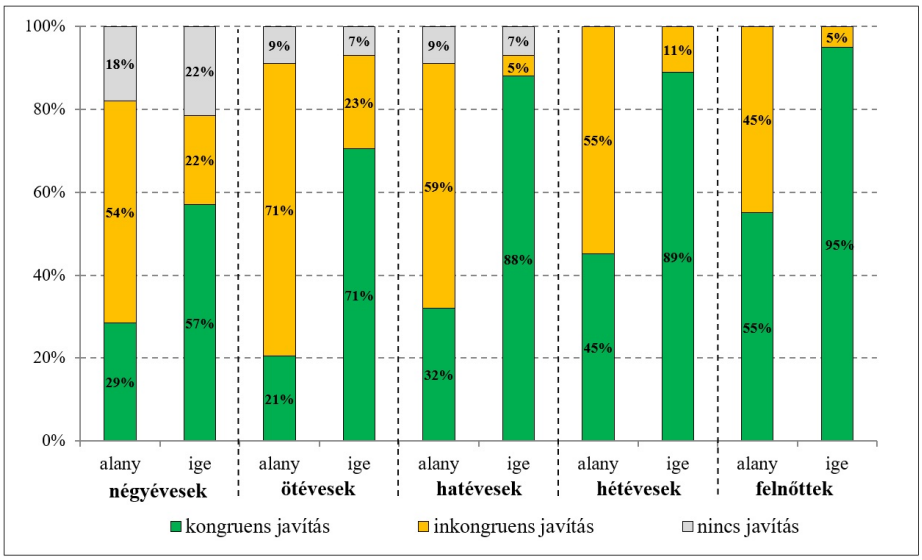
3.2.3. Predikciók

Ahogyan azt a 3.1. fejezetben előrevetítettük, a pusztán prozódiai fókuszjelölés értelmezésében késést jósolunk a magyarban a Szendrői et al. (2018) által vizsgált, szintaktikai fókuszjelölést csak opcionálisan alkalmazó nyelvekhez képest. Várakozásaink szerint ez a relatív késés a nyelvelsajátításban kétféleképpen mutatkozhat meg. Egyfelől az életkor hatását jósoljuk (első predikció, a továbbiakban P1): egészen pontosan azt, hogy a magyar anyanyelvű gyerekek fókusz-kongruens javításainak aránya az életkor előrehaladtával emelkedni fog, így különbség lesz az egyes korcsoportok teljesítménye között – szemben Szendrői et al. (2018) eredményeivel. Másfelől várjuk az elsajátítandó nyelv hatását is (második predikció, a továbbiakban P2): eszerint az alapértelmezettől eltérő szűk fókuszos mondatok (azaz az alanyi fókuszt tartalmazó állítások) esetében a kongruens válaszok aránya a magyar gyerekeknél elmarad majd az azonos korú angol, francia és német gyerekek eredményeitől, legalábbis a legfiatalabb életkori csoportban, a négyéves óvodásoknál.

3.3. Eredmények

A 2. ábra a különféle javítástípusok eloszlását mutatja korcsoportok szerinti bontásban.

A kapott válaszokat a statisztikai elemzéshez bináris adatokként kódoltuk attól függően, hogy kongruensek vagy nem-kongruensek voltak-e. (Ez utóbbi kategória részét képezték a túlnyomó többségben lévő inkongruens javítások mellett az alkalomszerűen előforduló elfogadó válaszok is, amelyek ily módon nem tartalmaztak semmilyen javítást sem.) Az elemzést binomiális általánosított kevert modellek segítségével végeztük az R-szoftver (R Core Team, 2019) és az *lme4* csomag (Bates et al., 2015) segítségével. A függő változónk a javítások kongruenciája volt, a modellszelekciót követően pedig rögzített független



2. ábra. A javítástípusok arányai a vizsgált öt életkori csoportban

változóként a FÓKUSZTÍPUS (alanyi vagy igefókusz) és a KORCSOPORT (négy-, öt-, hat- vagy hétévesek, illetve felnőttek) szerepelt, a résztvevőkkel és az egyes mondat–kép párokkal pedig mint random hatásokkal számoltunk.

Az elemzés kimutatta, hogy mind a FÓKUSZTÍPUS, mind a KORCSOPORT szignifikánsan befolyásolta a kongruens válaszok arányát, interakció viszont nem volt a két tényező között. Ami a FÓKUSZTÍPUSt illeti, az eredmények alapján az igefókuszos mondatok esetében – korcsoporttól függetlenül – több kongruens javítás született, mint az alanyi fókuszos kondícióban ($\chi^2(1) = 15,28$, $p < 0,001$). A FÓKUSZTÍPUS szignifikáns hatása abban nyilvánult meg, hogy a fókusz-kongruens válaszok aránya az életkorral párhuzamosan nőtt ($\eta^2(4) = 24,23$, $p < 0,001$), méghozzá mindkét fókusztípus esetén. Az Anova() függvénnyel (*car* csomag, Fox & Weisberg, 2019) végzett *poszt-hoc* tesztek arra is rávilágítottak, hogy egyedül a hétévesek teljesítménye nem különbözött szignifikánsan a felnőttekétől ($Z = 0,43$, $p = 0,664$), a gyerekek csoportjain belül pedig egyedül a négy- és ötévesek között nem volt jelentős eltérés ($Z = -0,14$, $p = 0,889$); minden más páronkénti összevetés során szignifikáns különbséget találtunk. Itt jegyezzük meg, hogy a négy-, öt- és hatéves korcsoportok felmérése

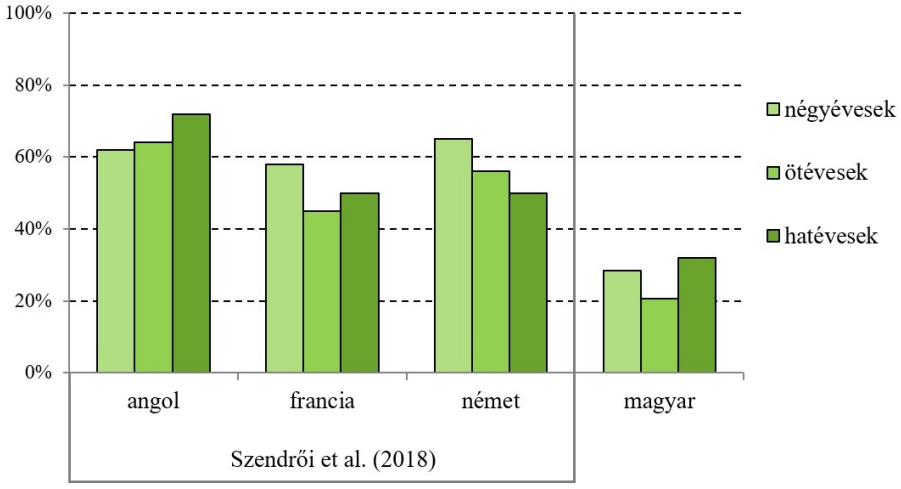
mellett a hétéves korcsoport vizsgálatba vonása azt szolgálta, hogy megállapítsuk, a magyar gyermekek milyen korra érik el a felnőttszerű szintet. Hogy a korcsoportok száma megegyezzen a Szendrői et al. (2018) kísérletében szereplő korcsoportok számával, hároméves korúakat nem vizsgáltunk.

3.4. Diszkusszió

Kísérletünkben azt a hipotézist teszteltük, hogy késést mutat-e a prozódiai fókusz felnőttszerű azonosításának elsajátítása a magyar nyelvben, ahol a prozódiai fókuszjelölés mellett a szintaktikai fókuszjelölés is kötelező, így az előbbi egyértelműsítő szerepe jóval kisebb, mint például az angol, francia vagy német nyelvekben.

Ennek a késésnek a megnyilvánulását egyrészt az életkor hatásában (P1) jósoltuk. Amint azt a 2. fejezetben bemutattuk, Szendrői et al. (2018) ugyanezt a feladattípust alkalmazva azt találták, hogy az angol, francia és német gyerekek már hároméves korukban felnőttszerűen azonosítják a pusztán prozódiailag jelölt fókuszt, és nincs különbség a vizsgált gyerekcsoportok (a három-, négy-, öt-, illetve hatévesek) teljesítménye között sem. Ezzel szemben a magyarban az életkor szignifikánsan befolyásolta a fókusz-kongruens válaszok arányát: a hatéves gyerekek mind a négy-, mind az ötévesekhez képest több kongruens javítást adtak, a hétévesek pedig mindhárom fiatalabb életkori csoportnál jobban teljesítettek, sőt az ő válaszaik már elérték a felnőttszerű szintet is. Első predikciónk tehát maradéktalanul teljesült.

Egy másik lehetséges út a késés bizonyítására annak a különbségnek a kimutatása volna, amelyet az alanyi fókusz kondícióban adott kongruens javítások között várunk a magyar, illetve az angol, francia és német gyerekek tesztelésekor – legalábbis a legfiatalabb résztvevőink, a négyévesek esetében (P2). Elsősorban azért ebben a kondícióban vizsgáljuk az eltérést, mert a másik (a magyar esetében igefókuszos, az angol, francia és német esetében pedig tárgyi fókuszos) kondícióban a fókuszált összetevő az egyes nyelvek alapértelmezett fókuszaként is funkcionál. Mivel nem volt előzetes feltételezésünk arra vonatkozóan, hogy milyen lehetséges különbségekhez vezethet az egyes nyelvekben az alapértelme-



3. ábra. Az alanyi fókusz kondícióban adott kongruens javítások aránya a két kísérletben (A diagram (Szendrői et al., 2018, 234.) 2. ábrája alapján készült.)

zett fókuszra való támaszkodás, a magyar igefókuszos mondatok korrekciójának arányait nem vetettük össze a korábban vizsgált nyelvek tárgyi fókuszos kondícióival.

Az alanyi fókuszos mondatok nyelvek között megfigyelhető különbségére irányuló predikciónkat (P2) nyilvánvalóan alátámasztják a kapott eredmények. A kongruens alanyifókusz-javítások aránya valóban körülbelül a fele volt az angol, francia és német gyerekek kongruens válaszainak – méghozzá nemcsak a négyéves, hanem még az öt- és hatéves óvodások csoportjaiban is (3. ábra). Noha ez a megállapítás csak a Szendrői et al. (2018) által publikált átlagértékekkel való hozzávetőleges összevetés eredménye, a különbség mértéke miatt mégis igencsak sokatmondó.

Az a tény, hogy a magyar gyerekek az iskoláskor előtt nem érik el a prozódiai fókuszjelölés értelmezésének felnőttszerű szintjét, miközben az angol, francia és német kortársaiknál ez egyértelműen kimutatható, valamint az a megfigyelés, hogy az alanyi fókusz kondícióban a fókusz-kongruens javításaik aránya az említett gyerekekének csak körülbelül a felét tették ki, egyaránt arra utal, hogy a

prozódiai fókusz azonosításának elsajátítása a magyar nyelvben valóban késést mutat a másik három nyelvhez képest.

Miután maga a feladat ugyanaz volt, mint Szendrői et al. (2018) kísérletében, a feltárt különbségek nem tudhatók be a feladat hatásának, azaz nem valószínű, hogy a két kísérlet eltérő módon korlátozta volna az azonos mögöttes tudás megnyilvánulását. Ehelyett a magyarban talált viszonylagos késés a prozódiai fókuszjelölés értelmezésében könnyedén magyarázható a nyelvben található szintaktikai fókuszjelölés gyakoriságának hatásaként. Ahogyan azt a 3.1. fejezetben részletesen kifejtettük, a fókuszált kifejezés kötelezően jelölve van az igét közvetlenül megelőző pozícióba történő mozgatása révén is. Ennek a szintaktikai fókuszjelölésnek a szisztematikus, felszíni megjelenése csökkenti a prozódiai jelölés funkcionális terhelését a gyermekek fókuszértelmezésében. Ez ugyanakkor azt is eredményezi, hogy a magyar anyanyelvű óvodások kevésbé megbízhatóan hasznosítják a prozódiai fókuszjelölést, mint a germán vagy újlatin nyelveket elsajátító kortársaik, ezekben a nyelvekben ugyanis a fókusz azonosításakor a prozódia szolgál annak legfőbb jelölőjeként, miközben a szintaktikai fókuszjelölés legfeljebb csak egy lehetőség.

Abból az eredményből, hogy a szisztematikus szintaktikai fókuszjelölés jelenléte számottevő mértékben késlelteti a prozódiai fókuszjelölés értelmezésének fejlődését a magyarban, arra következtetünk, hogy a gyerekeknek a prozódiai fókuszjelölés megértésére vonatkozó kompetenciája korai óvodáskorban még nem feltétlenül teljes – szemben azzal, amit Szendrői et al. (2018) az angol, német és francia adatok alapján feltételeznek.

Ez az eredmény azt a további kérdést is felveti, hogy egy adott nyelvben használatos szintaktikai fókuszjelölés pontosan milyen feltételek teljesülése esetén elegendően szisztematikus ahhoz, hogy a magyarhoz hasonlóan kimutathatóan késleltesse a prozódiai fókuszjelölés felnőttszerű értelmezésének elsajátítását. Elegendő-e, ha a szintaktikai jelölés nem kötelező ugyan (mint a magyarban), hanem csak erősen preferált, vagy ha csak részben kötelező: például ha csak az alanyi argumentum esetében elengedhetetlen (pl. hausa és tangale, lásd Green & Jaggar, 2003; Hartmann & Zimmermann, 2004). Ennek a kérdésnek a megvá-

laszolása további nyelvközi vizsgálatokat igényel. Annyit azonban Szendrői et al. (2018) francia nyelvből származó eredményei alapján is megállapíthatunk, hogy amennyiben a szintaktikai jelölés csak az alanyi argumentum esetében jellemző, és ott is csupán preferált alternatívaként van jelen a grammatikai rendszerben, az nem elegendő ahhoz, hogy a prozódiai fókuszjelölés elsajátítását jelentősen késleltesse. Szendrői et al. ugyanis annak ellenére nem találtak életkori hatást a francia gyerekek esetében, hogy a franciában az argumentumok közül az alany fókusz státuszát hangsúlyáthelyezés helyett tipikusan inkább szintaktikai eszközzel, szétszakított (cleft) szerkezetekkel jelölik (Lambrecht, 1994).

Az utolsó megvitatandó eredmény a fókusztípus hatása: azt találtuk ugyanis, hogy a kongruens javítások aránya az igefókuszos kondícióban minden életkori csoportban magasabb volt, mint az alanyi fókuszos kondícióban. Ez a konzisztens aszimmetria több tényezőből is fakadhat. Egyfelől az alapértelmezett információszerkezet a tág fókuszos értelmezés, másfelől pedig – amint azt a 3.1. fejezetben áttekintettük – a magyarban az ige előtti határozott alany topikértelmezést kap. Vagyis az általunk használt SV szerkezetű mondattípus alapértelmezett információszerkezetében az alany topikként, az ige pedig fókuszként funkcionál – hasonlóan a szűk igefókuszos értelmezés esetéhez. Feltételezhető, hogy a résztvevők a tesztmondatok feldolgozásakor az esetek egy részében ehhez az alapértelmezett jelentéshez nyúltak vissza, ami eltolhatta a tesztmondatokhoz társított interpretációk arányát mind az alanyi fókuszos, mind az igefókuszos kondícióban, méghozzá mindkét esetben az igefókuszos jelentésváltozat javára. Az alanyi fókuszos mondatok esetében ez a tényező csökkenthette, az igefókuszos mondatok esetében pedig növelhette a helyes javítások arányát. Az eredményeinkből – különösen a felnőtt korcsoport alanyi fókusz kondíciójában talált javítási arányokból – úgy tűnik, hogy az alapértelmezett információszerkezethez történő visszanyúlás annak ellenére volt pragmatikai szempontból megengedett stratégia a válaszadásban, hogy ennek alkalmazása az alanyi fókuszos kondícióban technikai értelemben véve (ld. Rooth, 1992; Roberts, 1996) inkongruens korrekciókhoz vezetett. E stratégia alkalmazásának több lehetséges oka is elképzelhető; kísérletünk szempontjából azonban elsősorban az bír jelentőséggel,

hogy a kongruens válaszok aránya hogyan alakul a vizsgált nyelv és az életkor függvényében.

Az alanyi fókuszos és igefókuszos kondícióink között találthoz hasonló aszimmetria figyelhető meg Szendrői et al. (2018) eredményeiben is, ahol a tárgyi fókusz kondícióban adott kongruens válaszok aránya rendre magasabb volt az alanyi fókuszos kondícióhoz viszonyítva. Az általuk vizsgált nyelvekben alapértelmezett esetben az SVO szórendű mondatok tárgya viseli a nukleáris hangsúlyt, hasonlóan a tárgyi fókuszos mondatokhoz, de eltérően az alanyi fókuszos mondatoktól. A szerzők az alanyi és tárgyi fókuszos kondíciók közötti eltérés egyik lehetséges magyarázataként utalnak rá, hogy az előbbieknek az alapértelmezettől eltérő prominencia mintázatát a résztvevők az esetek egy részében figyelmen kívül hagyták a mondatértelmezés során. Az általunk a magyarral kapcsolatban imént említett javaslat igen közel áll ehhez a felvetéshez; csupán abban térnek el, hogy míg mi az alapértelmezett információszerkezet felől közelítjük meg a kérdést, addig Szendrői et al. (2018) az alapértelmezett információszerkezettel társuló prominenciamintázat oldaláról.

Az igefókuszos értelmezés irányában tapasztalt eltolódás azonban pusztán prozódiai alapon is magyarázható, amint arra egyik anonim bírálónk rámutatott. Az SV felépítésű tág fókuszú magyar mondatokban ugyanis a deklináció mechanizmusának köszönhetően az alany dallamhangsúlyában mért alapfrekvencia-maximum jellemzően magasabb, mint az igén mért érték. Ebben a tekintetben a szűk alanyi fókuszos mondatok hasonlítanak a tág fókuszú megvalósításhoz, hiszen esetükben is magasabb az alapfrekvencia-csúcs az alanyon, mint az igén (vö. Melléklet, 1. ábra). A szűk igefókuszos mondatok ebből a szempontból különböznek mindkét előbb említett mondattípustól, esetükben ugyanis fordított a helyzet: az igén mérhető magasabb alapfrekvencia-csúcs, mint az alanyon (vö. Melléklet, 2. ábra). Így az igefókuszos mondatok dallammenete jobban elüt – és ezért könnyebben megkülönböztethető – a tág fókuszú mondatokra jellemző dallammenettől, mint az alanyi fókuszos mondatoké.

4. Összegzés

A tanulmány a gyerekek fókuszazonosítási képességeit vizsgálta a magyarban, amely nyelv a prozódiai fókuszjelölés mellett kötelező szintaktikai fókuszjelölést is használ. Annak érdekében, hogy a kapott eredmények felhasználásával legalább egy hozzávetőleges összevetést lehessen végezni más nyelvekkel, kísérletünkben ugyanazt a feladatot használtuk, amelyet Szendrői et al. (2018) is alkalmaztak nemrégiben az angol, francia és német óvodások tesztelésére. Annak a hipotézisnek az alapján, hogy a fókusz rendszerszerű szintaktikai jelölése a magyarban csökkenti a gyerekek számára a prozódiai jelölés egyértelműsítő szerepét, azt vártuk, hogy az olyan mondatok esetében, ahol a fókuszt csupán a prozódia jelöli egyértelműen, a prozódiai fókuszjelölés felnőttszerű értelmezése késést fog mutatni a Szendrői et al. (2018) által vizsgált nyelekhez képest. Ezt a feltételezést a kapott adatok alátámasztották: a négyéves magyar óvodások a szűk alanyi fókuszt tartalmazó mondatokat körülbelül fele olyan gyakran javították kongruens módon, mint az angol, francia és német kortársaik, akik ebben az életkorban már felnőttszerűen teljesítettek. A vizsgált magyar gyerekek ezt a szintet mindössze hétéves korban érték el.

Eredményeink így arra utalnak, hogy – a prozódiai fókuszjelölés produkciójának elsajátításához (Chen, 2018) hasonlóan – a prozódiai fókuszjelölés értelmezésének fejlődési menetét is nagymértékben befolyásolja a fókuszjelölésben megfigyelhető nyelvek közötti változatosság. Kísérletünk egészen pontosan arra világított rá, hogy hatással bírhat egy az adott nyelv fókuszjelölésében szerepet játszó alternatív, nem-intonációs (a magyarban: szintaktikai) eszköz gyakorisága. Ebből a szempontból a jelen munka Chen et al. (2019) tanulmányának fordítottjaként is tekinthető: az a cikk ugyanis azt vizsgálta, hogy mennyiben lehetnek relevánsak az elsajátítási folyamat szempontjából az adott nyelv által a fókusz jelölésére használt fonetikai sajátosságok egyéb alternatív, rendszerszerűen nem-intonációs (a mandarinban: lexikai tonális) funkciói.

Köszönetnyilvánítás

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal KH 130558 azonosítószámú pályázata támogatta. Köszönettel tartozunk Turi Gergőnek a kísérlet során használt hanganyagok előállításában és fonetikai elemzésében nyújtott segítségéért.

Hivatkozások

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. URL: https://doi.org/10.18637/jss.v067.i01. doi:10.18637/jss.v067.i01.

Berger, F., & Höhle, B. (2012). Restrictions on addition: children's interpretation of the focus particles auch 'also' and nur 'only' in german. *Journal of Child Language*, *39*, 383–410. doi:10.1017/S0305000911000122.

Bergsma, W. (2006). (un)stressed ook in Dutch. In V. Geenhoven (Ed.), *Semantics in acquisition* (p. 329–348). Dordrecht: Springer.

Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars*. Cambridge, MA: MIT Press.

Brown, R. (1973). *A First Language*. Cambridge, MA: Harvard University Press.

Chen, A. (2018). Get the focus right across languages: Acquisition of prosodic focus-marking in production. In P. Prieto, & N. Esteve-Gibert (Eds.), *Prosodic development in first language acquisition* (p. 295–314). Amsterdam: John Benjamins.

Chen, H-C., Szendrői, K., Crain, S., & Höhle, B. (2019). Understanding prosodic focus marking in mandarin chinese: data from children and adults. *Journal of Psycholinguistic Research*, *48*, 19–32. URL: https://doi.org/10.1007/s10936-018-9580-9. doi:10.1007/s10936-018-9580-9.

Chen, S-H. E. (1998). Surface cues and the development of given/new interpretation. *Applied Psycholinguistics*, *19*, 5535–5582. URL: https://doi.org/10.1017/S0142716400010365. doi:10.1017/S0142716400010365.

Chien, Y-C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition*, *1*, 225–295.

Costa, J., & Szendrői, K. (2006). Acquisition of focus marking in European Portuguese – Evidence for a unified approach to focus. In V. Torrens, & L. Escobar (Eds.), *The acquisition of syntax in Romance languages* (p. 319–330). Amsterdam: John Benjamins.

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language*, *12*, 643–661.

Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language*, *14*, 145–167.

É. Kiss, K. (1987). *Configurationality in Hungarian*. Budapest: Akadémiai Kiadó.

É. Kiss, K. (2002). *The Syntax of Hungarian*. Cambridge: Cambridge University Press.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression*. (3rd ed.). Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, *4*. URL: https://doi.org/10.1038/ncomms2430. doi:10.1038/ncomms2430.

Green, M., & Jaggar, P. (2003). Ex-situ and in-situ focus in hausa: Syntax, semantics, and discourse. In J. Lecarme (Ed.), *Research in Afroasiatic Grammar II* (p. 187–213). Amsterdam: John Benjamins.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001). At the semantics/pragmatics interface in child language. In R. Hastings, B. Jackson, & Z. Zvolenszky (Eds.), *Semantics and Linguistic Theory 11* (p. 231–247). Ithaca, NY: Cornell University.

Gualmini, A., Maciukaite, S., & Crain, S. (2003). Children's insensitivity to contrastive stress in sentences with 'only'. *University of Pennsylvania Working Papers in Linguistics*, *8*, 87–100.

Hartmann, K., & Zimmermann, M. (2004). Focussing strategies in Chadic: The case of Tangale revisited. In S. Ishihara, M. Schmitz, & A. Schwarz (Eds.), *Interdisciplinary Studies on Information Structure Vol 1. Working Papers of the SFB632* (p. 207–243). Potsdam: Universitätsverlag Potsdam.

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, *120*, 1887–1897. URL: https://doi.org/10.1016/j.lingua.2010.02.002. doi:10.1016/j.lingua.2010.02.002.

Hornby, P. A. (1971). Surface structure and the topic–comment distinction: a developmental study. *Child Development*, *42*, 1975–1988.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behavior & Development*, *32*, 262–274. URL: https://doi.org/10.1016/j.infbeh.2009.03.004. doi:10.1016/j.infbeh.2009.03.004.

Hüttner, T., Drenhaus, H., Vijver, R., & Weissenborn, J. (2004). The acquisition of the German Focus particle auch 'too': Comprehension does not always precede production. In A. Brugos, L. Micciulla, & C. E. Smith (Eds.), *Online Proceedings Supplement of the 28th Annual Boston University Conference on Language Development*. URL: http://www.bu.edu/bucld/.

Kenesei, I., & Vogel, I. (1989). Prosodic Phonology in Hungarian. *Acta Linguistica Hungarica*, *39*, 149–193.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge, MA: Cambridge University Press.

Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.

Menyuk, P. (1969). *Sentences children use*. Cambridge, MA: MIT Press.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, *86*, 253–282. URL: https://doi.org/10.1016/S0010-0277(02)00179-8. doi:10.1016/S0010-0277(02)00179-8.

R Core Team (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. URL: https://www.R-project.org/.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. Yoon, & A. Kathol (Eds.), *The Ohio State University Working Papers in Linguistics 49: Papers in Semantics* (p. 91–136). Columbus: The Ohio State University.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*, 3–11. doi:10.1037/0012-1649.33.1.3.

Sauermann, A., Höhle, B., Chen, A., & Järvikivi, J. (2011). Intonational marking of focus in different word orders in german children. In M. Washburn (Ed.), *Proceedings of the 28th West Coast Conference on Formal Linguistics* (p. 313–322). Somerville, MA: Cascadilla Project.

Schmitz, M., Höhle, B., Müller, A., & Weissenborn, J. (2006). The recognition of the prosodic focus position in German-learning infants from 4 to 14 months. In S. Ishihara, M. Schmitz, & A. Schwarz (Eds.), *Interdisciplinary studies on information structure 5 – Working papers of the SFB632* (p. 187–208). Potsdam: Universitätsverlag Potsdam.

Sekerina, I. A., & Trueswell, J. C. (2012). Interactive processing of contrastive expressions by Russian children. *First Language*, *32*, 63–87. doi:10.1177/0142723711403981.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Speer, S., & Ito, K. (2009). Prosody in first language acquisition – Acquiring intonation as a tool to organize information in conversation. *Language and Linguistics Compass*, *3*, 90–110. URL: https://doi.org/10.1111/j.1749-818X.2008.00103.x. doi:10.1111/j.1749-818X.2008.00103.x.

Szendrői, K., Bernard, C., Berger, F., Gervain, J., & Höhle, B. (2018). Acquisition of prosodic focus marking by English, French, and German three-, four-, five- and six-year-olds. *Journal of Child Language*, *45*, 219–241. doi:10.1017/S030500091700007.

Varga, L. (1983). Hungarian sentence prosody: an outline. *Folia Linguistica*, *17*, 117–151. doi:10.1515/flin.1983.17.1-4.117.

Varga, L. (2008). The calling contour in Hungarian and English. *Phonology*, *25*, 469–497.

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., & Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, *3*, 580. URL: https://doi.org/10.3389/fpsyg.2012.00580. doi:10.3389/fpsyg.2012.00580.

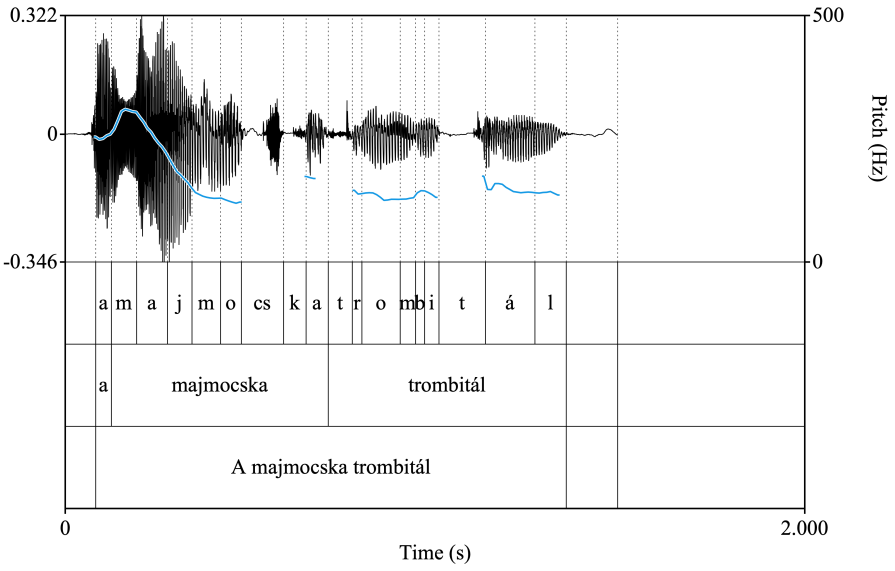
Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, *31*, 749–778. URL: https://doi.org/10.1017/S030500090400652X. doi:10.1017/S030500090400652X.

Wieman, L. (1976). Stress patterns in early child language. *Journal of Child Language*, *3*, 283–286.

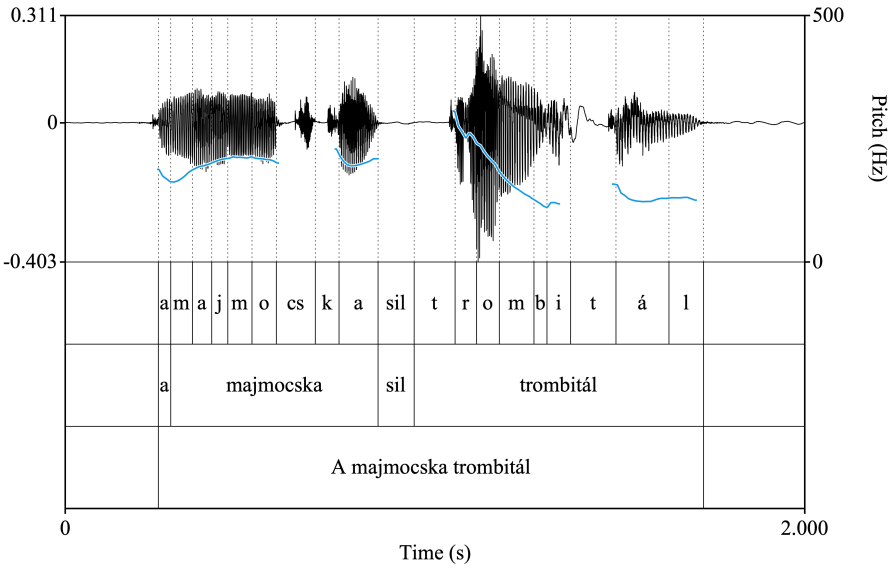
Yang, A., & Chen, A. (2014). Prosodic focus-marking in Chinese four- and eight-year-olds. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Social and Linguistic Speech Prosody* (p. 713–717). Dublin: Trinity College Dublin.

Zhou, P., Su, Y. E., Crain, S., Gao, L., & Zhan, L. (2012). Children's use of phonological information in ambiguity resolution: a view from Mandarin Chinese. *Journal of Child Language*, *39*, 687–730. doi:10.1017/S0305000911000249.

Melléklet



1. ábra. Az (5a) mondat akusztikai megvalósulása



2. ábra. Az (5b) mondat akusztikai megvalósulása

1. táblázat. A kritikus mondatok néhány releváns akusztikai paraméterének átlagértékei

		f0-maximum	**intenzitás**	**duráció**
alanyi fókusz	alany első hangsúlyos szótagjának magánhangzója	316 Hz	76 dB	96 ms
	ige első hangsúlyos szótagjának magánhangzója	136 Hz	65 dB	69 ms
igefókusz	alany első hangsúlyos szótagjának magánhangzója	208 Hz	70 dB	77 ms
	ige első hangsúlyos szótagjának magánhangzója	249 Hz	76 dB	90 ms

A néma szünetek sajátosságai az életkor és a beszédtípus függvényében

Gyarmathy Dorottya[1]

[1] *Nyelvtudományi Intézet*

Abstract

Speech is occasionally interrupted by silent pauses of various length. Pauses serve various functions in speech, like breathing, grammatical function, marking syntactic boundaries, providing time for speech planning processes, for self-repair and for perception as well. The realization of pauses depends on various factors, e.g. the speaker's age, the length and the complexity of the utterance or the speech style. Researches revealed connection between the speech situation and the pauses. The more complex a speech task was – the greater cognitive effort it required – the longer and more frequent the pauses became. Pause in a conversation has also various functions: it plays an important role in turn-taking system, can be connected with pragmatic or social meanings or with cognitive reasons. Furthermore, conversations can have pauses for thinking or for dramatic effect, the speaker can use them to highlight new information, and they can also be used to structure the discourse.

The aim of the study is to analyze the occurrence and duration of silent pauses in two age groups according to their position in conversations. Our hypotheses were that (i) silent pauses realize with different patterns according to age groups; (ii) the duration of silent pauses is determined by their position. 20 conversations and narratives from two age groups (20-35 years old and 40-55 years old) were selected from the Hungarian Spontaneous Speech Database, BEA. Three speakers participated in each conversation; the interviewer and one speaker were collegues, the third participant was the subject. Silent pauses were categorized based on the system developed by Gyarmathy & Horváth (2018).

Results showed that the strategies of pausing are determined by its functions, the speech type and the speakers' age and individual characteristics. Pauses occurred in a grammatically justified position in a greater ratio without breaking the unity of the utterance.

1. Bevezetés

A néma szünet a spontán beszédben számos funkciót betölthet; lehet fiziológiai szükséglet (levegővétel), szolgálhatja az értelmi tagolást, lehet gondolkodási vagy hatásszünet, a beszélő jelezheti vele az új információt, de a társalgásban

Email address: gyarmathy.dorottya@nytud.hu (Gyarmathy Dorottya)

Beszédtudomány – Speech Science 2020. 152–186. DOI: 10.15775/Besztud.2020.152-186

diskurzusszervezői szereppel is bírhat (Esposito et al., 2007). A korai kutatások elkülönítették egymástól a beszélő tervezési nehézségeiből adódó néma szünetet, illetve a szintaktikai szerkezet határán létrejövő junktúrát (Boomer, 1965; Lounsbury, 1965). Egy másik korai megkülönböztetés alapja az, hogy artikulációs okok vagy beszédtervezési probléma áll az adott szünet megjelenése mögött (Goldman-Eisler, 1968). A szünetek osztályozhatóak aszerint is, hogy grammatikai vagy nem grammatikai szerepet töltenek be a beszédben. Az elkülönítés ebben az esetben azon alapszik, hogy tartalmas vagy funkciószó előzi meg, illetve követi őket (Gee & Grosjean, 1983). Azok a szünetek, amelyek tartalmas szó és funkciószó között fordulnak elő, általában grammatikai funkciójúak, szintaktikai vagy prozódiai határt jelölnek. A funkciószót követő és a tartalmas szót megelőző szünetek ezzel szemben egy szintaktikai/prozódiai egységen belül realizálódnak, nem-grammatikai típusúak. A spontán beszédben és a társalgásban többféle szünetet különböztet meg a szakirodalom. A *pause* (magyarul 'szünet') egy beszédfordulón belüli jelkimaradás, a *gap* (magyarul 'rés, hézag') a társalgási egységek közötti szünet, lehetőséget kínál a beszélőváltásokra; a *lapse* szintén ('kihagyás, megszűnés') jelezheti a társalgás végét (Sacks et al., 1974; Levelt, 1989).

A beszédszünetek funkcióinak elkülönítése attól (is) függ, hogy a kutatók mely paradigmarendszer alapján vizsgálják azokat. Bruneau (1973) kommunikációs szempontból három típusú csendet definiált: pszichológiai, interaktív és szociokulturális. A pszichológiai típusú általában nagyon rövid időtartamban, hezitációs jelenséggel vagy tempólassulással valósul meg, és azt a célt szolgálja, hogy időtartama alatt a hallgató feldolgozhassa az elhangzottakat. Az interaktív szünet ennél általában hosszabb időtartamú, a személyek közötti kapcsolatban játszik szerepet, például a beszélőváltások lebonyolítására szolgál. A szociokulturális szünetek egyesítik az első két típus sajátosságait. Zellner (1994) a szünetek kétféle osztályozási rendszerét különböztette meg: 1. a fizikai és nyelvészeti osztályozást, és 2. a pszichológiai és pszicholingvisztikai osztályozást. Az első csoportosítás szerint a beszédszünet lehet intra-szegmentális vagy inter-lexikális, míg a második kategóriarendszer néma és kitöltött szüneteket is

megkülönböztet. Pragmatikai szempontú elemzésében Kurzon (2007) négyféle csendet különböztetett meg: társalgási, tematikus csendet (egy témával kapcsolatban a beszélő nem hajlandó beszélni, például politikai jellegű interjúban), de más típus az is, amikor a társalgási helyzetben egy vagy több résztvevő magában elolvas valamit – például egy osztálytermi helyzetben a tanár instrukciójára a diákok elolvasnak/átfutnak egy részt a tankönyvből. A negyedik típusú a szituációs csend, például egy koncert hallgatása vagy közös megemlékezés közben. Zellner (1994) különféle aspektusokból definiálja a néma szüneteket: beszédtechnológiai szempontból a szünet egy olyan amplitúdóval nem rendelkező egység, amely fizikai jelenség; lehet egy beszédhang része (például zöngétlen zárhangok néma fázisa) vagy megjelenhet szavak között. Pszicholingvisztikai szempontból a néma szünetek együtt járhatnak kilégzéssel, nyeléssel, hangos belégzéssel.

A néma szünet a spontán beszéd leggyakoribb jelensége, amit számos hazai és külföldi kutatás megerősített (Verzeano & Finesinger, 1949; Goldman-Eisler, 1958; Hargreaves & Starkweather, 1959; Boomer, 1965; Levin et al., 1967; Tannenbaum et al., 1967; Misono & Kiritani, 1990; Gósy, 2000; Menyhárt, 2003; Markó, 2005a; Bóna, 2007, 2013b; Neuberger, 2014). A teljes beszédidőhöz viszonyított aránya általában 20–30% körül alakul, időtartamát és gyakoriságát azonban számos tényező befolyásolhatja. Ezek közé tartozik a beszélő személye (Markó, 2005a; Gósy et al., 2011) aktuális fizikai állapota, pl. alkoholos befolyásoltság (Gyarmathy, 2007); a beszédkörnyezet tényezői, pl. zajhatás (Gyarmathy, 2008); a beszédben való jártasság, a beszédhelyzet, a téma (Markó, 2005b, 2014); a beszédstílus (Duez, 1982); a beszédműfaj (Imre, 2005; Olaszy, 2005); a beszédtípus (Markó, 2005a; Váradi, 2010; Bóna, 2013a); az életkor (Laczkó, 2009; Bóna, 2010, 2012) és a nem (Gocsál, 2001). Ezeken túl meghatározó lehet a nyelv (Zwirner & Zwirner, 1937; Trouvain & Möbius, 2014; Trouvain et al., 2016); különféle szintaktikai tényezők, mint a mondat hossza és összetettsége (Volkskaya, 2003; Krivokapic, 2007); illetve a szünet közlésben elfoglalt helye (Sallai & Szende, 1995; Vallent, 2005; Menyhárt, 2010). A kutatások szerint a beszédhelyzet, a szünet funkciója, gyakorisága és időtartama összefüggést mutat. Minél komplexebb egy beszédfeladat, minél nagyobb kognitív

erőfeszítést igényel, annál gyakoribb és hosszabb szüneteket tartanak a beszélők (Goldman-Eisler, 1968; Kowal et al., 1975). Politikai beszédben a néma szünetek gyakrabban és hosszabb időtartamban valósultak meg, a leghosszabb szünetek stilisztikai funkciót töltöttek be – erre a beszédtípusra nem voltak jellemzőek a kitöltött szünetek, amelyek interjúhelyzetben kimondottan gyakoriak voltak (Duez, 1982). Összefüggést találtak továbbá angol nyelvű beszélőknél a szünet pozíciója és időtartama között például a 'to+infinitive' nyelvtani szerkezetek esetében. A felolvasásokban a to elemet megelőző szünetek szignifikánsan hosszabbak voltak, mint az azt követők; míg a spontán beszédben ennek ellenkezője igazolódott feltehetően a beszédtervezés sajátosságai miatt (Bada & Genç, 2008). Oliveira (2002) a narratívák szintaktikai szerkezetének és a néma szünetek időtartamának összefüggéseit elemezte azt feltételezve, hogy a néma szünetek fontos szerepet töltenek be a szintaktikai struktúra jelzésében. A kutatási eredményei megerősítették, hogy a beszélők a narratív egységeik végét rendszerint hosszabb időtartamú szünetekkel jelzik. A beszédtípusnak a néma szünetekre gyakorolt hatását a magyar szakirodalomban is vizsgálták. A kutatások igazolták a spontán beszéd és a hangos olvasás során alkalmazott szünettartási stratégiák különbözőségét (Olaszy, 2005, 2007; Váradi, 2010). Olaszy (2005) novella, mese, hír- és reklámszövegek vizsgálata során megállapította, hogy a reklámok szinte alig tartalmaznak szüneteket; a négy vizsgált szövegtípus közül a legtöbb és átlagosan a leghosszabb mondaton belüli szünetet a novellákban adatolta, míg a mondatközi szüneteket is figyelembe véve a hírekben fordultak elő átlagosan a leghosszabb szünetek. A spontán beszédre kapott prozódiai mutatók alapján megállapította, hogy az leginkább a szünettartási stratégiákban különbözik a felolvasásoktól, a beszélő ugyanis nem mondategységekben, hanem nagyobb, gondolati egységekben valósítja meg a szöveget. A spontán beszédben a felolvasásokhoz képest három-négyszer több szünet fordult elő (Olaszy, 2007). Ezzel összecsengenek Váradi (2010) eredményei is, aki hat adatközlő felolvasásainak és spontán monológainak összevető elemzése alapján megállapította, hogy a felolvasásokban a néma szünetek előfordulása ritkább, de azok időtartamára a beszédtípus nem gyakorolt matematikailag igazolható hatást. A társalgásokban

a néma szünetek előfordulási gyakorisága a narratívákhoz viszonyítva ritkább, az időtartamuk pedig rövidebb volt (Markó, 2005a). Bóna (2010, 2013b,a) több kutatásban vizsgálta a néma szünetek sajátosságait fiatalok és idősek spontán narratíváiban, társalgásaiban és hallás alapú történet-visszamondásaiban. A három közül a legnehezebb feladatnak a történet-visszamondás bizonyult, ebben volt a legmagasabb a szünetek aránya mindkét életkori csoportban; az időtartamok azonban csak a fiataloknál mutattak szignifikáns növekedést. A különleges beszédtípusok prozódiai elemzése (Menyhárt, 2011) igazolta a beszédtípus és a műfaj meghatározó voltát. A Hegedűs-Archívum adatközlőinek spontán beszédét és különféle műfajú meséit (tündérmese, állatmese, reális mese) elemezve jellegzetes eltéréseket igazoltak a szünettartásban a műfajok között.

Az itt bemutatott kutatások döntő többsége felnőtt vagy idős adatközlők beszédén alapul, a felnőttkor különböző szakaszait azonban általában nem elemzik külön. A fejlődéspszichológia a felnőttkort hagyományosan (legalább) három szakaszra bontja: 19 és 35 éves kor közé tehető a korai felnőttkor, 35 és 60 éves kor közé a középső felnőttkor, míg 60 év felett késői felnőttkorról, illetve időskorról beszélhetünk (vö. Erikson, 1963). A korai felnőttkorra jellemző az egyén társadalomban való beilleszkedése mind közéleti, mind magánéleti szinten; a fizikai teljesítőképesség maximumának elérése; a családalapítás, illetve utódnemzés. Gyakorinak tekinthető továbbá a – hol egyéni, hol külső késztetésre bekövetkező – munkahelyváltás. A középső felnőttkorra az egyén stabilan beilleszkedik a társadalomba, igyekszik megvalósítani egyéni céljait. Ebben az időszakban realizálódnak a karrier- és a családi szerepek, valamint kialakulnak a stabil életmódbeli szokások. Az egyén életét főleg gyakorlatias célok irányítják, ami miatt ez az életszakasz egyfajta produktív periódusnak tekinthető. A késői felnőttkor az öregkorba való átmenet, gyakran válságperiódusként élik meg. Az egyén életében alapvető változások zajlanak, jellemző az érzékszervek romlása, pszichofizikai teljesítőképesség csökkenése, memóriazavarok, nehéz alkalmazkodó képesség, lassuló reakció-készség; valamint szintén erre a periódusra tehető a gyermekek önállósodása, ami szülőkben a feleslegesség érzetét keltheti. Az életkori szakaszok határai szerzőnként, szakterületenként, sőt a vizsgált populáció

földrajzi elhelyezkedése szerint is különbözhetnek; gondoljunk csak arra, hogy még az Európai Unión belül sem egységes a nyugdíjkorhatár. A WHO például az időskorra vonatkozóan a következő felosztást alkalmazza: 50–60-ig áthajlás kora, 60–75-ig idősödés kora, 75–90-ig időskor, 90 év fölött aggkor és 100 év felett matuzsálemi kor (Iván, 2002). Az imént ismertetett felosztásból kiindulva a jelen kutatásban a felnőtt beszélőket két korosztályba soroltuk; fiatal felnőtteknek tekintettük a 20-35 év közötti adatközlőinket, míg középkorúnak a 40-55 év közé tartozó beszélőinket.

A jelen tanulmány célja, hogy egy átfogó elemzést adjon a néma szünetek különböző típusainak előfordulási gyakoriságáról és időtartam-realizációiról az életkor és a beszédtípus függvényében. A kutatásunkban választ kerestünk arra, hogy 1. a beszéd típusa milyen hatással van a korábbi kutatásokban a spontán narratívákra igazolt szünettartási stratégiákra (vö. Gyarmathy, 2017b, 2019; Gyarmathy & Horváth, 2018); 2. kimutatható-e valamilyen különbség a fiatal felnőttek és a középkorú beszélők között az egyes szünettípusok gyakoriságát és időtartamát illetően.

Hipotéziseink szerint 1. a néma szünetek időtartamát és gyakorisági arányát elsődlegesen a szünet típusa, a közlésben betöltött pozíciója és funkciója határozza meg; 2. a különböző beszédtípusokban a beszélők eltérő szünettartási stratégiákat alkalmaznak, amely tetten érhető az időtartam-realizációkban; és 3. a fiatal felnőtt és a középkorú beszélők szünettartási stratégiái tendenciózusan megegyeznek a két beszédtípusban, de gyakorisági és a temporális paraméterekben adódnak közöttük különbségek.

2. Anyag, módszer, kísérleti személyek

A kutatáshoz a BEA spontánbeszéd-adatbázis (Gósy et al., 2012) 40 felvételét elemeztük; 20 narratívát és 20 társalgást. A társalgásokban 3 fő vett részt: az interjúkészítő, a társalgó partner, aki az interjúkészítő kollégája, illetve az adatközlő. Az interjúkészítő és a társalgó partner személye minden felvételen változatlan; mindketten (a felvételek elkészültekor) 28 éves női beszélők. A jelen

kutatásban kizárólag az adatközlő megnyilatkozásait elemeztük. Adatközlőinket két korcsoportból választottuk ki, 10 (5 férfi és 5 nő) 20-35 év közötti fiatal felnőtt, valamint 10 (5 férfi és 5 nő) 40-55 év közötti középkorú felnőtt; mindannyian ép halló, ép értelmű, köznyelvi beszélők. A fiatal felnőttek átlagéletkora 27,7 év, a középkorúaké 45,2 év volt. Az elemzett hanganyagok hossza összesen 6 óra 38 perc 32 másodperc volt, ebből a spontán narratíva 1 óra 41 perc 42 másodpercet (átlag: 5 perc/fő), míg a társalgás 4 óra 56 perc 50 másodpercet (átlag: 15 perc/fő) tett ki. A felvételek hossza a két életkori csoportban nem tért el jelentősen egyik beszédtípus esetében sem. A teljes hanganyagban összesen 4880 néma szünetet annotáltunk, amelyből 1617 a narratívákból (átlag: 81db/fő), 3263 a társalgásokból (átlag: 163 db/fő) származott (1. táblázat). A szünetek számát a beszédidőre vetítve a narratívákban átlagosan 4 (15,8 db/perc), míg a társalgásokban 5 másodpercenként (11 db/perc) tartottak néma szünetet az adatközlők. A szünetek összidőtartama mintegy 37 perc volt, ami a teljes beszédidő 9,2%-át tette ki. A fiatal felnőttek narratíváiban 5 (13 db/perc), társalgásaiban 6 másodpercenként (11,8 db/perc) követték egymást a néma szünetek, míg a középkorúaknál valamivel gyakrabban, a narratívákban 3 (19,4 db/perc), a társalgásokban 5 másodpercenként (11,1 db/perc).

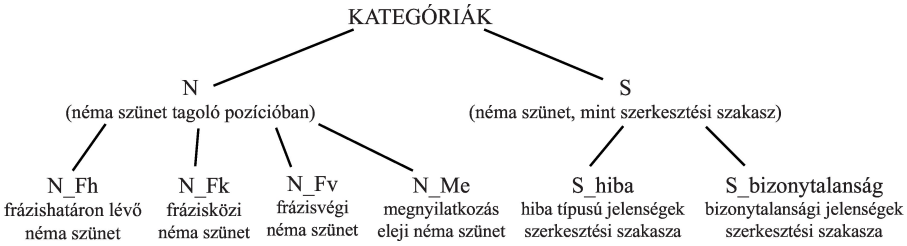
1. táblázat. a néma szünetek adatainak és a beszédidő alakulása a korcsoportok és a beszédtípusok alapján.

korcsoport	**beszédtípus**	**darab**	**átlag (ms)**	**szórás (ms)**	**beszédidő**
20–35 év	narratíva	721	529	432	55 m 26 s
	társalgás	1432	469	369	2 h 12 m 07 s
40–55 év	narratíva	896	442	413	46 m 16 s
	társalgás	1831	416	343	2 h 44 m 43 s

Az annotálást és a szünetidőtartamok meghatározását a Praat szoftver 6.1.09-es verziójával (Boersma & Weenink, 2018) manuálisan (a megelőző lexéma utolsó hangjának végétől a követő lexéma első hangjának kezdetéig) végeztük. A néma szünetek meghatározásakor nem alkalmaztunk minimális szünetidőtartamot, tehát minden, a hangszínképen detektálható néma szünetet annotáltunk. Mivel

a zöngétlen obstruensekkel kezdődő beszédszakaszok esetében lehetetlen megkülönböztetni, hogy meddig tart a néma szünet és hol kezdődik a beszédhang zárszakasza, ezekben az esetekben (korábbi kutatási eredményekre alapozva: vö. Gráczi, 2013, részben a Trouvain et al., 2016 által alkalmazott módszernek megfelelően) egy 50 ms-os állandó értéket határoztuk meg a zöngétlen felpattanó zárhangok néma fázisaként.

A szünetek osztályozását egy korábban általunk kidolgozott kategóriarendszer alapján végeztük (Gyarmathy, 2017a); mely elsődlegesen a tagolást szolgáló és a megakadásjelenségek szerkesztési szakaszaként realizálódó néma szüneteket különbözteti meg egymástól (1. ábra). Az előbbieket **N**-nel, az utóbbiakat **S**-sel jelöli. Megakadásjelenségek szerkesztési szakaszaként megvalósulónak csak azok a szünetek tekinthetőek, amelyeknél a felszíni szerkezetben detektálható az adott jelenség (hibák esetén annak javítása is). A tagoló néma szünetek a közlésbeli pozíciójuk alapján csoportosíthatóak, míg a szerkesztési szakaszokat aszerint, hogy hiba típusú (**S_hiba**: *ennyi pénzér amennyiér* **S_hiba** *amennyibe egy békávé bérlet kerül*), vagy bizonytalansági megakadásokhoz (**S_bizonytalanság**: *nagyon fontos hogy* **S_bizonytalanság** *hogy mi veszi körül*) köthetők-e. A tagolási pozícióban megjelenő néma szünetek (**N**) a közlésben elfoglalt helyük szerint lehetnek a megnyilatkozás eleji (**N_Me**) néma szünetek, amikor beszélőváltáskor az aktuális beszélő belekezd a közlésbe; ilyenkor a szünetet legfeljebb egy töltelékszó, vagy diskurzusjelölő előzi meg: Felvételvezető: *De most már annyira megemelték a bérlet árát is.* Adatközlő: *Hát* **N_Me** *relatív, mert ha azt számolod, hogy...* A frázis-határon lévő (**N_Fh**) néma szünetek közé tartoznak azok, amelyek az elemi mondatok határán, gyakran kötőszó előtt vagy után helyezkednek el: *Személyes hobbinak is tekintem, és* **N_Fh** *szerencsére vannak is lehetőségeim ebben a szakmában.* Frázisközi (**N_Fk**) szünetként jelölendők azok, amelyek grammatikai egységen belül, annak struktúráját megtörve fordulnak elő: *Egy havi nyolcezer forintos kiadás nem nagy* **N_Fk** *összeg.* Végül frázisvégi (**N_Fv**) szünetként azonosítandók a szemantikai, szintaktikai és grammatikai egységeket (írásban mondatokat) lezáró néma szünetek, amely után a beszélő új szintaktikai egységet kezd, gyakran



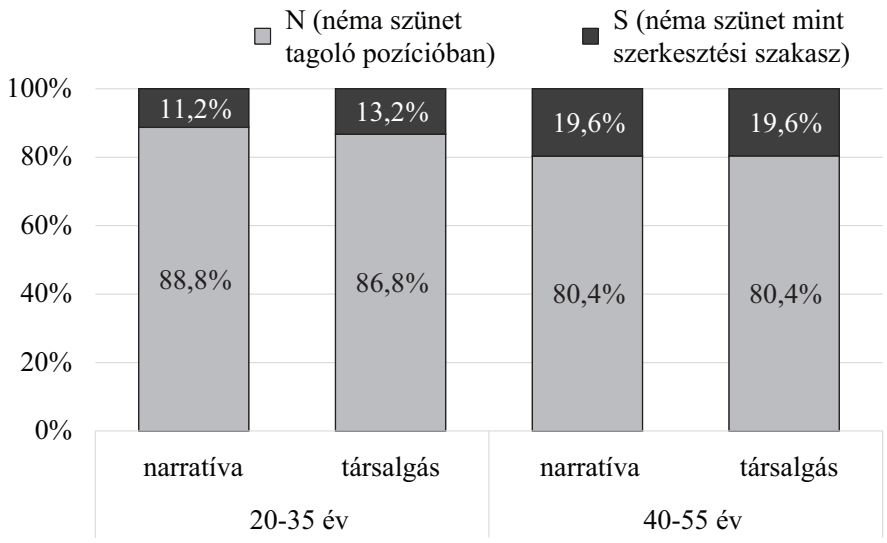
1. ábra. A néma szünetek kategóriarendszere Gyarmathy (2017b) alapján

egy új gondolati egységgel folytatja a közlését: *Előre nem közölt kritériumok alapján osztályoztak le.* **N_Fv** *Egyébként a szakkal kapcsolatban azt gondolom, hogy...* Mivel a frázisvégi és a frázishatáron lévő szünetek elkülönítése a spontán beszéd esetén problémásnak hathat, ezért a kategorizálás során szigorú kritériumokat követtünk. Ezek alapján csak az olyan szintaktikai egységeket lezáró néma szüneteket azonosítottuk frázisvégiként, amelyek esetében a követő vagy nem kötőszóval kezdődik, és/vagy teljesen új gondolati egységet vezet be. Azokat az eseteket, ahol a besorolás nem volt egyértelmű, nem vontuk be az elemzésünkbe.

A statisztikai elemzéseket az SPSS 20-as verziójával végeztük. Az adatok eloszlását binominális nemparaméteres teszttel és chi-négyzet goodness of fit teszttel, az előfordulási gyakoriságot általános lineáris modellel (GLM) vizsgáltuk; a részletes statisztikai elemzésekhez általános lineáris kevert modellt alkalmaztunk (GLMM), az adatok páronkénti összehasonlítását a modell részét képező pairwaise contrast-tal végeztük. Független változóink az egyes szünettípusok (az életkor és a beszédtípus függvényében), függő változóink az időtartamok voltak, random faktorként a beszélőket vettük fel.

3. Eredmények

Az elemzett néma szünetek átlagos időtartama és előfordulási gyakorisága korosztályonként és beszédtípusonként is eltér. A fiatal felnőtteknél adatolt néma szünetek átlagosan hosszabbak voltak (átlag: 488,7 ms, SD: 391,8 ms), mint a középkorúaknál előfordulók (átlag: 424,5 ms, SD: 367,6 ms). A narratívákban



2. ábra. A tagoló néma szünetek (N) és a szerkesztési szakaszok (S) egymáshoz viszonyított aránya korosztályok és a beszédtípusok szerint

a beszélők átlagosan hosszabb szüneteket tartanak (átlag: 480,4 ms, SD: 423,4 ms), mint a társalgásokban (átlag: 439,1 ms, SD: 355,4 ms), és ez a tendencia az életkortól függetlennek bizonyult.

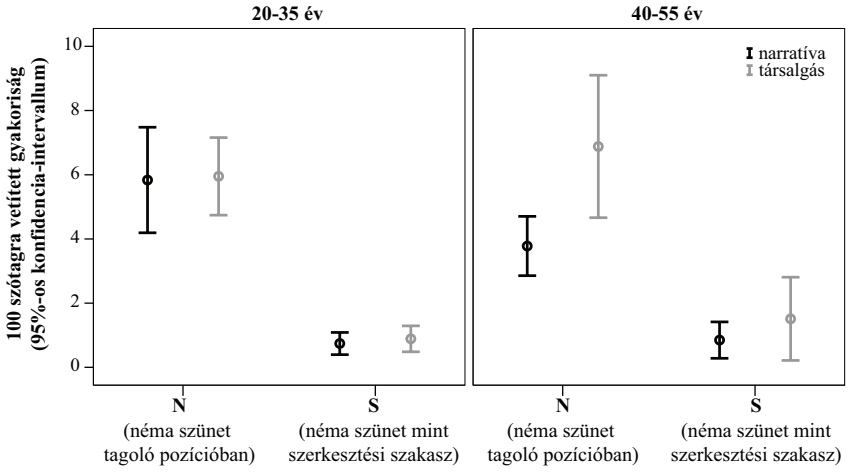
A teljes beszédanyagban adatolt 4880 néma szünet 17%-a szerkesztési szakaszként realizálódott, míg 83%-a tagoló pozícióban fordult elő. A beszédtípusok nem befolyásolták jelentősen az arányokat; a narratívákban a szünetek 16%-a realizálódott szerkesztési szakaszként, 84%-a tagoló pozícióban, míg a társalgásokban ez az arány 17% vs. 84% volt. A két korcsoportban ugyan hasonló arányt találtunk mind a narratív, mind a társalgásos beszédrészek esetében, a középkorú felnőtteknél azonban valamivel nagyobb arányú a szerkesztési szakaszként megjelenő néma szünetek előfordulása (2. ábra). Az adatok eloszlását vizsgáló binominális nemparaméteres teszt minden esetben igazolta, hogy az adatok nem véletlenszerűen rendeződnek kategóriákba ($p \leq 0,001$).

A beszédidőhöz viszonyított előfordulást elemezve valamivel árnyaltabb képet kaphatunk a két fő kategória előfordulási gyakoriságáról. A fiatal felnőttek narratíváiban 5 másodpercenként fordultak elő tagoló néma szünetek (11,5

db/perc), míg szerkesztési szakaszok mindössze 41 másodpercenként (1,5 db/perc); a társalgásaikban tagoló szünetek 6 másodpercenként (9,4 db/perc), míg szerkesztési szakaszok 42 másodpercenként (1,4 db/perc). A középkorú beszélőknél a tagoló néma szüneteket tekintve hasonló tendenciát találunk; a narratívákban 4 (15,6 db/perc), a társalgásokban 7 másodpercenként (8,9 db/perc) realizálódtak. A szerkesztési szakaszok előfordulása azonban a fiataloknál tapasztaltakhoz képest közel kétszer olyan gyakori; a narratívákban 16 (3,8 db/perc), a társalgásokban 28 másodpercenként (2,2 db/perc) voltak adatolhatók. Ez alapján elmondható, hogy a középkorú felnőttek gyakrabban küzdenek tervezési diszharmóniával.

A tempóbeli különbségek kiküszöbölésének céljából a szünetek előfordulását 100 szótagra vetítve is elemeztük. Az két vizsgált korcsoport között nem találtunk jelentős eltérést (fiatal felnőttek 3,4 szünet/100 szótag, középkorúak 3,3 szünet/100 szótag), azonban a szünettípus és a beszédtípus meghatározónak bizonyult (3. ábra). A fiatal felnőttek narratíváiban a tagoló néma szünetek 100 szótagra vetített aránya 5,8, társalgásaiban 6 volt; a középkorú beszélők narratíváiban 3,8, társalgásaiban 6,9. A szerkesztési szakaszok előfordulása 100 szótagra vetítve a fiatalok narratíváiban 0,7, társalgásaiban 0,9; míg a középkorúak narratíváiban 0,8, társalgásaiban 1,5 volt. A fiatal beszélőknél tehát nem tapasztalható lényeges eltérés a két beszédtípus között, a középkorú adatközlők társalgásaiban ezzel szemben lényegesen gyakoribb a szünetek előfordulása. Az adatok statisztikai elemzése igazolta, hogy a szünetek előfordulását a szünet típusa erős szignifikáns hatást gyakorolva egyértelműen meghatározza [GLM: $F(1, 80) = 142,652$ $p < 0,001$ $\eta^2 = 0,665$], valamint a beszéd típusa [GLM: $F(1, 80) = 6,789$ $p = 0,011$ $\eta^2 = 0,086$] szintén hatással van rá. Az időtartamra és a 100 szótagra vetített gyakorisági mutatók közt tapasztalható eltérések az artikulációs tempó, valamint a szünetidőtartamok különbségeiből fakadhatnak.

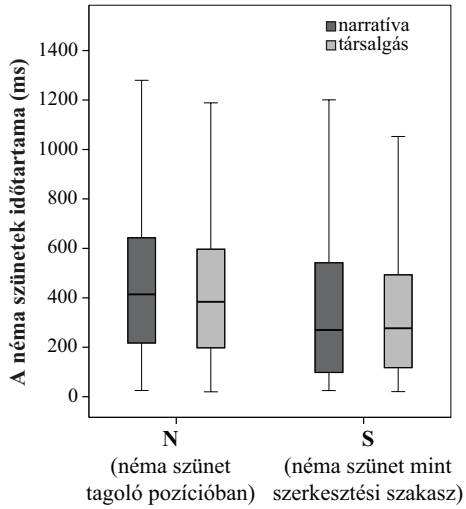
A szünetek időtartamát először a beszédtípus függvényében elemeztük. A tagoló néma szünetek mind a narratívák (átlag: 499,5 ms, SD: 432,3 ms), mind a társalgások (átlag: 457,4 ms, SD: 362,5 ms) esetében hosszabb időtartammal realizálódtak, mint a tervezési nehézségek feloldását szolgáló szerkesztési szaka-



3. ábra. A tagoló néma szünetek (N) és a szerkesztési szakaszok 100 szótagra vetített gyakorisága beszédtípusok szerint a kért korcsoportban

szok (narratíva átlag: 379,8 ms, SD: 356,8 ms; társalgás átlag: 348,3 ms, SD: 301,8 ms). Az interkvartilis tartományok alapján a csoportok homogenitása csaknem azonosnak mondható, a társalgásokban adatolt néma szünetek azonban típustól függetlenül rövidebb időtartammal realizálódtak (4. ábra). Az adatainkra épített általános lineáris kevert modell mindkét beszédtípus esetében megerősítette, hogy a szünettípus meghatározza az időtartamot [narratíva: $F(1, 4850) = 27,223\ p < 0,001$; társalgás: $F(1, 4850) = 26,977\ p < 0,001$].

Ha az időtartamok elemzését újabb változóval, a korcsoporttal bővítjük, a kép tovább árnyalódik. Mind a fiatal felnőttek, mind a középkorú beszélők esetében megállapítható, hogy a tagoló néma szünetek jóval hosszabb időtartamúak, mint a szerkesztési szakaszok (2. táblázat). A szélsőértékek elemzése alapján elmondható, hogy a legrövidebb és a leghosszabb szünetet is fiatal beszélők által tartott tagoló néma szünetek közé tartozott, az előbbi a társalgásaikban, az utóbbi a narratíváikban fordult elő. A legszélesebb tartományban az ugyanezen beszélők narratíváiban tartott tagoló szünetek realizálódtak, míg a legszűkebb intervallumban a középkorú beszélők narratíváiban előforduló tagoló szünetek. Általánosságban elmondható, hogy a tagoló szünetek a szerkesztési szakaszok-



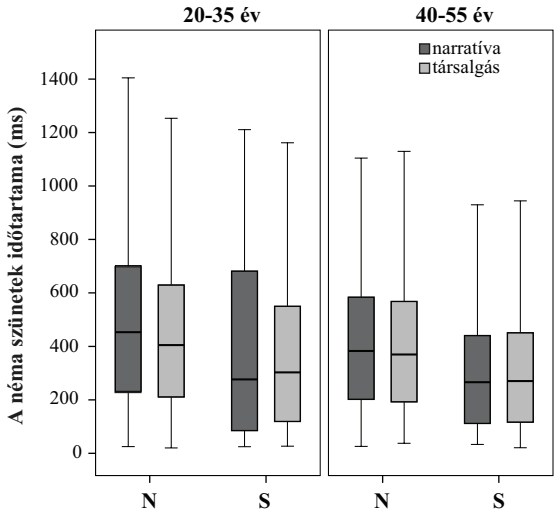
4. ábra. A tagoló néma szünetek (N) és a szerkesztési szakaszok (S) időtartama a beszédtípus függvényében (median és interkvartilis tartomány)

hoz képest nem csupán hosszabbak voltak, de az adatok szélesebb tartományban szóródtak.

2. táblázat. A tagoló pozíciójú (N) és a szerkesztési szakaszként (S) realizálódó néma szünetek időtartam-értékei a beszédtípus és a korcsoport függvényében.

korcsoport	beszédtípus	szünettípus	átlag (ms)	szórás (ms)	minimum (ms)	maximum (ms)
20-25 év	**narratíva**	**N**	538,9	429,8	25,3	4362,9
		S	446,6	438,7	24,8	2585,5
	társalgás	**N**	478,6	366,3	19,9	3322,3
		S	403,2	378,3	26,6	2372,9
40-55 év	**narratíva**	**N**	464,4	431,8	25,9	3715,3
		S	349,1	308,5	33,3	1584,9
	társalgás	**N**	439,5	358,5	37,44	3356,1
		S	319,4	247,9	20,8	1794,7

Az 5. ábrán az is jól látszik, hogy a középkorú beszélők által tartott szünetek időtartamértékei sokkal kiegyenlítettebbek, mint a fiatalok által tartottak; az előbbieknél az interkvartilis tartományok csaknem azonosnak mondhatóak, míg a fiataloknál különösen a szerkesztési szakaszok esetében nagyobb eltérések láthatóak. A középkorú beszélők tagoló néma szünetei és szerkesztési szakaszai



5. ábra. A tagoló néma szünetek (N) és a szerkesztési szakaszok (S) időtartama a beszédtípus függvényében (median és interkvartilis tartomány)

is közel azonos átlagos időtartammal realizálódtak a beszédtípustól függetlenül, míg a fiataloknál 40-60 ms-nyi eltéréseket figyelhetünk meg mindkét szünettípus átlagában a vizsgált beszédtípus szerint. A statisztikai elemzések azonban csupán a szünettípus tekintetében igazoltak szignifikáns különbséget minkét korcsoportnál mindkét beszédtípusban: GLMM: fiatal narratíva: $F(1, 4850) = 10,564$ $p = 0{,}001$; fiatal társalgás: $F(1, 4850) = 12,929$ $p < 0,001$; középkorú narratíva: $F(1, 4850) = 18,684$ $p < 0,001$; középkorú társalgás: $F(1, 4850) = 14,047$ $p < 0,001$.

Elemzésünket a néma szünetek alkategóriái mentén folytatva elsőként az adatok eloszlását vizsgáltuk külön a tagoló néma szünetek és a szerkesztési szakaszok alcsoportjaira vonatkoztatva. A tagoló néma szünetek a leggyakrabban, az esetek 51%-ában a közlés elemi szintaktikai egységeit elválasztva, frázishatáron (**N_Fh**) fordultak elő. 26%-uk a grammatikai és szintaktikai struktúrát megtörve, frázisközi helyzetben (**N_Fk**) realizálódott, míg szintaktikai, szemantikai és grammatikai egységet alkotó megnyilatkozásrészek lezárásaként, frázisvégen (**N_Fv**) 22%-uk jelent meg. A megnyilatkozás-kezdő (**N_Me**) pozíció csupán az esetek 1%-át jellemezte. Az adatok eloszlását vizsgáló χ^2 goodness of fit teszt

egyértelműen igazolta, hogy az eloszlás nem véletlenszerű, tehát az adatok szabályszerűen rendeződnek alkategóriákba [$\chi^2(3) = 2004,536$; $p \leq 0,001$]. A szerkesztési szakaszként megvalósuló néma szünetek döntő többsége (83%) a beszélő bizonytalanságából származó jelenségekhez kapcsolódott (**S_bizonytalanság**), míg egyhatod részük (17%) szolgált csupán valamely hiba javítására (**S_hiba**). A binominális nemparaméteres teszt eredménye alapján elmondható, hogy az adatok alkategóriákba rendeződése szabályszerű mintázatot követ ($p \leq$0,001). Az életkor és a beszédtípus függvényében az adatok eloszlása némiképp módosul, de a tendencia megegyezik az imént bemutatottal. A fiatalok narratíváiban a frázishatáron előforduló néma szünetek (**N_Fh**) aránya 51% volt, ezt követték a frázisköziek (**N_Fk)** 28%-kal és a frázisvégiek (**N_Fv**) 20%-kal, míg a megnyilatkozások elején (**N_Me**) csupán a néma szünetek 1%-a realizálódott. Aχ^2 goodness of fit teszt alapján az adatok szabályszerű mintázat alapján rendeződnek kategóriákba [χ^2 (3) = 322,338; $p \leq 0,001$]. Ugyanezen beszélők társalgásaiban 51% volt a frázishatáron lévő szünetek (**N_Fh**) aránya, 27% a frázisközieké (**N_Fk**), 21% a frázisvégieké (**N_Fv**), és 1% a megnyilatkozás elejieké **(N_Me**). A statisztikai elemzések szerint az adatok eloszlása nem véletlenszerű [$\chi^2(3) = 620,102$; $p \leq 0,001$]. A szerkesztési szakaszok két alkategóriájának egymáshoz viszonyított aránya a fiatal beszélők narratíváiban és társalgásaiban gyakorlatilag megegyezett. Az előbbiben a bizonytalanságokhoz köthetők (**S_bizonytalanság**) aránya 79%, a hibáké (**S_hiba**) 21%, míg az utóbbiban ugyanez 78% vs. 21%. A binominális nem paraméteres teszt szerint az adatok eloszlása egyik beszédtípusban sem véletlenszerű ($p \leq 0,001$). A középkorú beszélőknél szintén hasonló tendenciákat találunk. A narratívában a frázishatáron lévő szünetek (**N_Fh**) előfordulása 49%, a társalgásban 51%; a frázisközieké (**N_Fk**) 25% mindkét beszédtípusban; a szintaktikai struktúrákat lezáró frázisvégi szüneteké **(N_Fv**) 24% a narratívában, a társalgásban 23%; míg a megnyilatkozás eleji szüneteké (**N_Me**) egyaránt 1%. A χ^2 goodness of fit teszt a narratíva [$\chi^2(3) = 329,189$; $p \leq 0,001$] és a társalgás [$\chi^2(3) = 738,343$; $p \leq 0,001$] esetében is igazolta az adatok szabályszerű eloszlásmintázatát. A középkorú beszélők narratíváiban a bizonytalansági megakadásokhoz kapcsolódó

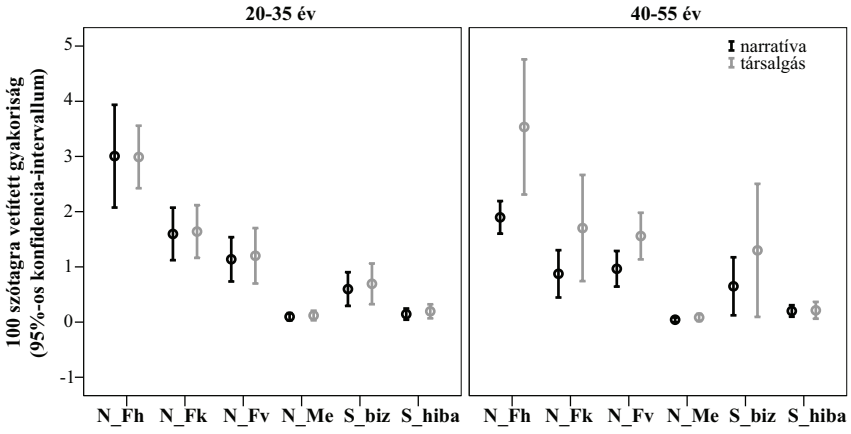
szerkesztési szakaszok (**S_bizonytalanság**) aránya 82% volt, a társalgásokban 88%; a hibajelenségek szerkesztési szakaszai (**S_hiba**) a narratívákban 18%-ot, a társalgásokban 12%-ot tettek ki. A binominális nem paraméteres teszt ebben a korosztályban is mindkét beszédtípusra igazolta az adatok eloszlásának szabályszerűségét ($p \leq 0,001$).

Az egyes alkategóriákba tartozó néma szünetek előfordulását a beszédidőhöz viszonyítva is elemeztük. A narratív beszédrészekben mindkét korosztálynál minden egyes altípusban magasabb elemszámot adatoltunk percenként, mint a társalgásokban (3. táblázat). A fiatal felnőttek narratíváiban a leggyakoribbak a frázishatáron lévő szünetek voltak, ezt követték a frázisköziek, majd a frázisvégiek, míg a megnyilatkozás elejiek csupán 6 perc 10 másodpercenként követték egymást. A bizonytalansági megakadások szerkesztési szakaszai több mint háromszor olyan gyakran fordultak elő, mint a hibajenségekhez kapcsolódók. Ugyanezen beszélők társalgásaiban a narratívákban tapasztaltakhoz hasonló tendencia figyelhető meg. A középkorú beszélők narratíváiban az egyes kategóriák valamivel gyakoribb előfordulást mutatnak a fiatalokéhoz képest, az előfordulási gyakoriság alapján felállítható sorrend azonban megegyezik a fent ismertetettel. A társalgásokat itt is alacsonyabb előfordulási arány jellemzi, a tendencia azonban csupán annyiban eltérést, hogy a bizonytalansági megakadások szerkesztési szakaszai mintegy hatszor gyakoribbak, mint a hibajelenségekéi.

A tempóértékekből adódó egyéni különbségek kiküszöbölése céljából az alkategóriák előfordulási gyakoriságát 100 szótagra vetítve is elemeztük. A két vizsgált korosztályban hasonló tendencia rajzolódik ki, ahogy az a 6. ábráról is leolvasható, a beszédtípus és a szünetkategóriák szerint azonban jelentős eltérések figyelhetőek meg (az értékeket lásd a 3. táblázatban). A fiatal felnőttek narratív beszédrészében a frázishatáron tartott néma szünetek (**N_Fh**) voltak a leggyakoribbak, ezt követték a grammatikailag, szintaktikailag inkorrekt helyen előforduló frázisközi (**N_Fk**) szünetek, majd a lezáró pozícióban előforduló frázisvégi (**N_Fv**) szünetek. A megnyilatkozás eleji szünetek (**N_Me**) megjelenése sporadikusnak mondható. A tervezési diszharmóniák feloldását elősegítő néma szünetek közül a beszélő bizonytalanságából adódó jelenségekhez köthetők (**S_bizonytalanság**) közel négyszer olyan gyakoriak voltak, mint a hibajavításra (**S_hiba**) szolgálók. A középkorú beszélők narratíváiban a tagoló néma szünetek 100 szótagra vetített aránya alatta marad a fiataloknál leírt értékeknek. A leggyakoribbak itt is a frázishatáron tartott néma szünetek, ezt követték a frázisvégi szünetek, illetve a frázisközi szünetek, míg a megnyilatkozás eleji szünetek itt is csak szórványos előfordulást mutattak. A szerkesztési szakaszok közül mind a bizonytalansági, mind a hiba típusú jelenségekhez kapcsolódók közel azonos arányban voltak adatolhatók, mint a fiatalabb beszélőknél. A társalgásokban a fiatalok megnyilatkozásaiban adatolt szünetek gyakorisága jórészt nem különbözik a narratív beszédrészben tapasztaltaktól, míg a középkorúakat gyakoribb szünettartás jellemezte. A fiatal beszélők társalgásaiban a leggyakoribb a frázishatáron előforduló néma szünet volt, ezt követték a frázisközi és a frázisvégi szünetek, majd a megnyilatkozás elejiek. A középkorú adatközlőknél – a narratív beszédrészben tapasztaltakkal ellentétben – a frázisvégi szünetek valamivel ritkább előfordulást mutattak, mint a frázisköziek. Megnyilatkozás eleji néma szünet az ő esetükben még a fiatalokénál is ritkábban volt adatolható. A szerkesztési szakaszok altípusainak gyakoriságában a hibajelenségeket illetően – a narratív beszédrészhez hasonlóan – itt sem találtunk jelentős eltérést a két korosztály között. A bizonytalansági megakadásokhoz kapcsolódó szünetek aránya azonban a középkorúaknál majdnem kétszerese a fiatal beszélőknél

3. táblázat. A néma szünetek alkategóriáinak előfordulása a beszédidőhöz viszonyítva, és 100 szótagra vetítve a korcsoport és a beszédtípus szerint

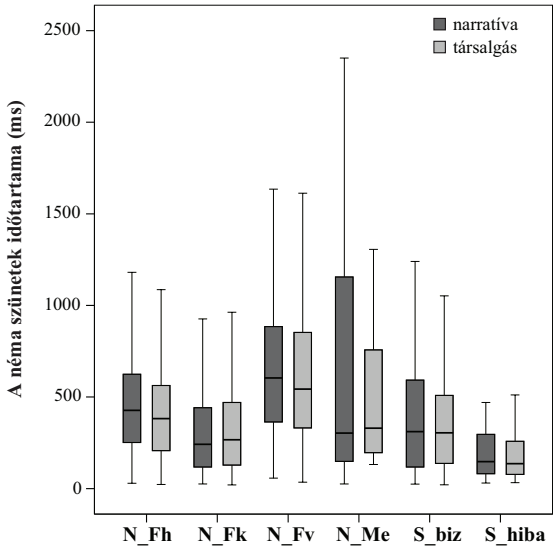
korcsoport	**beszédtípus**	**szünettípus**	**percenkénti előfordulás (db/perc)**	**előfordulási gyakoriság**	**100 szótagra vetített gyakoriság (db)**
20-35 év	**narratíva**	**N_Fh**	5,88	10 sec	3,01
		N_Fk	3,17	19 sec	1,6
		N_Fv	2,33	26 sec	1,1
		N_Me	0,16	6 min 10 sec	0,1
		S_biz	1,15	52 sec	0,6
		S_hiba	0,31	3 min 16 sec	0,14
	társalgás	**N_Fh**	4,8	13 sec	2,99
		N_Fk	2,51	24 sec	1,64
		N_Fv	1,96	31 sec	1,2
		N_Me	0,14	6 min 57 sec	0,12
		S_biz	1,12	54 sec	0,69
		S_hiba	0,13	3min 13 sec	0,2
40-55 év	**narratíva**	**N_Fh**	7,67	8 sec	1,9
		N_Fk	3,93	15 sec	0,87
		N_Fv	3,72	16 sec	0,97
		N_Me	0,24	4 min 12 sec	0,04
		S_biz	3,11	19 sec	0,65
		S_hiba	0,69	1 min 27 sec	0,2
	társalgás	**N_Fh**	4,58	13 sec	3,53
		N_Fk	2,22	27 sec	1,7
		N_Fv	2,03	30 sec	1,56
		N_Me	0,12	8 min 40 sec	0,08
		S_biz	1,91	31 sec	1,3
		S_hiba	0,26	3 min 50 sec	0,21



6. ábra. A néma szünetek 100 szótagra vetített gyakorisága a két korcsoportban az alkategóriák szerint

adatoltaknak. A két beszédtípus között a fiataloknál ebben az esetben sincs nagy különbség, a középkorú beszélők társalgásaiban azonban csaknem minden szünettípus nagyobb arányú előfordulást mutat. A statisztikai elemzések szerint a szünetek előfordulását a szünet típusa [GLM: $F(5, 240) = 71,847$, $p < 0,001$; $\eta^2 = 0,625$] és a beszédtípus [GLM: $F(1, 240) = 11,732$, $p = 0,001$; $\eta^2 = 0,052$] befolyásolta szignifikánsan.

A néma szünetek alkategóriáinak időtartam-realizációit először a beszédtípus mentén elemeztük. A 7. ábrán látható, hogy mind a narratívákban mind a társalgásokban hasonló mintázat rajzolódik ki, a kétféle beszédtípusban tehát az egyes szünetkategóriák időtartama azonos tendenciát követ. Általánosságban elmondható, hogy társalgásokban az egyes szünettípusok rövidebb időtartammal valósulnak meg, de két beszédtípus közti különbség nem szignifikáns. A narratív beszédrészben a megnyilatkozás eleji néma szünetek realizálódtak a leghosszabban (átlag: 767,7 ms, SD: 965,5 ms), míg a társalgásokban a lezáró funkcióval bíró frázisvégiek (átlag: 643,9 ms, SD: 444,5 ms). A második leghosszabb időtartam-realizáció a narratívákban a frázisvégi szüneteket (átlag: 715,1 ms, SD: 531,8 ms) jellemezte, míg a társalgásokban a megnyilatkozás ele-



7. ábra. A szünettípusok időtartama a beszédtípus szerint (medián és interkvartilis tartomány)

jieket (átlag: 599,7 ms, SD: 635,8 ms). A rangsor ezután a két beszédtípusban már azonos mintázatot követ. A harmadik leghosszabb átlagos időtartammal a frázishatáron lévő néma szünetek (narratíva átlag: 485,6 ms, SD: 379,9 ms; társalgás átlag: 434,7 ms, SD: 322,6 ms), a negyedikkel a bizonytalansági megakadások szerkesztési szakaszaihoz köthetők valósultak meg (narratíva átlag: 411,2 ms, SD: 368,8 ms; társalgás átlag: 368,9 ms, SD: 303,8 ms). Ezt követték a frázisközi szünetek (narratíva átlag: 329,6 ms, SD: 276,8 ms; társalgás átlag: 335,7 ms, SD: 262,6 ms), végül a legrövidebbek a hiba típusú jelenségek szerkesztési szakaszai voltak (narratíva átlag: 246,6 ms, SD: 264,6 ms; társalgás átlag: 235,1 ms, SD: 264,8 ms). A statisztikai elemzések szerint a kétféle beszédtípus között nincs lényegi különbség a néma szünetek hosszában, a szünetek típusa azonban mindkét beszédtípusban egyértelműen meghatározza azok időtartamát [GLMM: narratíva: $F(4, 4850) = 69,358$, $p < 0,001$; társalgás: $F(4, 4850) = 52,141$, $p < 0,001$].

Elvégeztük az adatok páronkénti összehasonlítását a beszédtípus mentén. Az elemzések szerint a narratívákban 11 esetben adódott szignifikáns különbség

az egyes csoportpárok között, míg a társalgásokban 12 esetben (4. táblázat). A legerősebb szignifikáns különbség a frázisközi (**N_Fk**) és a frázisvégi (**N_Fv**) szünetek, illetve a frázishatáron lévő (**N_Fh**) és a frázisvégi (**N_Fv**) szünetek között volt tapasztalható a narratívákban, míg a társalgásokban a frázishatáron megjelenő és a frázisvégi szünetek, illetve frázishatáron lévő és a frázisköziek között.

Az elemzésbe egy további változót, a korcsoportokat bevonva megállapítható, hogy a fiatal és a középkorú beszélők között nincs lényegi különbség az egyes szünettípusok időtartamát tekintve. Az életkorok mentén elvégzett statisztikai elemzések alapján megállapítható, hogy a szünetkategóriák mindkét vizsgált korosztályban meghatározzák a szünetek időtartamát (20-25 éves: $F(4, 4850) = 50,388$, $p < 0,001$; 40-55 éves: $F(4, 4850) = 64,722$, $p < 0,001$). Az adatok páronkénti összevetését a két vizsgált korcsoport szerint is elvégeztük. Az életkorok mentén végzett elemzések szerint a narratívákban és a társalgásokban egyaránt 11 esetben adódott szignifikáns különbség az egyes kategóriapárok között (5. táblázat). A legerősebb szignifikáns különbség mind a fiatal, mind a középkorú beszélőknél frázisközi (**N_Fk**) és a frázisvégi (**N_Fv**) szünetek, illetve a frázishatáron (**N_Fh**) lévő és a frázisvégi (**N_Fv**) szünetek között volt tapasztalható.

A 8. ábrán jól látszik, hogy a néma szünetek altípusainak időtartam-realizációi mindkét korosztály mindkét elemzett beszédtípusában azonos mintázatot rajzolnak ki. Sem az életkor, sem a beszédtípus nem befolyásolja matematikailag igazolhatóan a néma szünetek időtartamát. A statisztikai elemzések azonban mindegyik korcsoport narratíváiban és társalgásaiban is igazolták, hogy a néma szünet típusa szignifikáns hatást gyakorol az időtartamukra: GLMM: 20-35 éves narratíva: $F(4, 4850) = 22,418$, $p < 0,001$; 20-35 éves társalgás: $F(4, 4850) = 32,396$, $p < 0,001$; 40-55 éves narratíva: $F(4, 4850) = 31,841$ $p < 0,001$; 40-55 éves társalgás: $F(4, 4850) = 37,009$, $p < 0,001$.

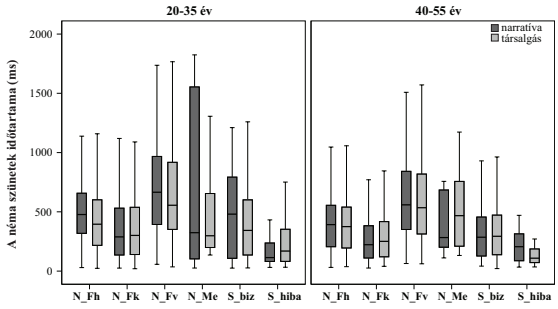
A fiatal felnőtt beszélők narratíváiban és társalgásaiban is egyaránt a frázisvégi néma szünetek realizálódtak átlagosan a leghosszabb időtartammal. Ezeknél alig valamivel rövidebbek a megnyilatkozás eleji néma szünetek mindkét

4. táblázat. A páronkénti összehasonlítás eredménye a beszédtípusok szerint

beszédtípus	szünettípus	t-érték	szignifikancia értéke
narratíva	**N_Fh – N_Fk**	6,346	0,000
	N_Fh – N_Fv	12,288	0,000
	N_Fh – N_Me	2,659	0,008
	N_Fh – S_biz	3,956	0,000
	N_Fh – S_hiba	2,659	0,008
	N_Fk – N_Fv	15,923	0,000
	N_Fk – N_Me	4,397	0,000
	N_Fk – S_biz	5,363	0,000
	N_Fk – S_hiba	4,397	0,000
	N_Me – S_biz	3,059	0,002
	S_biz – S_hiba	3,059	0,002
társalgás	**N_Fh – N_Fk**	6,579	0,000
	N_Fh – N_Fv	9,335	0,000
	N_Fh – N_Me	3,144	0,002
	N_Fh – S_biz	4,373	0,000
	N_Fh – S_hiba	3,144	0,002
	N_Fk – N_Fv	3,718	0,000
	N_Fk – N_Me	4,922	0,000
	N_Fk – S_biz	5,812	0,000
	N_Fk – S_hiba	4,922	0,000
	N_Fv – S_biz	2,100	0,036
	N_Me – S_biz	3,111	0,002
	S_biz – S_hiba	3,111	0,002

5. táblázat. A páronkénti összehasonlítás eredménye a korcsoportok szerint

korcsoport	szünettípus	t-érték	szignifikancia értéke
	N_Fh – N_Fk	6,401	0,000
	N_Fh – N_Fv	9,359	0,000
	N_Fh – N_Me	2,170	0,030
	N_Fh – S_biz	3,966	0,000
	N_Fh – S_hiba	2,170	0,030
20-35 évesek	**N_Fk – N_Fv**	13,531	0,000
	N_Fk – N_Me	3,918	0,000
	N_Fk – S_biz	5,333	0,000
	N_Fk – S_hiba	3,918	0,000
	N_Me – S_biz	3,640	0,000
	S_biz – S_hiba	3,640	0,000
	N_Fh – N_Fk	6,258	0,000
	N_Fh – N_Fv	11,579	0,000
	N_Fh – N_Me	3,698	0,000
	N_Fh – S_biz	4,374	0,000
	N_Fh – S_hiba	3,698	0,000
40-55 évesek	**N_Fk – N_Fv**	15,187	0,000
	N_Fk – N_Me	5,433	0,000
	N_Fk – S_biz	5,832	0,000
	N_Fk – S_hiba	5,433	0,000
	N_Me – S_biz	2,347	0,019
	S_biz – S_hiba	2,347	0,019



8. ábra. A szünettípusok időtartama a beszédtípus és a korcsoportok szerint (medián és interkvartilis tartomány)

beszédtípusban; ezt követik a frázishatáron tartott szünetek, míg a tagoló néma szünetek közül a legrövidebb időtartam a frázisközi szüneteket jellemezte. A szerkesztési szakaszok közül mind a narratívákban, mind a társalgásokban a beszélő bizonytalanságát jelző megakadásokhoz kapcsolódók voltak a hosszabbak. A középkorú beszélőknél hasonló rangsort találunk; a társalgásokban a szintaktikai és szemantikai egységeket lezáró frázisvégi szünetek a leghosszabbak, ezeket követik a megnyilatkozás eleji néma szünetek, a narratívákban ezzel szemben az utóbbiaknál adatoltuk átlagosan a leghosszabb időtartamot. A rangsor első két helyétől eltekintve a mintázat már mindkét beszédtípusban megegyezik a fiatal felnőtteknél bemutatottakkal. A tagoló néma szünetek közül a legrövidebbek itt is frázisközi szünetek, a szerkesztési szakaszok közül pedig a hibajelenségekhez köthetőek (6. táblázat). A legrövidebb időtartamú egy a fiatal felnőtt beszélők társalgásaiban a grammatikailag inkorrekt helyen tartott frázisközi néma szünet volt, míg a leghosszabb egy az ugyanezen beszélők narratív beszédrészében adatolt frázishatáron tartott néma szünet. A fiatal beszélők narratíváiban a beszélők bizonytalanságából adódó megakadások szerkesztési szakaszai között adatoltuk a legrövidebb szünetet, a leghosszabbat a frázishatáron tartott néma szünetek között; a társalgásokban a legrövidebbet a frázisközi szünetek között, a leghosszabbat a megnyilatkozás elejiek közt. A középkorú beszélők narratíváiban a legrövidebb szünet a frázisközi szünetek közé, a leghosszabb a megnyilatkozás elejiek közé tartozott; a társalgásokban a legrövidebb szünetet

a bizonytalansági megakadásokhoz kapcsolódó szerkesztési szakaszok csoportjában adatoltuk, míg a leghosszabbat a frázisvégiek között. A legnagyobb szórás a fiatal felnőtt beszélők narratíváiban előforduló frázishatáron lévő szüneteket jellemezte, míg a leghomogénebb csoportnak ugyanezen beszélők narratíváiban adatolt hiba típusú megakadások szerkesztési szakasza bizonyult.

6. táblázat. A szünettípusok időtartam-értékei a beszédtípus és a korcsoport függvényében

korcsoport	beszédtípus	szünettípus	átlag (ms)	szórás (ms)	minimum (ms)	maximum (ms)
20-35 év	**narratíva**	**N_Fh**	542	406,3	29,6	4362,9
		N_Fk	376,5	307,2	25,3	1639,6
		N_Fv	741,6	510,3	57,2	2819,7
		N_Me	696,8	742	25,5	1824,3
		S_biz	515,5	461,8	24,8	2585,5
		S_hiba	187,3	180,4	30,9	713,3
	társalgás	**N_Fh**	457	332,7	23	2198,1
		N_Fk	361,8	260,6	19,9	1306,4
		N_Fv	668,4	430,4	35,4	2266
		N_Me	644,1	806,2	135,5	3322,3
		S_biz	436,2	393,4	26,6	2372,9
		S_hiba	283,9	292,2	32,5	1449,5
40-55 év	**narratíva**	**N_Fh**	433,8	346,7	31,4	3399,8
		N_Fk	284,2	235,8	25,9	1842,5
		N_Fv	695,1	548	63,7	3687,4
		N_Me	825,7	1149,9	110,8	3715,3
		S_biz	364,8	309,6	42,1	1584,9
		S_hiba	278,1	297,8	33,3	1324
	társalgás	**N_Fh**	415,9	312,9	37,4	2895,1
		N_Fk	311,9	262,4	40,3	1846,1
		N_Fv	625	454,8	61,2	3356,1
		N_Me	555,2	420,5	131,8	1741,8
		S_biz	337,2	245,4	20,8	1794,7
		S_hiba	188,5	229,4	34,5	1291,5

4. Következtetések

Kutatásunkban a néma szünetek különböző típusainak (vö. Gyarmathy, 2017b) spontán beszédbeli realizációit elemeztük fiatal és középkorú beszélők spontán narratíváiban és társalgásaiban. A 4880 néma szünet elemzéséből kiderült, hogy – jóllehet, az életkor nem befolyásolta matematikailag igazolhatóan a szünetek időtartamát –, a fiatal beszélők általánosságban hosszabb szüneteket tartottak. A két elemzett beszédtípus közül a narratívákat jellemezte a hosszabb szünettartás, ami azzal magyarázható, hogy ezekben az esetekben – társalgásokkal ellentétben – a beszélőnek nem kell kvázi folyamatosan arra (is) figyelnie, hogy a beszédjogot magánál tartsa. A korábbi kutatások eredményeinek megfelelően (Gyarmathy, 2017b,a, 2019; Gyarmathy & Horváth, 2018, 2019) a néma szünetek döntő többsége szintaktikai szerepet betöltve, tagoló néma szünetként realizálódott, csak kisebb hányaduk kapcsolódott a felszíni szerkezetben manifesztálódott tervezési diszharmóniákhoz. Az adatok eloszlásmintázatát vizsgáló statisztikai eljárás igazolta, hogy az egyes szünetek kategóriákba rendeződése nem véletlenszerű. A két korcsoport között különbség volt adatolható a tagoló néma szünetek és a szerkesztési szakaszok előfordulásában; a középkorú beszélőknél mind a szünetek egymáshoz viszonyított arányát tekintve, mind a beszédidőre vetített gyakoriság alapján nagyobb arányú volt a szerkesztési szakaszok megjelenése; az eltérés azonban nem volt szignifikáns. Ezek alapján valószínűsíthető, hogy ezek a beszélők gyakrabban szembesültek tervezési diszharmóniával, mint a fiatalabb társaik. Tekintve, hogy a 100 szótagra vetített gyakorisági mutatókban nem találtunk ilyen jellegű eltérést a két korosztály között, feltehetőleg ezeket a tervezési nehézségeket rutinosabban, rövidebb idő alatt képesek leküzdeni, mint a fiatalok. A narratívákban és a társalgásokban előforduló szünetek gyakorisági mutatóiban szintén csak a középkorú beszélőknél találtunk lényegi különbséget; ők a társalgásokban több szünetet tartottak, ami csak részben magyarázható a szerkesztési szakaszok magasabb arányával. A két fő szünetkategória előfordulási gyakoriságára irányuló statisztikai elemzések

alapján igazolódott, hogy az előfordulást elsősorban a szünettípus, illetve kisebb mértékben a beszéd típusa határozza meg.

A főkategóriák időtartamának elemzéséből megállapítható volt, hogy a tagoló pozíciójú néma szünetek életkortól és beszédtípustól függetlenül hosszabban realizálódnak, mint a szerkesztési szakaszok; a beszélők tehát törekszenek arra, hogy a tervezési nehézségeiket, esetleges hibáikat minél gyorsabban javítsák, ezzel ne akadályozzák a hallgató megértési folyamatait. Jellemző volt továbbá mindkét korosztálynál, hogy a társalgásokban mind a tagoló néma szünetek, mind a szerkesztési szakaszok rövidebb időtartammal valósultak meg, mint a narratívákban, ami a társalgások dinamikusabb jellegével, a beszédpartnerek aktív jelenlétével magyarázható.

A néma szünetek alkategóriái szerint elemezve az előfordulási gyakoriságot, megállapítható volt, hogy mind a szünetek egymáshoz viszonyított arányát tekintve, mind pedig a 100 szótagra vetített gyakoriságot, beszédtípustól és életkortól függetlenül az elemi szintaktikai és szemantikai egységeket elválasztó, frázishatáron lévő néma szünetek a leggyakoribbak. Ezt követik a grammatikailag inkorrekt helyen lévő, és ezzel az értelmezést adott esetben potenciálisan megnehezítő frázisközi néma szünetek, majd a szemantikai egységeket lezáró frázisvégi szünetek. A megnyilatkozás eleji szünetek az interjúhelyzetből fakadóan csak csekély előfordulást mutattak. A szerkesztési szakaszok közül minden összesítésben a bizonytalansági jelenségekhez köthetőek voltak a gyakoribbak. Az adatok eloszlásvizsgálata minden esetben megerősítette azok statisztikailag igazolható, szabályszerű mintázatát. A fiatalokat ezúttal is minden alkategóriát illetően is gyakoribb szünettartás jellemezte, de esetükben a narratív és társalgásos szövegrészek között nem találtunk különbséget. A középkorúaknál ezzel szemben a két beszédtípus közt lényegi különbségek körvonalazódtak; a társalgásaikban majdnem minden alkategóriát magasabb előfordulási gyakoriság jellemez, mint a narratív beszédrészekben. A statisztikai elemzések megerősítették, hogy a gyakorisági mutatókat elsősorban a szünettípus befolyásolja, kisebb mértékben pedig a beszéd típusa. Az alkategóriák temporális paramétereinek részletes elemzéséből kiderült, hogy életkortól és beszédtípustól függetlenül a

leghosszabb időtartammal a szintaktikailag, szemantikailag és grammatikailag is adekvát helyen megjelenő néma szünetek realizálódtak. A leghosszabbak a frázisvégi és a megnyilatkozás eleji szünetek voltak, amely megfelel a korábbi szakirodalmi eredményeknek (Oliveira, 2002; Gyarmathy, 2017b,a), ezeket követték a frázishatáron megvalósulók. A grammatikai struktúrát megtörő, és ezzel potenciálisan a feldolgozást nehezítő frázisközi szünetek időtartamukban a szerkesztési szakaszokhoz idomultak, ami miatt felmerül a kérdés, hogy vajon megjelenésük hátterében milyen folyamatok állhatnak. Elképzelhető, hogy amíg a szerkesztési szakaszok a felszíni szerkezetben megjelenő tervezési és kivitelezési zavarok fémjelzői, addig a frázisközi szünetek a rejtetten zajló tervezési diszharmóniák manifesztációi. Ennek kiderítése, felfejtése azonban további szisztematikus elemzéseket igényel.

Összegzésképpen megállapíthatjuk, hogy kiinduló hipotéziseink közül az első, miszerint a néma szünetek időtartamát és gyakoriságát elsődlegesen a szünet típusa, a közlésben betöltött pozíciója és funkciója határozza meg, teljes mértékben igazolódott, amit a statisztikai elemzések is alátámasztanak. Második hipotézisünk, mely szerint a különböző beszédtípusokban a beszélők eltérő szünettartási stratégiákat alkalmaznak, amely tetten érhető az időtartam-realizációkban, csak részben igazolódott, hiszen a beszéd típusa csak a szünetek előfordulási gyakoriságára volt hatással, időtartamukra nem. Végül ugyan az életkor meghatározó szerepe statisztikailag nem volt bizonyítható, a harmadik előfeltevésünk helyesnek bizonyult, hiszen a két vizsgált korosztályban azonos tendenciákat találtunk a szünettartási stratégiákat illetően beszédtípustól függetlenül.

A jelen kutatásban mind az előfordulási gyakoriságra mind az időtartam-realizációkra kapott eredmények megfelelnek a témakörben közölt korábbi tanulmányok adatainak (Gyarmathy, 2017b,a, 2019; Gyarmathy & Horváth, 2018, 2019).

Köszönetnyilvánítás

A kutatást a Bolyai János Kutatási Ösztöndíj és az NKFI 128810 számú pályázata támogatta.

Hivatkozások

Bada, E., & Genç, B. (2008). Pausing preceding and following to in to-infinitives: A study with implications to reading and speaking skills in elt. *Journal of Pragmatics*, *40*, 1939–1949.

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.37, retrieved 14 March 2018 from http://www.praat.org/.

Bóna, J. (2007). A felgyorsult beszéd produkciós és percepciós sajátosságai.

Bóna, J. (2010). Beszédtervezési folyamatok az életkor és a beszédstílus függvényében. *Magyar Nyelvőr*, *134*, 332–341.

Bóna, J. (2012). A spontán beszéd sajátosságai idősödő, idős és matuzsálemi korban. In A. Markó (Ed.), *Beszédtudomány. Az anyanyelv-elsajátítástól a zöngekezdési időig. ELTE BTK–MTA Nyelvtudományi Intézet* (p. 100–115).

Bóna, J. (2013a). A beszédszünetek fonetikai sajátosságai a beszédtípus függvényében. *Beszédkutatás*, (p. 60–76).

Bóna, J. (2013b). *A spontán beszéd sajátosságai az időskorban*. Budapest: ELTE Eötvös Kiadó.

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

Bruneau, T. J. (1973). Communicative silences: forms and functions. *Journal of Communication*, *23*, 17–46.

Duez, D. (1982). Silent and non-silent pauses in three speech styles. *Language and Speech*, *25*, 11–25.

Erikson, E. H. (1963). *Childhood and Society*. (2nd ed.). New York: Norton.

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

Gee, J. P., & Grosjean, F. (1983). Performance structures: a psycholinguistics and linguistics appraisal. *Cognitive Psychology*, *15*, 411–458.

Gocsál, A. (2001). Gyorsabban beszélnek-e a nők, mint a férfiak? *Beszédkutatás*, (p. 61–72).

Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quarterly. Journal of Experimental Psychology*, *10*, 96–106.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. London: Academic Press.

Gósy, M. (2000). A beszédszünetek kettős funkciója. *Beszédkutatás*, (p. 1–14).

Gósy, M., Beke, A., & Horváth, V. (2011). Temporális variabilitás a spontán beszédben. *Beszédkutatás*, (p. 5–30).

Gósy, M., Gyarmathy, D., Horváth, V., Gráczi, T. E., Beke, A., Neuberger, T., & Nikléczy, P. (2012). Bea: Beszélt nyelvi adatbázis.

Gráczi, T. E. (2013). Explozívák és affrikáták időviszonyai. *Beszédkutatás*, (p. 94–120).

Gyarmathy, D. (2007). Az alkohol hatása a spontán beszédprodukcióra. *Beszédkutatás*, (p. 108–121).

Gyarmathy, D. (2008). Különböző zajok hatása a beszédprodukcióra. *Alkalmazott Nyelvtudomány*, *VIII*, 135–147.

Gyarmathy, D. (2017a). Anyanyelvi és idegennyelvi szünettartási stratégiák. alkalmazott nyelvtudomány xvii. évf.

Gyarmathy, D. (2017b). A néma szünetek funkciói a spontán beszédben. *Beszédkutatás*, (p. 67–92).

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

Gyarmathy, D., & Horváth, V. (2018). A néma szünetek sajátosságai óvodások és kisiskolások spontán beszédében. *Beszédkutatás*, (p. 134–155).

Gyarmathy, D., & Horváth, V. (2019). Pausing strategies with regard to speech style. In *In. Proceedings of DiSS* (p. 12–13). ELTE Eötvös Loránd University.

Hargreaves, W. A., & Starkweather, J. A. (1959). Collection of temporal data with the duration tabulator. *Journal of the Experimental Analysis of Behavior*, *2*, 179.

Imre, A. (2005). Különböző műfajú szövegek szupraszegmentális jellemzői. *Magyar Nyelvőr*, *129*, 510–520.

Iván, L. (2002). Az öregedés aktuális kérdései. *Magyar Tudomány*, *47*, 412–418.

Kowal, S., O'Connell, D. C., & Edward, J. S. (1975). Development of temporal patterning and vocal hesitations. *Journal of Psycholinguistic Research*, *4*, 195–207.

Krivokapic, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, *35*, 162–179.

Kurzon, D. (2007). Towards a typology of silence. *Journal of Pragmatics*, *39*, 1673–1688.

Laczkó, M. (2009). Középiskolai tanulók spontán beszédének temporális jellegzetességei. *Magyar Nyelvőr*, *133*, 447–467.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation. A Bradford Book*. Cambridge (Massachusetts)–London (England: The MIT Press.

Levin, H., Silverman, I., & Ford, B. (1967). Hesitations in children's speech during explanation and description. *Journal of Verbal Learning and Verbal Behavior*, *6*, 560–564.

Lounsbury, F. G. (1965). Transitional probability, linguistic structure and system of habit-family hierarchies. In C. Osgood, & T. Sebeok (Eds.), *Psycholinguistics. A survey of theory and research problems* (p. 93–101). Indiana University Press.

Markó, A. (2005a). *A spontán beszéd néhány szupraszegmentális jellegzetessége*. Budapest: PhD-értekezés. ELTE.

Markó, A. (2005b). A temporális szerkezet jellegzetességei eltérő kommunikációs helyzetekben. *Beszédkutatás*, (p. 63–77).

Markó, A. (2014). A beszéd temporális szerkezete a beszédmód és a beszédhelyzet függvényében. In S. Bátyi, J. Navracsics, & M. Vígh-Szabó (Eds.), *Nyelvelsajátítási, nyelvtanulási és beszédkutatások. Pszicholingvisztikai tanulmányok IV. Gondolat Kiadó. Pannon Egyetem MFTK* (p. 33–45). Budapest – Veszprém.

Menyhárt, K. (2003). A spontán beszéd megakadásjelenségei az életkor függvényében. In L. Hunyadi (Ed.), *Kísérleti fonetika – laboratóriumi fonológia a gyakorlatban* (p. 125–138). Debrecen: Debreceni Egyetem Kossuth Egyetemi Kiadója.

Menyhárt, K. (2010). A beszédsebesség objektív mérési és szubjektív észlelési eredményeinek összefüggései mai és 60 évvel ezelőtti beszélőknél. *Beszédkutatás*, (p. 110–124).

Menyhárt, K. (2011). Régi mesék prozódiája: Palkó Józsefné meséi. *Beszédkutatás*, (pp. 96–109).

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

Neuberger, T. (2014). *A spontán beszéd sajátosságai gyermekkorban*. Budapest: ELTE Eötvös Kiadó.

Olaszy, G. (2005). Prozódiai szerkezetek jellemzése a hírfelolvasásban, a mesemondásban, a novella és a reklámok felolvasásában. *Beszédkutatás*, (p. 21–50).

Olaszy, G. (2007). Beszédstratégiák a prozódia tükrében. *Magyar Tudomány*, *167. évf. 1*, 58–61.

Oliveira, M. (2002). The Role of Pause Occurrence and Pause Duration in the Signaling of Narrative Structure. In E. Ranchhod, & N. J. Mamede (Eds.), *Advances in Natural Language Processing. PorTAL 2002. Lecture Notes in Computer Science* (p. 43–51). Berlin, Heidelberg: Springer volume 2389.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language*, *50*, 696–735.

Sallai, J., & Szende, T. (1995). Spontán közlések beszédszüneteinek pszicholingvisztikai értelmezése (egészséges és skizofrén közlők szövegeiben). *Általános Nyelvészeti Tanulmányok*, *XVIII*, 209–222.

Tannenbaum, P. H., Williams, F., & Wood, B. S. (1967). Hesitation phenomena and related encoding characteristics in speech and typewriting. *Language and Speech*, *10*, 203–215.

Trouvain, J., Fauth, C., & Möbius, B. (2016). Breath and non-breath pauses in fluentand disfluent phases of German and French L1 and L2 read speech. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), *Proceedingsof Speech Prosody (SP8)* (p. 31–35). Boston.

Trouvain, J., & Möbius, B. (2014). Individuelle Ausprägung von Atmungspausen in der Mutter-und in der Fremdsprache als Anzeichen kognitiver Belastung. In *Elektronische Sprachsignalverarbeitung 2014: Tagungsband der 25* (p. 177–184). Konferenz, Dresden.

Vallent, B. (2005). A spontán beszéd ötven éve és ma. *Esettanulmány. Beszédkutatás*, (p. 99–111).

Váradi, V. (2010). A felolvasás és a spontán beszéd temporális sajátosságainak összehasonlítása. *Beszédkutatás*, (p. 100–109).

Verzeano, M., & Finesinger, J. (1949). An automatic analyzer for the study of speech in interaction and in free association. *Science*, *110*, 45.

Volkskaya, N. B. (2003). Virtual and real pauses at clause and sentence boundaries. In *Proceedings of the 15h International Congress of Phonetic Sciences.* (pp. 499–502). Barcelona.

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (p. 41–62). Chichester: John Wiley.

Zwirner, E., & Zwirner, K. (1937). Phonometrischer beitrag zur frage der lesepausen. *Archives Néerlandaises de Phonétique Expérimentale*, *XIII*, 111–128.

Az egyszerre beszélések jellemzői háromfős társalgásokban

Horváth Viktória[1]

[1] *Nyelvtudományi Intézet*

Abstract

Overlapping speech in spontaneous conversations Overlapping speech is claimed to be an „error" in the conversations by some authors, yet it is a frequent phenomenon is spontaneous conversations. Overlapping speech was investigated by just a few studies in Hungarian. Furthermore, the correlation of the main functions and the duration of overlapping speech was also less investigated as well as the realization of the phenomenon with regard to the participants' age and role in conversations.

The aim of the study was to analyze the characteristics of overlapping speech (frequency, duration, types) in spontaneous conversations. 10 conversations were selected from the Hungarian Spontaneous Speech Database. Three speakers participated in each conversation; two of them were constant in each case: the interviewer (Int) and the second speaker (S2). S1 speaker was the experimental person: 5 female speakers ages between 25 and 35 years, while 5 female speakers ages between 60 and 65 years.

The total material was more than 2 hours long, it was manually annotated using Praat. The frequency, the duration and the function of overlapping speech were analyzed. The statistical analysis was conducted using SPSS 20.0.

Results showed that the frequency of overlapping speech changed during the conversations: it was more frequent at the end of the conversational units – overlapping speech play an important role in turn-takings. The two main functional categories (starting a turn/backchanneling) was separable by their duration: the backchanneling overlaps were shorter than overlaps in the other main function. In addition, unsuccessful turn takings were preceded by longer overlapping speech than successful turn-takings. Short overlaps preceding turn-takings reflect to the extensive prediction in turn-takings.

1. Bevezetés

A hagyományos diskurzuselemzés a kezdetektől fogva foglalkozott a társalgásokban megjelenő egyszerre beszélések különféle típusaival (Sacks et al., 1974). Sacks és munkatársai kiemelték, hogy az átfedő beszéd előfordulhat akkor, ha a beszédpartner hibásan azonosítja a lehetséges beszélőváltási pontot, emiatt szól

Email address: horvath.viktoria@nytud.hu (Horváth Viktória)

Beszédtudomány – Speech Science 2020. 187–213. DOI: 10.15775/Besztud.2020.187-213

bele az aktuális beszélő közlésébe. A félbeszakítás szándéktalanságát jelzik az „elnézést", „bocsánat" udvariassági kifejezések. Az átfedő beszéd ekkor csupán egy „taktikai" hiba a beszélőváltás során. A szerzők különbséget tettek az átfedő beszéd és a közbevágás között, ez utóbbi udvariatlanság megnyilvánulása. Az átfedések oka azonban lehet a beszédlépés-átadás egyik szabályos módja is: aki gyorsabban kezd beszélni, az nyeri el a következő beszédlépés jogát. Az önkiválasztó beszélő az idő előtti kezdés miatt nem akar információt veszteni, ezért ún. beszédindítást megelőző (prestart), a beszédlépését lezáró beszélő pedig beszédzárást megelőző (recompleter) és beszédzárást követő (postcompleter) technikákat alkalmaz. Ezek a beszédlépések között, az ún. „átmeneti zónában" fellépő elemek gyakran átfedik egymást, a megelőző beszédlépést, vagy az önkiválasztó beszélő éppen megkezdett beszédlépését (vö. Sacks et al., 1974; Schegloff, 1979). Az átfedés tehát teljesen „normális" jelenség egy tipikus beszélőváltás esetén is.

Az újabb, korpuszokon alapuló elemzések kimutatták, hogy az átfedő beszéd relatíve nagyarányú a beszélgetésekben. Angol spontán társalgásokban és a telefonbeszélgetésekben 10–13% körüli volt az átfedő beszéd aránya a teljes beszédidőben (Shriberg et al., 2001; Çetin & Shriberg, 2006). Beattie a beszélőváltásokat elemezve (1982) kimutatta, hogy a két résztvevős angol társalgásban 11%-ban fordult elő egyszerre beszélés, több beszélőnél ez az arány már 31% volt. Ennél is nagyobb arányban fordultak elő az egyszerre beszélések a japán társalgásokban (Maynard, 1997).

A hagyományos funkcionális megközelítés szerint a társalgásokban az egyszerre beszélések kompetitív jellegűek (French & Local, 1983). Más kutatások hangsúlyozzák, hogy az átfedő beszédet a beszédpartnerek kooperatív jelleggel (is) használják a társalgásokban (Tannen, 1983; Cantrell, 2013). Ennek egyik típusa, amikor a beszédpartner ugyan párhuzamosan szólal meg az eredeti beszélővel, de csak azért, hogy együtt fejezzék be a közlést, nincs szóátvételi szándék, csak a támogató figyelem jelzése. Előfordulhat, hogy a két beszélő nagyjából azonos tartalmi közlést hoz létre egy időben, formai eltéréssel – az eredeti beszélő közlését kommentálja a partner, mintegy összefoglalva, figyelmét, támogatását kifejezve (Tannen, 1983).

Néhány nemzetközi kutatás összekapcsolta a diskurzuselemzés funkcionális aspektusait és a fonetikai szempontokat az egyszerre beszélések vizsgálatában. A beszélők a ritmus, a hangerő és dallam paramétereit kombinálva jelzik, hogy megszólalásuk a partner beszéde közben kompetitív vagy kooperatív jellegű-e. A kompetitív jellegű egyszerre beszélések (közbevágások) kevésbé ritmusosan, nagyobb hangerővel és magasabb alapfrekvenciával valósultak meg (French & Local, 1983). A versengő jellegű, szóátvételre irányuló átfedő beszéd esetén más kutatás is hangsúlyozta az emelkedő hangerőt és magasabb alapfrekvenciát (Schegloff, 2000; Hilton, 2016). A kompetitív jellegű párhuzamos megszólalások nemcsak ezekkel a beszédjellemzőkben bekövetkező változásokkal valósulnak meg, hanem az is jellemzi őket, hogy nem egy lehetséges beszélőváltási pontnál következnek be. A lehetséges beszélőváltási pontot megelőző, illetve a beszélő hezitációinak környezetében bekövetkező párhuzamos megszólalások ugyanakkor nem mutattak nagyobb intenzitást és magasabb alapfrekvenciát (Wells & Macfarlane, 1998).

Az egyszerre beszélések időtartamát is elemezték a funkcióval összefüggésben. Az átfedő beszéd előfordulhat egy szótagra, szóra korlátozódva (ilyenkor általában együttműködést jelez), de nem ritka a hosszabb közléseken átívelő átfedés sem – ez utóbbi stratégia lehet a beszédjog megszerzésére. A szóátvétel ilyenkor gyakran a téma változásával is együtt jár (Hutchby & Woofitt, 2006). Az egyszerre beszélés időzítése is összefügg azzal, hogy a párhuzamos megszólalás kompetitív vagy kooperatív jellegű. A megszólalás kompetitív, ha a beszédpartner az előző beszélő közlésének indítása után, de a lehetséges beszélőváltási pont előtt kezd el beszélni (Hilton, 2016). Ha a partnernek már elegendő információja van az elhangzottak alapján, mert az eredeti beszélő már majdnem a közlés végére ért, egyszerre beszélés formájában közösen is befejezhetik a megnyilatkozást a bejóslás alapján; ez inkább támogató jellegű átfedő beszéd. Kompetitív jellegű, amikor a két beszélő önkiválasztással egyszerre szólal meg, ilyenkor az egyik beszélő egy kis idő után abbahagyja a beszédet, megszüntetve ezzel a párhuzamos megszólalást (Levinson & Torreira, 2015).

A beszélők nemcsak bejósolni tudják az eredeti beszélő közlésének végét; gyakran megismétlik az elhangzottak egy részét, mintegy összefoglalva azt; ez is lehet átfedő megnyilatkozás. Ebben az esetben azonban az egyszerre beszélés nem számít közbevágásnak – az eredeti beszélő folytatja a közlést, nincs szóátvételi szándék. Ezek a párhuzamos közbevetések inkább támogató jellegűek, és „metaüzenet"-nek tekinthetők a beszédpartnerek közötti viszonyról: olyan jól megértik egymást, hogy akár azt is tudják, miként fog folytatódni a másik fél közlése. Az egyszerre beszélések ezen típusai tehát a támogató figyelem jelzései a hallgató részéről (Duranti, 1997; Tannen, 1983, 2012).

Az elemzések az átfedő beszédek külön csoportjaként kezelik az aktuális beszélő közlése közbeni háttércsatorna-jelzéseket a hallgató részéről: azokat a verbális vagy nonverbális jelzéseket, amelyeket egy társalgásban a hallgató a beszélő beszéde közben ad figyelem, megértés jelzésre, és/vagy tovább-beszélést ösztönző céllal; szóátvételi szándék nélkül (Yngve, 1970; Schegloff, 2000; Ward & Tsukahara, 2000; Levinson & Torreira, 2015). Az átfedő beszéddel megvalósuló (tehát nem néma szünetben elhangzó) háttércsatorna-jelzések gyakran egy lehetséges beszélőváltási pontot követően, vagy hallgatást követően adatolhatók, ez arra utal, hogy időzítésük érzékeny az aktuális beszélő beszédében lévő esetleges fordulóvég jelzésére (Gravano & Hirschberg, 2009).

Az egyszerre beszélések és a megakadásjelenségek összefüggését vizsgáló korpuszalapú kutatás azt mutatta, hogy átfedő beszéd esetén kétszer annyi megakadás adatolható ahhoz képest, mint amikor egy résztvevő beszél, ez főként az ismétlések növekvő számában volt kimutatható – a beszélő nem akarta, hogy átvegyék tőle a szót (Anna-Decker et al., 2008). Gyakori, hogy az aktuális beszélő közlésében valamilyen megakadásjelenség fordul elő az egyszerre beszélést megelőzően (rövid néma szünet, ismétlés), illetve csökken a beszédtempója (Levinson & Torreira, 2015). A hallgató ekkor valamilyen háttércsatorna-jelzést produkál – mintegy a megakadásra válaszként – azon a ponton, ahol a beszélő folytatja az aktuális közlést a megakadást követően; így jön létre az átfedő beszéd.

Magyar nyelvű társalgásokban eddig viszonylag kevés kutatás vizsgálta az egyszerre beszéléseket. Több résztvevős társalgásokban a jelenség aránya 2–12% körül adódott (Markó, 2006; Bata, 2009; Beke, 2014). Az átfedő beszéddel kapcsolatban a kutatások többsége a jelenség kooperatív funkcióját hangsúlyozta (pl. Markó, 2006; Bata, 2009; Dér, 2012); feladat-orientált társalgásokban (diákoknak párban meg kellett vitatniuk egy adott témát) igazolták, hogy az egyszerre beszélés gyakoribb volt egyetértés esetén, mint egyet nem értéskor (Weidl, 2019). Dér (2012) az egyszerre beszélések tipikus megnyilvánulási formáit foglalta össze 30 társalgás alapján. Egy kategóriába sorolta a háttércsatornajelzéseket; a rövid kommentárokat; azokat az eseteket, amikor a beszédpartner megismétli a beszélő mondandójának egy részét, kifejezve figyelmét és/vagy egyetértését. Sok esetben a hallgató képes befejezni a beszélő megnyilatkozásának záró részét, amely szintén átfedő beszédet eredményez. Külön kategória, amikor a hallgató megkísérli a szóátvételt, de ez sikertelen. Szintén külön csoportot alkotnak az előbbiek kombinációjából adódó egyszerre beszélések.

Markó (2006) fonetikai szempontú elemzésben vizsgálta az egyszerre beszélések temporális jellemzőit és típusait is. A felosztás alapja az volt, hogy az egyik beszélő közlésébe a másik beszélő mikor szól bele, ezzel átfedő közlést létrehozva. A leggyakoribb eset az volt, amikor két beszélő egyszerre kezdett beszélni (két önkiválasztás egy időben). Hasonló arányban jelent meg a „vége előtti megszólalás" esete is, amikor a partner már elegendő információ birtokában van ahhoz, hogy megszólaljon és akár átvegye a szót. Az elemzés alapján Markó is hangsúlyozta, hogy az egyszerre beszélések oka nem az udvariatlanság vagy versengés, hanem az együttműködés a beszélőváltás lebonyolításában. A partner fordulójának vége előtti megszólalás miatti átfedő beszéd gyakoribb volt ismerősökkel készült társalgások esetén, mint amikor idegen személy volt a kísérleti személy; továbbá a jelenség gyakoribb volt a társalgások utolsó szakaszában, mint az elején – a partnerek összeszokásának köszönhetően (Gráczi & Bata, 2010).

Egy pilot kutatás részletesen elemezte a társalgási jelenségek előfordulását és temporális viszonyait (Hámori & Horváth, 2019). Az egyszerre beszélésekkel

kapcsolatban azt találták, hogy a beszélőváltások fele egyszerre beszélést követett: két vagy több résztvevő is egyszerre beszélt, majd az átfedő beszédet követte a szóátvétel az eredeti beszélőtől. Az egyszerre beszélések időtartama a váltás előtt átlagosan 503 ms volt.

Egyetlen kutatás foglalkozott a magyarban bizonyos társalgási jelenségek életkorspecifikus megvalósulásával (Bata, 2009). A szerző Markó (2006) felosztását használva elemezte az egyszerre beszéléseket a BEA adatbázis háromfős társalgási felvételein, hangsúlyozva azok támogató jellegét. Eredményei szerint az egyszerre beszélések mennyisége nem függött a beszédpartnerek életkorától. Más kutatások a társalgásokban előforduló megakadásjelenségek és temporális jellemzők életkorspecifikus jellemzőit tárgyalták, de ezek kizárólag fonetikai szempontú elemzéseket tartalmaztak, és a társalgásokat összevetésben elemezték a narratívákkal, tartalomösszegzésekkel együtt (pl. Bóna, 2014, 2015).

A jelen kutatás célja az egyszerre beszélések elemzése háromfős társalgásokban, a fonetikai és funkcionális aspektusok ötvözésével.

Új megközelítésre volt szükség a hagyományos kooperatív/kompetitív dichotómia helyett a korpuszalapú vizsgálathoz. A kooperatív/kompetitív megkülönböztetés egyrészt nem objektív kategóriákkal dolgozik: az elemző/annotátor által nem elkülöníthető, hogy milyen jellegű volt az átfedő beszéd. Az, hogy mi számít a résztvevők számára támogató vagy versengő egyszerre beszélésnek, szubjektív konstrukció, amely a diskurzusbeli közös jelentéslétrehozás által alakul ki; illetve függ a beszélők társalgási stílusától, valamint szociálisan meghatározott (Tannen, 2012; Hilton, 2016). A tisztán formai kategóriák alkalmazása (rövid/hosszú az egyszerre beszélés) szintén nem megfelelő, mert ezek semmilyen információval nem szolgálnak a jelenség társalgásbeli funkciójáról. A jelen kutatásban alkalmazott elemzési kategóriák ezért formailag megjelenő, objektíven megragadható jelenségekből indulnak ki, funkcionális beágyazottsággal (olyanok, amelyekben nincs – vagy minimális – a szubjektív értékelés).

A beszélő életkorát a fonetikai kutatások és a labovi szociolingvisztikai irányzat a biológiai beágyazottság miatt objektív változóként kezelik; a modern szociolingvisztikában a kort nem statikus jellemzőként kezelik, hanem egy társas és

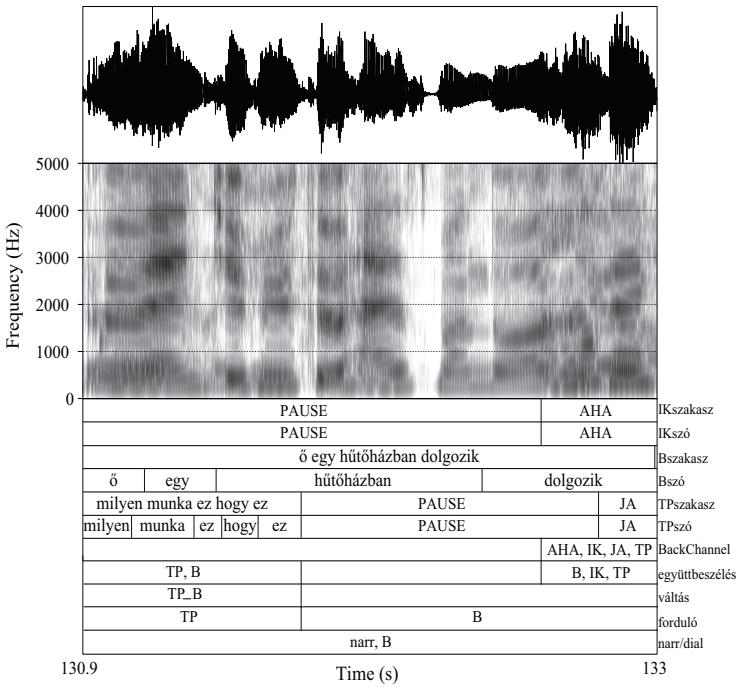
személyes konstrukcióként, amely az adott kontextusban jön létre. „Ugyanakkor az egyén önprezentációja is összetett: nagy eltérések lehetnek az »idősek« csoportjának tagjai között, és az egyén önprezentációja is jelentős változatosságot mutathat egyazon interakció keretei között, illetőleg kontextusonként is" (Bartha et al., 2016, 203.).

A jelen kutatásban változóként kezeljük a kísérleti személy életkorát, mert fontos kérdés, hogy a társalgásokban megjelenő átfedő beszédet hogyan befolyásolja a partnerek biológiai kora; de evidenciaként kezeljük, hogy a jelenség megvalósulását számos más tényező is befolyásolja. Az egyik ilyen tényező a partnerek társalgásban betöltött szerepe, ezért ezt is vizsgáltuk a kutatás során.

Hipotéziseink szerint az egyszerre beszélések gyakoriságát és időtartamát befolyásolja a kísérleti személy életkora – kevesebb lesz az átfedő beszéd az idősebb beszélőkkel rögzített társalgásokban. Feltételeztük, hogy az egyszerre beszélés funkciója (háttércsatorna-jelzés vagy szóátvételi szándék) befolyásolja a jelenség időtartamát; illetve az átfedő beszéd jellemzői (gyakoriság, időtartam) dinamikusan változnak a globális struktúra függvényében (narratív és dialogikus szakaszok váltakozása). Feltételezésünk szerint továbbá a beszélői szerepek (az interjú készítője, a kísérleti személy és társalgó partner) is alakítják az átfedő beszéd megjelenését a társalgásokban.

2. Anyag és módszer, kísérleti személyek

A kutatás alapja 10 darab, 3 résztvevős társalgás a BEA-adatbázisból (Horváth et al., 2019). A meglévő 6 szintű annotációt (beszédszakasz és szó szint a 3 beszélőtől) kiegészítettük további címkesorokkal a Praat programban (Boersma & Weenink, 2018, vö. 1. ábra): jelöltük a háttércsatorna-jelzéseket, az egyszerre beszéléseket, a fordulókat, a beszélőváltásokat, a globális szerkezeten belül azt, hogy az egyes fordulók döntően narratív vagy inkább dialogikus szakaszokhoz tartoznak (Hutchby & Woofitt, 2006). A teljes anyag összesen 134 perc időtartamú (átlagos hossz 13 perc, 8,5–23,5 perc).



1. ábra. Az annotálási rendszer (Horváth et al., 2019, 268.)
(IK = interjú készítője, B = beszélő, a kísérleti személy, TP = társalgó partner)

Összesen 687 darab egyszerre beszélést elemeztünk manuálisan a kontextus alapján a funkció szempontjából. A funkcionális elemzésben elsőként azt állapítottuk meg (Yngve, 1970; Tannen, 2012; Dér, 2012; Cutrone, 2013 alapján), hogy az egyszerre beszélés **háttércsatorna-jelzés** miatt realizálódott, vagy **szóátvételi szándék** volt jelen. Az első esetben nincs szóátvételi szándék, nem jön létre új forduló. A második csoportot tovább bontottuk aszerint, hogy a szóátvétel sikeres volt, vagy nem. **Sikertelen szóátvétel** esetén az aktuális beszélő közlésébe egy másik partner beleszólt, párhuzamosan beszéltek egy ideig, de a szóátvétel nem volt sikeres, a partner elhallgatott, és az eredeti beszélő folytatta a fordulóját. A **sikeres szóátvétel rövid időre** is létrejöhetett: az egyszerre beszélést követően nem az eredeti beszélő, hanem egyik partnere beszélt, de csak egy rövid megjegyzés vagy kérdés erejéig, utána újra az eredeti beszélő folytatta

a közlést. Az egyszerre beszélést **sikeres szóátvétel** is követhette, ebben az esetben a beszélőváltás után új beszélő új fordulója következett.

Példák a négy típusra a korpuszból ([] jelöli az átfedő beszéddel realizálódó részeket, a SIL és a PAUSE a különböző típusú szüneteket):

háttércsatorna-jelzés miatti egyszerre beszélés:

(1) a. B: *túlképzés* [*van*]
TP: [*biztos is*]

(bea113, az átfedő beszéd időtartama: 550 ms)

b. B: *valaki autózik főleg* [*hogyha*] *nem egyedül utazik*
TP: [*ühüm*]

(bea135, az átfedő beszéd időtartama: 234 m)

egyszerre beszélés sikertelen szóátvétellel:

(2) B: *és hányszor éreztük azt hogy ez* [*igazságtalanság és hogy kitoltak velünk soha nem adott a mama se*] *igazat nem fordult elő hogy megvétózza a tanárt...*
IK: [*és* SIL *igen és akkor izé öh ah* PAUSE]

(bea113, az átfedő beszéd időtartama 2957 ms)

egyszerre beszélést követő rövid idejű szóátvétel:

(3) B: *novemberben is elég sok munkájuk* [*volt*]
TP: [*de akkor az*] *valamilyen szempontból jó, nem?* PAUSE
B: *igen meg hogy nagyon jó neki hogy biztos állása van...*

(bea088, az átfedő beszéd időtartama: 511 ms)

egyszerre beszélést követő szóátvétel:

(4) TP: *és nagyon felgyorsult minden szerintem* SIL [*nagyon közel van*]
IK: [*de egyébként*] SIL *a az ikeában az tetszik hogy így ki van írva ilyen...*

(bea113, az átfedő beszéd időtartama: 591 ms)

A funkcionális elemzés mellett elemeztük az egyszerre beszéléseket időtartamuk és a sűrűség függvényében. Az időtartamok kinyerése automatikusan történt Praat-szkript segítségével. Az egyszerre beszélések előfordulásának dinamikus változását a társalgásban a következőképpen vizsgáltuk. A teljes társalgás időtartamát Praat-szkripttel 5 egyenlő egységre osztottuk: például egy 20 perces társalgást az első négy percre (1-20%), a második négy percre (21-40%) stb. Ezekben az egységekben mértük az egyszerre beszélések darabszámát.

A globális szerkezetben való előfordulás vizsgálatához a teljes társalgást felosztottuk több beszédfordulóból álló, nagyobb szakaszokra a szövegtípus alapján: döntően narratív vagy dialógus jellegűek-e. A társalgásokba gyakran ékelődnek ugyanis hosszabb, narratív, általában történetmesélő jellegű részek; ezek monologikus szövegekhez nagyon hasonló felépítésűek (Tolcsvai Nagy, 2001; Hutchby & Woofitt, 2006). A jelen társalgásokban is előfordultak a párbeszédes részek mellett hosszabb, döntően narratív részek, például történetmesélés (ezekben is előfordultak háttércsatorna-jelzések, rövid kommentek a beszédpartnerek részéről, de döntően monologikus szakaszok voltak). A változó időtartamú narratív és dialógus jellegű szakaszokat is 5 azonos részre osztottunk fel azok teljes időtartama szerint Praat-szkript segítségével: például egy 500 másodperces szakaszt 5 db 100 másodperces egységre szegmentáltunk. Ezt követően megnéztük, hogy egy adott egységben hány darab és milyen időtartamú egyszerre beszélés realizálódott.

A társalgásokban 3 beszélő vett részt. Az interjú készítője (IK) és a társalgó partner (TP) minden felvételen ugyanaz volt: a felvételek időpontjában 28–35 éves nők, azonos végzettséggel, kollégák; egynyelvűek, budapestiek. A kísérleti

személy (beszélő, B) személye változott a felvételeken: 5 fiatal nő (25–35 évesek) és 5 idősebb nő (60–72 évesek), egynyelvűek, budapestiek. A témát az interjú készítője jelöli ki (pl. húsvéti szokások a családban, karácsonyi készülődés), megkérdezi a partnerek véleményét az adott témáról, ezt követően a társalgás spontán szerveződik; de a beszélgetést IK zárja le.

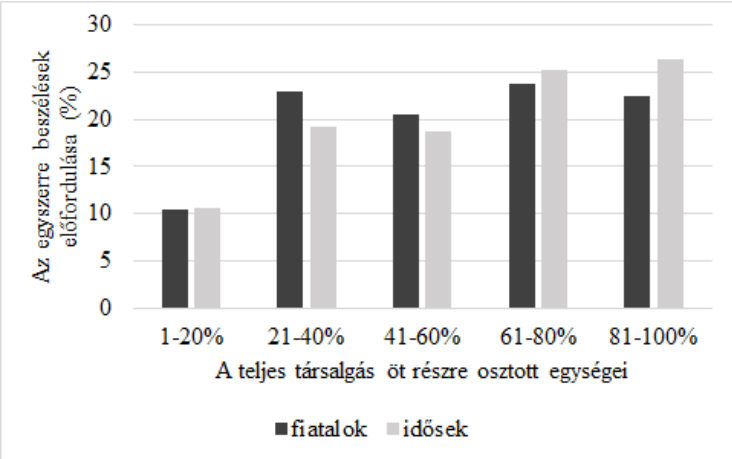
A statisztikai elemzéseket az SPSS programban végeztük (GLM, függő változók az időtartam és az előfordulási gyakoriság, független változók az egyszerre beszélések funkcionális típusai, beszélők életkora, random faktor a beszélők).

3. Eredmények

A több mint kétórányi spontán társalgás 687 egyszerre beszélést tartalmazott, átlagosan 5,4 ilyen jelenség fordult elő percenként (min.: 2,5 db/perc, max.: 8,8 db/perc). A fiatalabb nőkkel rögzített beszélgetésekben percenként átlagosan 5,3 darab átfedő beszéd jelent meg (min.: 2,5 db/perc, max.: 8,7 db/perc); az idősebbekkel készült felvételeken nagyon hasonló értékeket adatoltunk (átlag: 5,6 db/perc, min.: 3,3 db/perc, max.: 8,8 db/perc). A kísérleti személy (B) életkora nem befolyásolta az egyszerre beszélések gyakoriságát. Az egyszerre beszélések összes időtartama (6,4 perc) a teljes anyag 4,8%-át tette ki.

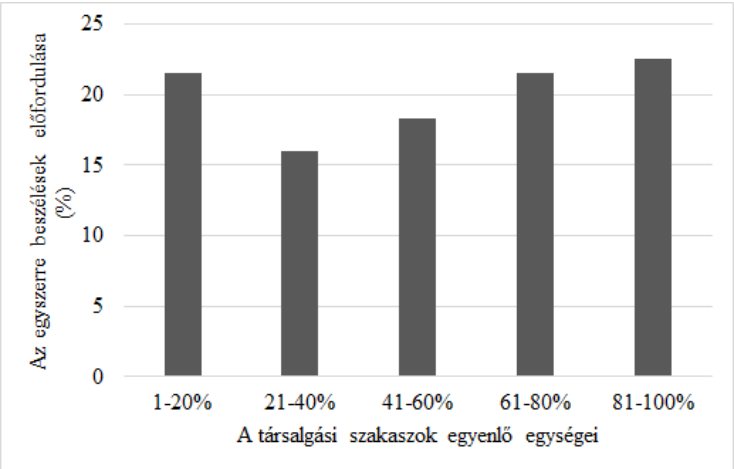
Az egyszerre beszélések gyakoriságának dinamikus változását is elemeztük a társalgásokon belül a teljes időtartamot öt egyenlő hosszúságú egységre bontva (2. ábra). Az eredmények azt mutatták, hogy a fiatalokkal rögzített társalgásokban a teljes felvétel időtartamának első egyötödében fordult elő egyszerre beszélés a legkisebb arányban, a felvétel többi részében ehhez képest kétszer nagyobb, és nagyjából hasonló arányban jelent meg az átfedő beszéd. Az idősekkel készült felvételeken szintén legkevésbé a társalgások elején fordult elő egyszerre beszélés, legnagyobb arányban pedig a társalgás utolsó egyötöd részében.

A következőkben azt vizsgáltuk a nagyobb társalgási szakaszokat, vagyis több fordulót is tartalmazó narratív és dialogikus egységeket 5 egyenlő részre osztva, hogy a szakaszban előforduló összes átfedő beszéd hány százaléka fordul elő a szakasz első 20%-ában, második 20%-ában stb. A 10 társalgást elemezve



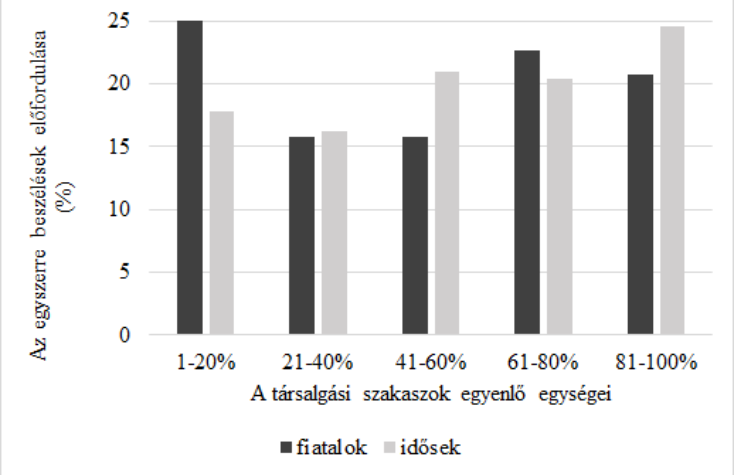
2. ábra. Az egyszerre beszélések előfordulása a teljes társalgás öt egyenlő részre osztott egységeiben

az eredmények azt mutatták, hogy az egyszerre beszélések 21,6%-a a társalgási szakaszok időtartamának első 20%-ában jelent meg (3. ábra). Előfordulásuk ezt követően a társalgási szakaszok belsejében csökkent, majd fokozatosan emelkedő tendenciát találtunk a szakaszok végéhez közeledve (16–22,6%).



3. ábra. Az egyszerre beszélések előfordulása a társalgási szakaszok egységeiben

A fiatal és az idősebb beszélőkkel folytatott társalgásokban külön-külön is elemeztük az egyszerre beszélések előfordulásának változását a társalgási szakaszokban (elsőként függetlenül attól, hogy a szakaszok döntően narratív vagy dialógus jellegűek voltak-e). A fiatalokkal készült felvételeken az egyszerre beszélések egynegyede a társalgási szakaszok elején jelent meg, ennél 10%-kal kevesebb volt az átfedő beszéd aránya a szakaszok belsejében. A szakaszok végéhez közeledve ismét gyakoribbá váltak az egyszerre beszélések. Az idősebbek felvételein a társalgási szakaszok elején még kisebb arányban fordult elő az átfedő beszéd, majd előfordulásuk gyakoribbá vált, és a legtöbb egyszerre beszélést pedig az utolsó egységben adatoltuk (4. ábra).



4. ábra. Az egyszerre beszélések előfordulása a társalgási szakaszok egységeiben a fiatalokkal és az idősekkel folyatott társalgásokban

Az egyszerre beszélések társalgási szakaszokon belüli változása nagyon hasonló tendenciát mutatott a narratív és a dialogikus szakaszokban: legkevésbé a szakaszok közepén jelentek meg, nagyobb volt az átfedő beszéd aránya a szakaszok vége felé közeledve (1. táblázat).

Az egyszerre beszélések előfordulási arányát a jelenség típusa és a kísérleti személy életkorának függvényében is elemeztük (2. táblázat). Az összes háromfős spontán társalgásban az átfedő beszéd döntő többségében akkor jelent meg,

1. táblázat. Az egyszerre beszélések előfordulása a dialogikus és narratív szakaszokban

Az egyszerre beszélések előfordulása (%)						
	összes társalgás		társalgások fiatalokkal		társalgások idősebbekkel	
	szakaszok jellege		szakaszok jellege		szakaszok jellege	
a szakasz részei	dialogikus	narratív	dialogikus	narratív	dialogikus	narratív
1-20%	20,8	21,4	24,6	24,4	16,7	18,2
21-40%	17,1	15,6	16,7	15,6	17,6	15,7
41-60%	16,7	19,1	12,3	17,6	21,6	20,8
61-80%	24,5	20,4	27,1	20,8	21,6	19,9
81-100%	20,8	23,5	19,3	21,6	22,5	25,4

amikor egy résztvevő beszéde alatt egy másik partner hümmögött vagy verbális háttércsatorna-jelzést adott (*ja, persze, jézusom*, stb.). Hasonló volt a rövid idejű sikeres szóátvétel és a valós szóátvételt megelőző átfedő beszéd aránya. 10% alatt volt azon egyszerre beszélések aránya, amikor a párhuzamos közlést követően sikertelen volt a szóátvétel, és az eredeti beszélő folytatta a közlést.

A kísérleti személy életkorától függetlenül a háttércsatorna jelzés funkciójú átfedő beszéd jelent meg a legnagyobb arányban (2. táblázat); az idősebb nőkkel folytatott társalgásokban a résztvevők gyakrabban alkalmaztak ilyen jelzéseket, mint a fiatal nők egymás közötti beszélgetéseiben. A sikertelen szóátvételt megelőző egyszerre beszélések azonos arányban jelentek a társalgásokban a kísérleti személy életkorától függetlenül. Nagyobb arányban fordult elő fiatalok esetében, hogy az egyszerre beszélést követően a partner csak egy rövid kérdés vagy komment idejére vette át a szót – az idősekkel folytatott beszélgetésekben ez volt az egyszerre beszélések legritkább típusa. Az egyszerre beszélést követően megvalósuló „valódi" szóátvétel aránya 10% körül adódott a kísérleti személy életkorától függetlenül.

Az egyszerre beszélések előfordulását a résztvevők (beszélői szerepek) szerint is elemeztük (3. táblázat). A kísérleti személyek (B) életkorától függetlenül az volt a leggyakoribb eset, hogy az interjúkészítő (IK) és a kísérleti személy beszélt egy időben; az idősek csoportjában ez nagyobb arányban fordult elő, mint a fiataloknál. Mindkét csoportban az volt a második leggyakoribb eset,

2. táblázat. A különböző típusú egyszerre beszélések előfordulása az életkor függvényében

Az egyszerre beszélések előfordulása (%)			
egyszerre beszélés típusa	**összes társalgás**	**társalgások fiatalokkal**	**társalgások idősebbekkel**
háttércsatorna-jelzés	72,3	68,8	76,2
sikertelen szóátvétel	6,1	6,1	6,1
rövid idejű szóátvétel	10,1	14,2	5,5
szóátvétel	11,5	10,9	12,2

hogy a kísérleti személy és a társalgó partner (TP) beszélt egy időben, de ennek előfordulása nagyobb arányú volt a fiatal nőkkel készült felvételeken. Az egymást jól ismerő IK és TP kicsivel nagyobb arányban beszélt átfedő jelleggel azokban a társalgásokban, ahol a harmadik partner is fiatal nő volt. Az egyszerre beszélések kb. 1%-ánál fordult elő, hogy mindhárom résztvevő egy időben beszélt.

3. táblázat. Az egyszerre beszélések előfordulása a beszélői szerepek és az életkor függvényében

Az egyszerre beszélések előfordulása (%)		
beszélői szerepek	**társalgások fiatalokkal**	**társalgások idősebbekkel**
B–IK	52,9	72
B–TP	29,8	14,9
IK–TP	15,9	11,3
B–IK–TP	1,4	1,8

Tovább elemeztük a beszélő szerepek és az átfedő beszéd összefüggését aszerint, hogy ki volt az eredeti beszélő, és melyik partner szólalt meg ezzel egy időben. A fiatalokkal rögzített társalgásokban az összes átfedő beszéd 39%-ában a B beszélő (kísérleti személy) beszédébe szólt bele IK. Jóval ritkább volt (19%), amikor TP szólt bele IK beszédébe. Az esetek 0,8%-ában fordult elő,

hogy B beszélt, TP és IK is megszólalt. 14%-ban fordult elő, hogy B szólalt meg, miután IK kezdett beszélni.

Az idősekkel rögzített társalgásokban az összes átfedő beszéd 59%-a jött létre úgy, hogy B közlésébe szólt bele IK. 7%-ban TP szólt bele IK beszédébe; míg az esetek 0,9%-ában B beszélt, IK és TP is megszólaltak. Az idősebb B beszélők 26%-ban szólaltak meg IK beszéde közben.

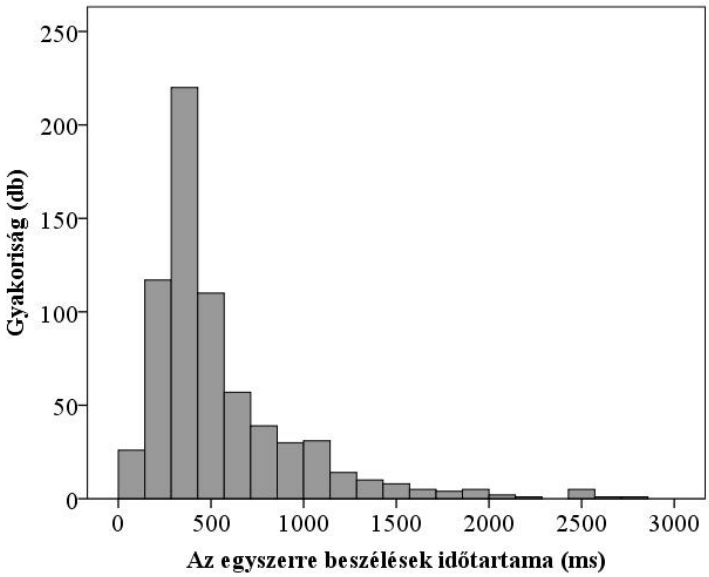
Ezeket az eredményeket befolyásolja, hogy az egyszerre beszélések döntő többsége háttércsatorna-jelzés volt (vö. 2. táblázat) – IK feladata, hogy biztosítsa a támogató légkört, ösztönözze a beszélgetés folytatását; ezért a beszélői szerepek megszólalása szerinti elemzést elvégeztük úgy is, hogy csak a szóátvétel-jellegű átfedő beszédeket vizsgáltuk.

A fiatalokkal készült felvételeken a szóátvételre irányuló átfedő beszédek 24%-ában B beszélt, IK is megszólalt. 19%-ban TP szólt bele IK közlésébe. 2% volt az aránya annak, amikor B beszélt, IK és TP is megszólaltak párhuzamosan. A fiatal B beszélő az esetek 25%-ban IK, 14%-ában TP megnyilatkozásába szólt bele. IK és TP egymás közléseibe megegyező arányban szóltak bele párhuzamosan (8%).

Az idős B beszélők megnyilatkozásaival párhuzamosan IK az esetek 28%-ában szólt közbe, TP jóval ritkábban (6%). Nem fordult elő, hogy B beszédével párhuzamosan mindkét másik beszélő is megszólalt. Az idősebb B beszélő az átfedő beszédek 42%-ában IK, 12%-ában TP megnyilatkozásába szólt bele. Az idősekkel folytatott társalgásokban is az volt a jellemző, hogy IK és TP egymás közléseibe megegyező arányban szóltak bele párhuzamosan (6%).

Az egyszerre beszélések időtartamát is részletesen elemeztük. Az egyszerre beszélések átlagos időtartama 559 ms volt (SD: 442 ms), az értékek 26 ms és 4282 ms között szóródtak. Az átfedő beszéd az esetek 63%-ában 500 ms alatt realizálódott (5. ábra).

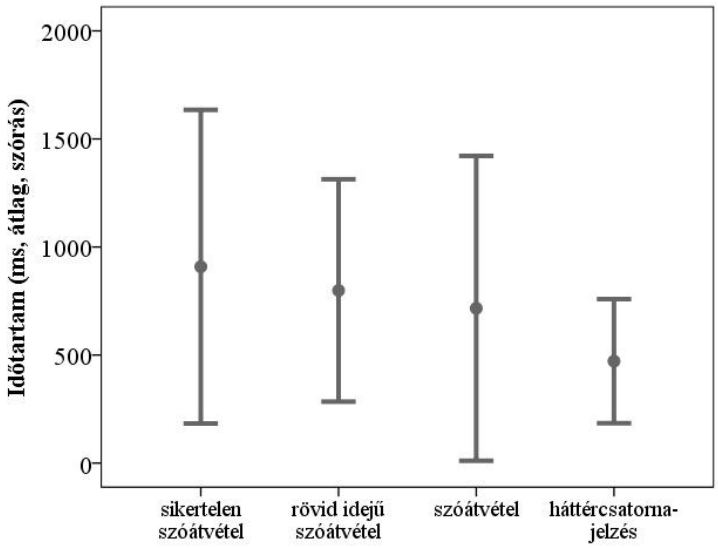
Az egyszerre beszélések időtartamát a társalgásokban a kísérleti személy életkora nem befolyásolta szignifikánsan. A fiatalokkal készült beszélgetésekben az átfedő beszéd időtartama átlagosan 542 ms (SD: 403 ms) volt, az idősekkel készült társalgásokban 578 ms (SD: 481 ms).



5. ábra. Az egyszerre beszélések időtartama

Az egyszerre beszélések időtartamát befolyásolta azok típusa ($F = 13,541$ $p < 0,001$ $\eta^2 = 0,480$, vö. 6. ábra). A háttércsatorna típusú átfedő beszéd (átl.: 471 ms, SD: 287 ms) szignifikánsan rövidebb volt a másik három, szóátvételi céllal megvalósuló jelenségnél ($p < 0,001$ mindhárom esetben). A szóátvételi szándékkal összefüggésben megjelenő egyszerre beszélések szignifikáns mértékben nem különböztek egymástól, de átlagosan a leghosszabbak a sikertelen szóátvételt megelőző jelenségek voltak (átl.: 908 ms, SD: 725 ms), ezeknél átlagosan rövidebb időtartamban valósultak meg a rövid idejű szóátvételt megelőző egyszerre beszélések (átl.: 798 ms, SD: 514 ms). A legrövidebb pedig akkor volt az egyszerre beszélés, ha sikeres szóátvétel követte a jelenséget (átl.: 716 ms, SD: 705 ms).

Az egyszerre beszélések időtartamát a kísérleti személy életkora és a típus után aszerint is elemeztük, hogy a jelenség a társalgási szakaszok melyik részében helyezkedik el (első 20%-ában, második 20%-ában stb.). Az eredmények azt mutatták, hogy az egyszerre beszélések időtartamát nem befolyásolta jelentős



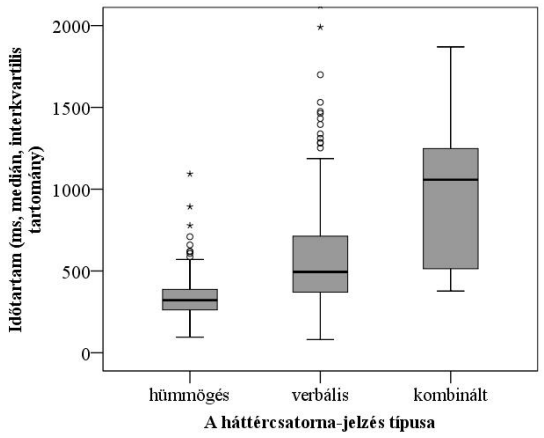
6. ábra. Az egyszerre beszélések időtartama a típus függvényében

mértékben az, hogy a társalgási szakaszok melyik szakaszában jött létre (4. táblázat).

A leggyakoribb átfedő beszéd háttércsatorna-jelzés miatt jött létre (vö. 2. táblázat), ezért ennek a típusnak a megvalósulását és időtartamát részletesen is elemeztük. A háttércsatorna-jelzések döntő többsége vagy verbális típusú volt (49,9%, pl. *ja, persze, igen, hát igen*) vagy a hümmögés különféle formája (48,3%). 1,8% volt azon jelzések aránya, amikor az előző két típus kombinálódott (pl. *mhm hát igen, aha igen*). A háttércsatorna-jelzés realizációja meghatározta az időtartamot (Kruskal-Wallis: $\chi^2 = 141,145$; $p < 0,001$, vö. 7. ábra). A hümmögések voltak a legrövidebbek (átl.: 335 ms, SD: 121 ms), ezeknél a verbálisak átlagosan mintegy 250 ms-mal hosszabbak voltak (átl.: 586 ms, SD: 322 ms); a kombinált típus átlagos időtartama pedig 946 ms volt (SD: 503 ms). A különbség minden típus között statisztikailag szignifikáns (Mann-Whitney: hümmögés–verbális: $Z = -11,483$ $p < 0,001$; verbális–kombinált: $Z = -2,276$ $p = 0,023$; hümmögés–kombinált: $Z = -4,507$ $p < 0,001$).

4. táblázat. Az egyszerre beszélések időtartama a társalgási szakaszok egyes részeiben

Az egyszerre beszélések időtartama (ms)				
	társalgások fiatalokkal		társalgások idősebbekkel	
a szakasz részei	**átlag**	**SD**	**átlag**	**SD**
1-20%	544	445	613	465
21-40%	571	441	615	571
41-60%	513	394	554	380
61-80%	535	359	603	454
81-100%	569	405	550	538



7. ábra. A háttércsatorna-jelzések időtartama a típus függvényében

Az egyszerre beszélések fontos szerepet töltenek be a beszélőváltások során. A jelen korpuszban nagyjából minden tizedik egyszerre beszélést beszélőváltás követett (vö. 2. táblázat). Az átfedő beszéd átlagos időtartama ebben az esetben 716 ms volt, az adatok nagy szóródást mutattak (min.: 26 ms, max.: 4282 ms). A beszélőváltásokat megelőző egyszerre beszélések nagyjából fele 600 ms alatti időtartamban realizálódott.

A hagyományos kooperatív/kompetitív dichotómia az egyszerre beszéléseknél az esetek nagy részében nem alkalmazható (Tannen, 2012; Hilton, 2016), illetve az adott helyzetben konstruálódik, hogy a résztvevők számára mi számít kooperatív vagy kompetitív együtt beszélésnek. A tanulmánynak ezért nem volt célja minden egyes átfedő beszéd ilyen jellegű elemzése, ennek ellenére a jelen korpuszban is adatoltunk olyan eseteket, amikor a kooperatív, támogató jelleg egyértelműen kimutatható volt nemcsak a háttércsatorna-jellegű egyszerre beszélések esetében. A támogató jellegű egyszerre beszélések egyik típusa, amikor a partner megismétli, összefoglalja a beszélő közlésének egy részét – ez nem félbeszakítás, hanem az aktív figyelem jelzése (Duranti, 1997; Dér, 2012; Tannen, 2012). Példák a korpuszból:

(5) IK: *és akkor hogy neki gyűjtsünk mindenféle MM* (704 ms), SIL (554 ms) [*ismeretségeket meg ötleteket*]
TP: [*ötletek*]
(bea006, átfedő beszéd időtartama: 676 ms)

(6) B: *hogy ilyeneket ahogy mi az amire* [*mi nem is gondoltunk*]
TP: [*nem is gondolsz*]
(bea006, átfedő beszéd időtartama: 816 ms)

4. Következtetések

A jelen tanulmányban a társalgásokban megjelenő egyszerre beszéléseket elemeztük korpusz alapon, funkcionális keretben. A kutatáshoz felhasznált 10

darab háromfős társalgásban az átfedő beszéd összes időtartama a teljes anyag 4,8%-át tette ki. Ez az arány megegyezik politikai jellegű beszélgetésekben adatolt értékkel (Anna-Decker et al., 2008). Magyar nyelvre is adatoltak hasonló előfordulást több résztvevős társalgásokban (Markó, 2006), míg más kutatások (Bata, 2009) kevesebb átfedő beszédet állapítottak meg; nagyobb korpuszon pedig 12% volt az egyszerre beszélés aránya (Beke, 2014).

Az egyszerre beszéléseket elkülönítettük aszerint, hogy háttércsatorna-jelzés miatt jött létre (nem volt szóátvételi szándék) vagy szóátvételi szándékból adódott az átfedő beszéd. A korpuszban az összes átfedő beszéd majdnem háromnegyede háttércsatorna-jelzés volt, hasonlóan egy nagy korpuszon végzett korábbi kutatás eredményeihez (Levinson & Torreira, 2015). A korpuszalapú vizsgálatok megerősítik, hogy az egyszerre beszélések sokszor támogató funkcióban jelennek meg a társalgásokban, nem versengő jelleggel. Ráadásul a hagyományos kooperatív/kompetitív felosztás nem megfelelő a jelenség részletes vizsgálatához, mert az átfedő beszéd jellege csak az adott kontextusban értelmezhető, az adott társalgásban konstruálódik a résztevők által és dinamikus változik az egyezkedés során. Az összes egyszerre beszélés nagyjából egyharmada valósult meg szóátvételi szándékkal összefüggésben: egyaránt 10% körül adódott a rövid idejű szóátvételt és a szóátvételt megelőző átfedő beszéd aránya; és az esetek 6%-ban a párhuzamos megszólalást nem követte szóátvétel, az eredeti beszélő folytatta a közlést.

A jelen korpuszban az egyszerre beszélések átlagos időtartama 559 ms volt (az értékek 26 ms és 4282 ms között szóródtak); átfedő beszéd az esetek kétharmadában 500 ms alatt realizálódott. Nagyon hasonló értékeket adatolt Markó (2006) négyfős magyar társalgásokban; illetve 610 ms volt az átlagérték spontán holland társalgások egyszerre beszéléseiben is (Heldner & Edlund, 2010). A hasonló időtartam-értékek felvetik olyan kutatások szükségességét, hogy különböző nyelvekben miként alakul az egyszerre beszélések időtartama több résztvevős társalgásokban – mennyire univerzálisak, vagy kulturálisan meghatározottak az átfedő beszéd időzítései paraméterei.

Az egyszerre beszélések funkcionális típusait és időtartamát elemezve az eredmények azt mutatták, hogy a szóátvételi szándékkal megjelenő és a szóátvételi szándék nélküli (háttércsatorna-jelzés) átfedő beszéd időtartamában szignifikánsan elkülönül egymástól. A szóátvételi szándékkal realizálódó egyszerre beszélések altípusai is különböznek egymástól időtartamukban, de csak tendenciaszerűen. Átlagosan azok a leghosszabb egyszerre beszélések, amelyek után nem következik be beszélőváltás. Ezekben az esetekben a leghosszabb az egyezkedés a beszédjogért, és végül nem jön létre beszélőváltás, a szóátvételi kísérlet sikertelen, az eredeti beszélő folytatja a közlést, a párhuzamosan megszólaló partner pedig elhallgat. Ezeknél átlagosan rövidebbek azok az egyszerre beszélések, amelyeket követően a partner rövid időre átveszi a szót (megjegyzés, kérdés erejéig); a legrövidebbek pedig azok, amelyet sikeres szóátvétel követ. Az összes átfedő beszéd kb. 10%-a szolgált ilyen jellegű szóátvételre; és arra utal, hogy a jelenség a gördülékeny szóátvétel egyik eszköze a társalgásokban.

A kutatás során vizsgáltuk, hogy a kísérleti személy (B) életkora, a beszélői szerepek és a társalgás globális szerkezete hogyan befolyásolja az egyszerre beszélések megjelenését.

Hipotézisünkkel ellentétben a kísérleti személy életkora nem befolyásolta az egyszerre beszélések gyakoriságát és időtartamát sem szignifikáns mértékben: az átfedő beszéd percenkénti gyakorisága és időtartama nem különbözött a fiatal és idősebb nőkkel készült társalgásokban. A részletesebb elemzés a beszélői szerepek és a kísérleti személy (B) életkorának függvényében azonban igazolt néhány különbséget. Az interjú készítője (IK) nagyobb arányban szólalt meg párhuzamosan a B beszélővel, ha az idősebb volt, mint akkor, ha fiatal. Az egyszerre beszélések előfordulása a dialogikus és narratív szakaszokbanz azzal magyarázható, hogy az interjú készítője gyakrabban produkált valamilyen háttércsatorna-jelzést az idősebb beszédpartner közlése alatt, feltehetően ebben szerepet játszik az udvariasság is. A szóátvételre irányuló párhuzamos megszólalás arányát IK részéről ugyanakkor nem befolyásolta a B beszélő életkora, de a társalgó partner (TP) figyelembe vette a B beszélő életkorát: jóval ritkábban szólalt meg a be-

szélővel párhuzamosan idősebb női beszélő esetén (ezt természetesen nemcsak a partner biológiai életkora, hanem a téma, a bevonódás stb. is befolyásolhatja).

A társalgásokon belül az egyszerre beszélések megjelenése váltakozást mutatott. A beszélő életkorától függetlenül az átfedő beszéd legkevésbé a társalgások első részében volt jellemző – az interjúkészítő által megjelölt témáról először a másik két beszélő kifejti a véleményét, inkább narratív közlések jellemzők. Az egyszerre beszélés a társalgások vége felé mintegy kétszer nagyobb gyakorisággal jelent meg; ez feltehetően a beszédpartnerek összeszokásával, illetve a bevonódás hatásával magyarázható. A vizsgált tényezők közül a globális szerkezet (narratív és dialogikus szakaszok váltakozása) szignifikáns mértékben nem befolyásolta az átfedő beszéd jellemzőit. Tendenciaszerűen azonban igazoltuk, hogy az egyszerre beszélések legnagyobb arányban az egyes társalgási szakaszok vége felé fordultak elő. Ez arra utal, hogy az átfedő beszéd segíti az egyes szakaszok lezárását – függetlenül azok narratív vagy dialogikus jellegétől –, hozzájárulva a társalgások szerveződéséhez. Ezt támasztja alá a jelen kutatás azon eredménye is, miszerint a sikeres beszélőváltásokat megelőző átfedő beszéd nagyjából fele 600 ms alatt realizálódott – ez az extenzív predikció (vö. Levinson & Torreira, 2015) nagymértékű jelenlétére utal a háromfős társalgásokban. A jelen kutatás eredményei tehát nemcsak az új funkcionális megközelítés, a beszélői szerepek vizsgálata és a fonetikai szempontok ötvözése miatt fontosak; az eredmények hangsúlyozzák, hogy az átfedő beszéd a társalgás gördülékeny lebonyolításának fontos eszköze.

Köszönetnyilvánítás

Köszönöm Hámori Ágnes, Krepsz Valéria és Markó Alexandra, valamint az anonim lektorok javaslatait, segítségét.

A kutatást az NKFI K-128810 számú pályázat támogatta.

Hivatkozások

Anna-Decker, M., Barras, C., Adda, G., Paroubek, P., de Mareüil, P. B., & Harbert, B. (2008). Annotation and analysis of overlapping speech in political interviews. In *LREC 2018*. Marrakech, Morocco.

Bartha, C., Hámori, Á., & Huppert, A. (2016). Az öregség/idősség és kor konstruálása és prezentációi: Szociolingvisztikai és diskurzuselemzési alapvetések és gyakorlati elemzések. In G. Balázs, & Á. Veszelszki (Eds.), *Generációk nyelve* (p. 203–220). Budapest: ELTE BTK. Mai Magyar Nyelvi Tanszék – Inter Nonprofit Kft. – MSZT.

Bata, S. (2009). A társalgás fonetikai jellemzőinek alakulása a beszédpartnerek életkorának függvényében. In T. Váradi (Ed.), *III. Alkalmazott Nyelvészeti Doktorandusz Konferencia* (p. 3–13). Budapest: MTA Nyelvtudományi Intézet.

Beattie, G. W. (1982). Turn-taking and interruption in political interviews: Margaret thatcher and jim callaghan compared and contrasted. *Semiotica*, *39*, 93–114.

Beke, A. (2014). *Beszélődetektálás magyar nyelvű spontán társalgásokban. PhD-értekezés*. Budapest: ELTE.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer. URL: http://www.fon.hum.uva.nl/praat/.

Bóna, J. (2014). Megakadásjelenségek az életkor, a nem és a beszédtípus függvényében. *Beszédkutatás*, (p. 123–143).

Bóna, J. (2015). Különböző beszédtípusok temporális sajátosságai az életkor és a nem függvényében. *Magyar Nyelvőr*, *139*, 201–213.

Cantrell, L. (2013). The power of rapport: an analysis of the effects of interruptions and overlaps in casual conversation. *Innervate*, *6*, 74–85.

URL: https://www.nottingham.ac.uk/english/documents/innervate/13-14/06-lucy-cantrell-q33103-pp-74-85.pdf.

Çetin, O., & Shriberg, E. (2006). Analysis of overlaps in meetings by dialog factors, hot spots, speakers, and collection site: insights for automatic speech recognition. In *Proceedings of INTERSPEECH 2006* (p. 293–296). URL: https://pdfs.semanticscholar.org/14a1/18c6427e3d56471a18eaf119e8bccea08045.pdf.

Cutrone, P. (2013). *Assessing pragmatic competence in the Japanese EFL Context: Towards the learning of listener responses*. Newcastle: Cambrigde Scholar Publishing.

Duranti, A. (1997). *Linguistic Anthropology*. Cambridge: Cambridge University Press.

Dér, Cs. I. (2012). Beszélőváltások során használt diskurzusjelölők a magyar spontán beszédben. *Beszédkutatás*, (p. 130–141).

French, P., & Local, J. (1983). Turn-competitive incomings. *Journal of Pragmatics*, *7*, 701–715.

Gravano, A., & Hirschberg, J. (2009). Backchannel-inviting cues in task-oriented dialogue. In *Proceedings of Interspeech 2009* (p. 253–261). URL: https://www.isca-speech.org/archive/archive_papers/interspeech_2009/papers/i09_1019.pdf.

Gráczi, T. E., & Bata, S. (2010). The effect of familiarization on temporal aspects of turn-taking: a pilot study. *Acta Linguistica Hungarica*, *57*, 307–328.

Hámori, Á., & Horváth, V. (2019). Társalgás, beszélőváltás és diskurzusszerveződés új megközelítésben – fonetikai jellemzők és pragmatikai tényezők összefüggései magyar társalgásokban (pilot study. *Beszédkutatás*, *27*, 134–153.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*, 555–568.

Hilton, C. (2016). The perception of overlapping speech: the effect of speaker prosody and listener attitudes. In *Proceedings of Interspeech 2016* (p. 1260–1264). doi:https://www.isca-speech.org/archive/Interspeech_2016/pdfs/1456.PDF.

Horváth, V., Krepsz, V., Gyarmathy, D., Hámori, Á., Bóna, J., Dér, Cs. I., & Weidl, Zs. (2019). Háromfős társalgások annotálása a bea-adatbázisban: elvek és kihívások. *Nyelvtudományi Közlemények*, *115*, 255–274.

Hutchby, I., & Woofitt, R. (2006). *Conversation analysis: principles, practices and applications*. Cambridge: Polity Press.

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

Markó, A. (2006). Beszélőváltás a társalgásban. előadás ix. pszicholingvisztikai és alkalmazott nyelvészeti nyári egyetemen. *Balatonalmádi*, (p. 21–24). URL: http://fonetika.nytud.hu/wp-content/uploads/2016/04/ma_2.pdf.

Maynard, S. K. (1997). Analyzing interactional management in native/non-native english conversation: A case of listener response. *International Review of Applied Linguistics in Language Teaching*, *35*, 37–60.

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

Schegloff, E. (1979). The relevance of repair to syntax-for-conversation. In T. Givon (Ed.), *Syntax and Semantics* (p. 261–286). New York: Academic Press volume 12.

Schegloff, E. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, *29*, 1–63.

Shriberg, E., Stolcke, A., & Baron, D. (2001). Observations on overlap: findings and implications for automatic processing of multi-party conversation. In *Proceedings of EUROSPEECH* (p. 1359–1362). URL: http://www.icsi.berkeley.edu/ftp/global/pub/speech/.

Tannen, D. (1983). When is an overlap not an interruption? one component of conversational style. In R. Pietro, W. Frawley, & A. Wedel (Eds.), *In the first Delaware symposium on language studies* (p. 119–129). Newark: University of Delaware Press.

Tannen, D. (2012). Turn-taking and intercultural discourse and communication. In C. Paulston, S. Kiesling, & E. Rangel (Eds.), *The handbook of intercultural discourse and communication* (p. 135–157). Chicester: Wiley–Blackwell.

Tolcsvai Nagy, G. (2001). *A magyar nyelv szövegtana*. Budapest: Nemzeti Tankönyvkiadó.

Ward, N., & Tsukahara, W. (2000). Prosodic features which cue back-channel responses in english and japanese. *Journal of Pragmatics*, *32*, 1177–1207.

Weidl, Zs. (2019). Egyet nem értések fonetikai jellemzői középiskolások beszédében. In T. Váradi, Zs. Ludányi, & T. E. Gráczi (Eds.), *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2019. XIII. Alkalmazott Nyelvészeti Doktoranduszkonferencia* (p. 133–142). Budapest: MTA Nyelvtudományi Intézet.

Wells, B., & Macfarlane, S. (1998). Prosody as an interactional resource: turn-projection and overlap. *Language and Speech*, *41*, 265–294.

Yngve, V. H. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting Chicago Linguistics Society* (p. 567–78).

Disfluent whole-word repetitions across the lifespan: Durational patterns and functions

Judit Bóna[1], Tímea Vakula[1]

[1] *ELTE Eötvös Loránd University, Budapest, Hungary*

Abstract

Frequency of disfluencies in speech is often analysed in different age groups, but functions of certain types of disfluencies are rarely examined. The aim of this study is to analyse durational patterns and functions of the repeated words in diverse age groups (from preschool children to elderly speakers). Functions of repetitions are well determinable from their durational patterns and pauses around the first and second instances of the repeated words. Results show that disfluent word-repetitions are good predictors for detecting some age-dependent changes in speech production process.

Keywords: whole-word repetition, durational patterns, function, age

1. Introduction

During spontaneous speech, speakers often produce disfluencies. They can occur for several reasons, and they are of various types. They reflect the speakers' speech planning process, as well (Levelt, 1989). The occurrence of disfluencies is influenced by numerous factors such as age (de Andrade & de Oliveira Martins, 2007, 2010). Across the lifespan, cognitive processes change which also affects language abilities of a person (Craik & Bialystok, 2006). During first language acquisition, the vocabulary of children is constantly growing, and more complex grammatical structures are acquired. As spontaneous speech becomes more complex, children learn to hesitate by copying disfluencies of adults (Haynes & Hood, 1977; Ratner & Sih, 1987). This means that they learn how to fill in the time resolving the speech planning problems. In the speech of young children, long silent pauses and multiple whole-word repetitions are typ-

Email addresses: bona.judit@btk.elte.hu (Judit Bóna), vakula.timi@gmail.com (Tímea Vakula)

ical (Kowal et al., 1975; DeJoy & Gregory, 1985; Horváth, 2006). After young adulthood, more cognitive changes take place. During aging, lexical retrieval abilities deteriorate because of the slowing cognitive processes and the change of the capacity of the memory (Burke et al., 1991). These might also influence the occurrence of speech errors and disfluencies. However, there are contradictory findings regarding significant differences in the frequency of disfluencies when comparing young and elderly speakers. Some authors didn't find any differences while others found that the elderly typically produce more disfluencies (Duchin & Mysak, 1987; Leeper & Culatta, 1995; Yairi & Clifton, 1972).

It is our assumption that not only frequency and/or types of disfluencies can vary in different ages, but their functions differ as well. For the analysis of this question, disfluent whole-word repetition will be examined as it is one of the most frequent disfluencies in speech (Shriberg, 1995; Branigan et al., 1999). Repetitions can occur at different levels of the speech planning process. For example, they can indicate word-finding problem, difficulty in conceptual planning, or covert self-monitoring (Plauché & Shriberg, 1999; Gyarmathy, 2009). In addition, their function could be various which is well determinable from their durational patterns and pauses around the first and second instances of the repeated words.

Whole-word repetitions could be single and multiple repetitions. In case of single repetitions, the repeated word occurs twice. In case of multiple repetitions, it can occur three or four, or occasionally more times. In the latter case, the speaker has more serious difficulties with speech planning. In Example (1) (Bóna, 2013), the speaker tried to remember a heard text, but she didn't succeed (FIL = filled pause, double letters indicate prolongation). Multiple repetitions might occur when at the beginning of the interview the speaker is thinking about what to say (Example 2) (Bóna, 2013). In the examples, disfluencies are bold.

(1) valami ilyesmiről volt szó **SIL** és akkor ez mit hozott ki mi lett belőle SIL **FIL** SIL érdeke**sss** kutatási dolog d**ee** SIL **ez** ezt **úgy úgy** SIL **úgy** nem tudom

'the topic was something like that SIL and than how this ended what this resulted in SIL FIL SIL interestinggg research topic buuut SIL I don't quite quite SIL quite know i it'

(2) **FIL** akkor eddig SIL **FIL hát én én én** ugye ezerkilencszáz-harminchétben születtem tehát hetvenegy éves vagyok

'FIL until now SIL FIL well I I I you know was born in 1937 so I am 71'

The main parts of the repetitions are the following: the original utterance, the first instance of the repeated word, the second instance of the repeated word and the continuation of the utterance (Plauché & Shriberg, 1999). Optional pauses may also occur next to the main parts (Plauché & Shriberg, 1999). Example (3) shows the main parts of a disfluent whole-word repetition (SIL = silent pause, P = pause, R1 = first instance of the repeated word, R2 = second instance of the repeated word):

(3) *There is a book* *SIL* ***on*** *SIL* ***on*** *SIL the table.*
Original utterance P1 R1 P2 R2 P3 Continuation

The duration of the first and second instances of the repeated words and the occurrence of pauses are related to the function of the repetitions. Heike (1981) suggested two types of repetitions – bridging a gap and hesitating. In the first case, the repeated word provides a bridge to the continuation of the utterance after hesitating (retrospective repeat). In the second case, the repeated word is the hesitation itself (prospective repeat). This means that it fills the gap between the original utterance and the continuation. The occurrence of these two functions are indicated by pauses which appear next to the repeated items. In retrospective repeats, there is a pause before the repeated word, but there

isn't one after it. In prospective repeats, there is a pause after the repeated word.

Plauché & Shriberg (1999) suggested more other functions according to the phonetic features of the repetitions. They analysed the durational patterns, f0-variation and glottalization, and on the basis of these results they found three main types of functions: canonical repetition, covert self-repair, and stalling repetition (Table 1).

In canonical repetitions (Plauché & Shriberg, 1999), the duration of R1 is much longer than in the utterance of the same word in fluent speech. The duration of R2 is similar to the fluent word. There might be a pause before R1, there is a long pause between R1 and R2, and there is no pause after R2. Both R1 and R2 are characterized by falling intonation, and R1 is often characterized by creak-like voicing modality (similar to filled pause). In this case the speaker has difficulties in speech, stops during the pronunciation of the word (R1), lengthens it, and after having solved the problem they continue speaking with repeating the last lengthened word. This type corresponds to Heike's *retrospective repeat* (1981).

In covert self-repairs (Plauché & Shriberg, 1999), P1 often occurs, but P2 and P3 don't occur. R1 and R2 are slightly long, they are slightly longer than they are in fluent speech, and their duration is similar to each other. R1 and R2 are both characterized by rising pitch. R1 is pronounced sometimes with glottalization. In this case the speaker detects a problem during the pronunciation of R1, this is shown by a possible preceding pause and glottalization. The speaker makes an effort to correct it, and "R2 usually marks the beginning of a new utterance or a corrected version of the previous one" (Plauché & Shriberg, 1999, 1516.).

In stalling repetitions (Plauché & Shriberg, 1999), there is no pause before R1, but P2 and P3 might occur. The duration of R1 is slightly longer than in fluent speech, and the duration of R2 is much longer. R1 is characterized by dropping in pitch. The speech is fluent during the pronunciation of R1, the speaker has a problem during and/or after the production of R2. This is usually

marked by P3 or other possible disfluencies after R2. This type looks as if it was the inverse of canonical repetitions, and corresponds to Heike's *prospective repeat* (1981).

Table 1: The structures of the three types of word-repetitions (examples with 'the') '+' = a longer than fluent duration. ' - ' = no pause. (Based on Plauché & Shriberg (1999))

Type	**Structure**
Canonical repetition	(Original Utterance) (Possible Pause) **the**+++ (Long Pause) **the** (-) (Continuation)
Covert self-repair	(Original Utterance) (Often Pause) **the**+ (-) **the**+ (-) (Continuation)
Stalling repetition	(Original Utterance) (-) **the**+ (Possible Pause) **the**+++ (Possible Pause) (Continuation)

Our starting point was that the characteristics of repetitions (like other disfluencies) are influenced by speakers' age. In the case of young children, the numerous and multiple repetitions refer to problems of lexical retrieval and grammatical formulation (Horváth, 2017a). The function of the repeated instance is mainly stalling and hesitation. Young adults speak more fluently, they have significantly less problems with speech planning, and they need less time to resolve them. In their case, the function of the repeated instance is bridging the gap between the original utterance and the continuation, or self-monitoring. In case of elderly, word retrieval becomes difficult again (Burke et al., 1991; Burke & Shafto, 2004). They repeat words for gaining time to solve the problem.

Durational patterns of word-repetitions have been analysed mainly in the speech of young adults (e.g. Shriberg, 1995; Benkenstein & Simpson, 2003; Gyarmathy, 2009), few similar measurements are known at different ages (Bóna, 2015; Horváth, 2017b; Bóna & Vakula, 2018). Horváth (2017b) analysed the durational patterns of word-repetitions in the speech of 6-9-year-old children. She found that age didn't affect the duration of editing phases, but there are big individual differences in the durations in each age group. In another study (Bóna

& Vakula, 2018), durational patterns and functions of repetitions were compared in four age groups and two speech tasks. The age groups were the following: schoolchildren (9-year-olds), adolescents (13-14-year-olds), young adults (20-25-year-olds), and old speakers (75+). The two speech tasks were spontaneous narratives about the own lives and narrative recalls of heard texts. Results show that speech tasks didn't influence the characteristics of repetitions (the ratio of the repeated words and the duration of editing phases were similar in both speech tasks) while the age of speakers did. There were differences in the durational patterns (duration of the repeated words and pauses) and functions between the age groups in both speech tasks.

The above mentioned paper (Bóna & Vakula, 2018) analysed four age groups whose speech planning processes were examined by several papers. However, our knowledge is scarce about the disfluencies of preschool children, middle-aged, and young-old speakers. In this paper therefore we try to examine their speech, as well. The aim of this study is to analyse durational patterns, functions and linguistic characteristics (function or content word) of the repeated words in diverse age groups – from preschool children to elderly speakers.

Our hypotheses are that (i) all speakers mostly repeat function words, but (due to their slower or more difficult word-retrieval processes) the ratio of the repetitions of content words is higher in the speech of children and senior speakers than in that of young adults. (ii) Across the lifespan, the durational patterns of repetitions change: a) the pauses between the two instances of the repeated words and b) the ratio of the duration of the two repeated instances. The children and the elderly will have longer editing phases, and the difference between the duration of the two instances will be smaller. (iii) The ratio of the diverse functions differs between the age groups. The children and the elderly will have a higher rate of stalling function than young adults.

2. Procedure

2.1. Material

Speech samples were selected from the GABI Hungarian Children Speech Database and Information Repository (Bóna et al., 2014) and BEA Hungarian Speech Database (Gósy, 2012). Participants were asked to speak about their own lives, hobbies and families. Their speech was rarely interrupted, only when they were having difficulties in continuing, so spontaneous narrative could be recorded. Altogether 380 min of recordings were analysed.

2.2. Participants

Speech samples were selected from 140 speakers altogether. They were from seven age groups: pre-schoolers (4-5-year-olds), school children (9-year-olds), adolescents (13-14-year-olds), young adults (20-25-year-olds), middle-aged adults (45-53-year-old), senior adults (60-65-year-olds), and elderly adults (75+-year-olds). In every age group there were 20 speakers – 10 women and 10 men. They were native Hungarian speakers with normal hearing and without any known mental or speech disorders. Even senior adults didn't show any mental disorders. The adults were all of similar educational background. They all completed at least 12 years of education.

2.3. Method

Disfluent whole-word repetitions were annotated in speech samples by Praat (Boersma & Weenink, 2008). Duration of the first instance of the repeated word (R1), the second instance of the repeated word (R2), the pause between R1 and R2 (P2), and the pause between R2 and the continuation (P3) were measured as well. The durations of the repeated instances were measured from the beginning of the word to the end of the word.

Functions and the types of repeated words (function word or content word) were classified. Functions were defined on the basis of Plauché & Shriberg (1999). Since they found that phonetic parameters (durational patterns, pausing, f0-variation) show connection with functions, we based our analysis on these

factors. Those phenomena in which R1 was significantly longer than R2, and there was a pause between them, but there was no pause after R2, and thus were classified as canonical repetition. Those phenomena in which the duration of R1 and R2 was similar, in which there was no pause between R1 and R2 and after R2, were classified as covert-self repair. Those phenomena in which R2 was longer than R1, were classified as stalling repetition. In these cases, often, but not necessarily, there was a pause between R2 and the continuation. Those phenomena which could not be classified into these main categories, were considered as "Others".

Next, durational patterns of the certain types were analysed in the age groups. The duration of the components of repetitions (pauses and first and second instances of repeated words) were measured by Praat (Boersma & Weenink, 2008). The time interval between two words was considered as pause irrespective of its duration, and whether it was silent or filled with sound (not words) (Fletcher, 2010).

The ratio of R2 and R1 was calculated in each repetition. Calculating the ratio was necessary for the analysis to disclose whether there were significant differences between the age groups in the duration of R1 and R2 and their relations to each other. If purely their duration had been compared, it would not have resulted in relevant data. On the one hand, tempo differences between the age groups would have caused differences in durational patterns. On the other hand, alternatively to Shriberg (1999) and Plauché & Shriberg (1999), not only the repetitions of the words *I* and *the* were analysed. This wouldn't have been possible in Hungarian, since Hungarian is an agglutinative language and as such it doesn't require to use the personal pronoun with verbs. Each occurrence of any kind of word-repetition was analysed, this is why the raw data of the different groups could not be compared. Speakers of different groups repeated different words, so differences were caused by the divergent length and phonetic characteristics of words. For example, if young speakers had repeated multisyllabic words in faster speech rate and elderly speakers repeated monosyllabic words in slower speech rate, the comparison of the raw duration of R1 and R2

would have resulted in false results. The ratio resolved these problems, and it could be compared between the age groups. The ratio of the second instance of the repeated word and the first instance of the repeated word and the pauses (before) between and after them were analysed.

Statistical analysis was carried out by SPSS on 95% confidence level.

2.4. Reliability

The annotation of repetitions, the coding of types of words and the definition of functions were carried out by the two authors individually. The results were compared with 94% agreement. In cases where there was not agreement, a third party was involved in the decision making.

3. Results

There were 566 disfluent whole-word repetitions in the analysed speech samples. There were far fewer repetitions in the speech of children and adolescents than in the speech of adults (perhaps due to their speech consisting of shorter connected sections). The number of occurrences of word-repetitions was 45 in 4-5-year-olds, 32 in 9-year-olds, 21 in 13-year-olds, 160 in 20-30-year-olds, 102 in 45-53-year-olds, 112 in 60-65-year-olds, and 94 in 75+-year-olds. The frequency cannot be inferred from the numbers as they are of speech samples of different lengths and from speakers with various speech- and articulation rate. Children, adolescents and young-old speakers repeated content words in higher ratio than young, middle-aged and old-old adults (Table 2). Example (6), (7), (8) contain repetitions of function words. Example (9), (10), (11) contain repetitions of content words.

(4) van egy hely és SIL és FIL és ott mindent lehet játszani (5-year-old boy)
'there is a place and SIL and FIL and there you can play anything you want'

(5) pedig fizetnek azért amit 1082 amit hallgatnának (20-year-old man)
'but they pay for what SIL what they study'

(6) alkalom a SIL FIL SIL a SIL egyetemi tanulmányaim folytatására (75+-year-old woman)
'there was no chance to continue the SIL FIL SIL the SIL university studies'

(7) szoktam játszani játszani a barátokkal (5-year-old girl)
'I usually play play with friends'

(8) találkozott egy kiskacsa kiskacsa a másik kiskacsával (9-year-old boy)
a duckling duckling met another duckling'

(9) mondjuk kedvenc SIL kedvenc országom Olaszország és Franciaország (75+-year-old woman)
'let's say my favourite SIL favourite countries are Italy and France'

Table 2: Types of the repeated words

Age-group	**Content word**	**Function word**
4-5-year-olds	11.1%	88.9%
9-year-olds	28.1%	71.9%
13-year-olds	23.8%	76.2%
20-30-year-olds	3.8%	96.3%
45-53-year-old	2.0%	98.0%
60-65-year-olds	11.6%	88.4%
75+-year-olds	6.4%	93.6%

Duration of R1 and R2 was analysed in all repetitions (Table 3). There was no significant difference between R1 and R2 in the speech of 4-5-year-old, 9-year-old, and 75+-year-old speakers. There were significant differences between

the durations of R1 and R2 in the speech of 13-year-old [repeated measures ANOVA: $F(1, 20) = 28.755$; $p < 0.001$; $\eta^2 = 0.590$], 20-30-year-old [repeated measures ANOVA: $F(1, 159) = 91.871$; $p < 0.001$; $\eta^2 = 0.366$], 45-53-year-old [repeated measures ANOVA: $F(1, 101) = 16.190$; $p < 0.001$; $\eta^2 = 0.138$], and 60-65-year-old (Wilcoxon Signed Ranks Test: $Z = -3.120$; $p = 0.002$) speakers.

Table 3: Duration of R1 and R2 depending on age (ms) (Mean and Standard Deviation)

	Spontaneous narratives	
	R1	**R2**
4-5-year-olds	716 (408)	647 (350)
9-year-olds	430 (199)	378 (212)
13-year-olds	448 (179)	255 (113)
20-30-year-olds	344 (144)	232 (109)
45-53-year-olds	320 (133)	264 (118)
60-65-year-olds	364 (212)	322 (167)
75+-year-olds	362 (237)	334 (181)

To be able to compare how the duration of R1 and R2 relate to each other in different age groups and speech tasks, the ratio of R2 and R1 was calculated (Figure 1). If the ratio was less than 100%, R1 was longer than R2. If the ratio was more than 100%, than R2 was longer than R1. The average ratio of R2 and R1 was 93% (SD: 35.2) in 4-5-year-olds, 92% (SD: 9.1) in 9-year-olds, 63% (SD: 6.5) in 13-year-olds, 75% (SD: 3.2) in 20-30-year-olds, 91% (SD: 43.6) in 45-53-year-olds, 98% (SD: 44.2) in 60-65-year-olds, and 107% (SD: 6.3) in 75+-year-old speakers. This means that the lowest mean values were calculated in 13-year-olds and 20-30-year-olds, while the highest values were calculated in 75+-year-olds. Table 4 shows the significant differences as results of the statistical analysis. There were no significant differences between the age groups in any other cases.

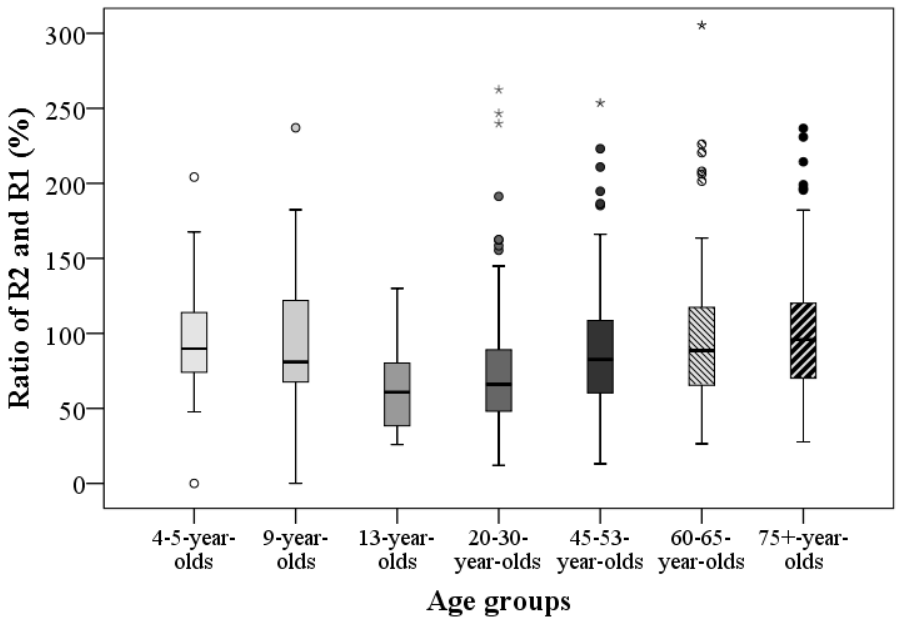


Figure 1: Ratio of the durations of R2 and R1 (R2 = duration of the second instance of the repeated word, R1 = duration of the first instance of the repeated word)

Table 4: Significant differences between the age groups in the ratio of R2 and R1 (Results of the Mann–Whitney-test)

	4-5-year-olds	**9-year-olds**	**13-year-olds**	**20-30-year-olds**	**45-53-year-olds**	**60-65-year-olds**	**75+-year-olds**
4-5-year-olds	_	_	$Z = -6.505$; $p < 0.001$	$Z = -3.840$; $p < 0.001$	_	_	_
9-year-olds	_	_	$Z = -2.237$; $p = 0.025$	$Z = -2.805$; $p = 0.005$	_	_	$Z = -2.365$; $p = 0.018$
13-year-olds	$Z = -6.505$; $p < 0.001$	$Z = -2.237$; $p = 0.025$	_	$Z = -7.438$; $p < 0.001$	$Z = -7.199$; $p < 0.001$	$Z = -7.250$; $p < 0.001$	$Z = -3.866$; $p < 0.001$
20-30-year-olds	$Z = -3.840$; $p < 0.001$	$Z = -2.805$; $p = 0.005$	$Z = -7.438$; $p < 0.001$	_	$Z = -3.501$; $p < 0.001$	$Z = -5.197$; $p < 0.001$	$Z = -5.437$; $p < 0.001$
45-53-year-olds	_	_	$Z = -7.199$; $p < 0.001$	$Z = -3.501$; $p < 0.001$	_		$Z = -2.047$; $p = 0.041$
60-65-year-olds	_	_	$Z = -7.250$; $p < 0.001$	$Z = -5.197$; $p < 0.001$	_	_	_
75+-year-olds	_	$Z = -2.365$; $p = 0.018$	$Z = -3.866$; $p < 0.001$	$Z = -5.437$; $p < 0.001$	$Z = -2.047$; $p = 0.041$	_	_

Editing phases (p2) of all repetitions were also analysed (Figure 2). The longest editing phases were produced by 9-year-olds. According to the statistical analysis, there were significant differences between 9-year-olds and 20-30-year-olds (Mann–Whitney-test: $Z = -2.805$; $p = 0.005$), between 9-year-olds and 45-53-year-olds (Mann–Whitney-test: $Z = -3.176$; $p = 0.001$), between 9-year-olds and 60-65-year-olds (Mann–Whitney-test: $Z = -2.383$; $p = 0.017$), and between 9-year-olds and 75+-year-olds (Mann–Whitney-test: $Z = -2.365$; $p = 0.018$) in the durations of editing phases. There were no significant differences between the other age groups.

The majority of editing phases was realized as silent pause in each age group. In 20-30-year-olds and 75+-year-olds, the ratio of silent editing phases was higher than in the other age groups (Figure 3). There were some cases in each group where editing phase also contained filled pauses. Due to the rare occurrence of these cases, Figure 5 shows in one category the cases in which the editing phase was only a filled pause and the cases in which filled pause and silent pause occurred together (one after the other).

Functions of repetitions varied among the age groups. Adolescents, young and middle-aged adults produced canonical repetitions in higher ratio than the other groups. Children, middle-aged adults and elderly (both 60-65 and 75+) speakers produced more stalling repetitions than adolescents and young adults. Other types of repetitions (which could not be categorized into the three main types) occurred in higher ratio in pre-schoolers and elderly speakers (Table 5).

Durational patterns of the three main types of repetitions were analysed (Table 6). In case of *canonical repetitions*, editing phases (p2) were significantly longer in the speech of children than in the speech of other speakers (Table 7). There were no significant differences between the age groups above the age of 13 (the only exception was the significant difference shown between 20-30-year-olds and 45-53-year-olds). In the ratio of R2 and R1, there were significant differences only between 20-year-olds and 4-5-year-olds [UNIANOVA showed significant differences between the groups: $F(6, 237) = 5.477$; $p < 0.001$, according to the Tukey post hoc test between 20- and 4-5-year-olds: $p = 0.001$], between 20-

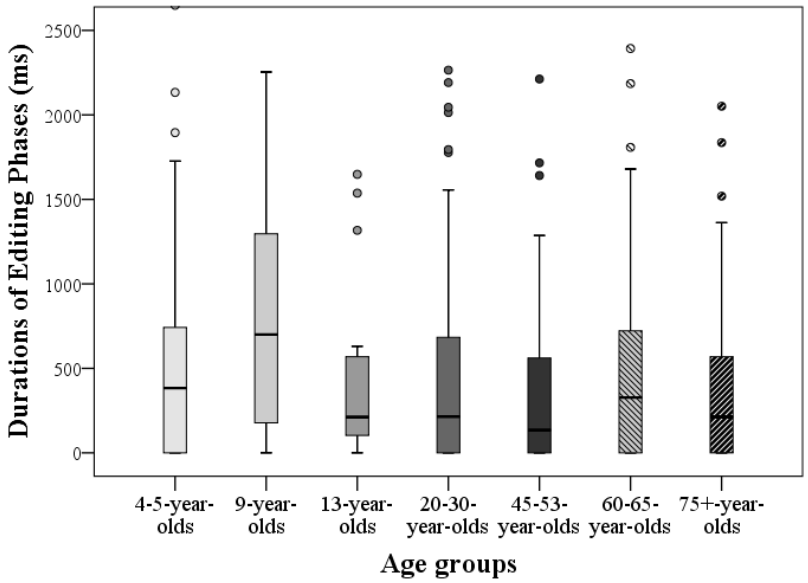


Figure 2: Durations of editing phases of every repetitions (ms)

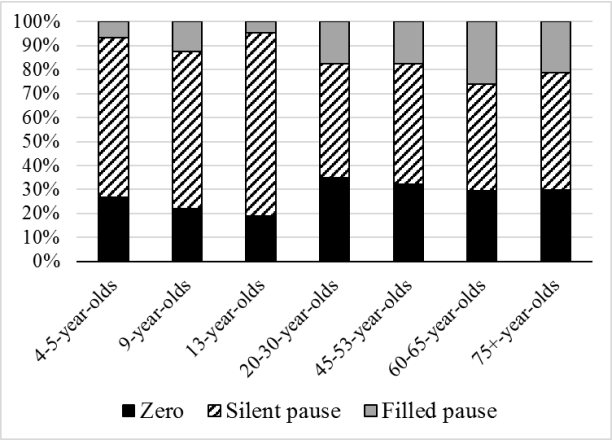


Figure 3: Types of editing phases (p2)

Table 5: Functions of repetitions

Age-group	Canonical repetitions	Covert self-repairs	Stalling repetitions	Others
4-5-year-olds	35.6%	17.8%	26.7%	20.0%
9-year-olds	46.9%	18.8%	31.3%	3.1%
13-year-olds	61.9%	19.0%	19.0%	0.0%
20-30-year-olds	50.0%	27.5%	16.3%	6.3%
45-53-year-old	48.0%	18.6%	32.4%	1.0%
60-65-year-olds	35.7%	16.1%	39.3%	8.9%
75+-year-olds	31.9%	16.0%	39.4%	12.8%

year-olds and 60-65-year-olds ($p = 0.006$), and between 20-year-olds and 75+-year-olds ($p = 0.005$). In case of *covert-self monitoring*, editing phases were 0 ms. There were no significant differences between the age groups in the ratio of R2 and R1. In case of *stalling repetitions*, there were no significant differences between the age groups in editing phases. As regards the ratio of R2 and R1, there was significant difference only between 20-30-year-olds and 75+-year-olds ($Z = -2.011$; $p = 0.044$).

4. Discussion

In the analysis, disfluent whole-word repetitions were analysed in different age groups. Our question was whether functions of whole-word repetitions are different in diverse age groups and whether this occurs in the durational patterns of repetitions.

The first hypothesis was confirms as results show that although repetition of function words occurred in the highest ratio in each age group, children, adolescents and elderly speakers repeated content words in much higher ratio than young and middle-aged adults. According to the literature (e.g. Burke et al., 1991; McGregor, 1997; Barresi et al., 2000), the former age groups have more difficulties with lexical access than the latter age groups. Thus, the ratio of

Table 6: Duration of editing phases and ratio of R2 and R1

	p2 (Editing phase) (ms)	Ratio of R2 and R1 (%)
Canonical repetitions		
4-5-year-olds	943 (218)	76 (4)
9-year-olds	1421 (313)	62 (8)
13-year-olds	630 (148)	53 (6)
20-30-year-olds	636 (60)	53 (2)
45-53-year-old	454 (65)	63 (3)
60-65-year-olds	508 (73)	68 (3)
75+-year-olds	607 (158)	69 (4)
Covert self-repairs		
4-5-year-olds	–	75 (6)
9-year-olds	–	82 (5)
13-year-olds	–	59 (7)
20-30-year-olds	–	77 (4)
45-53-year-old	–	86 (5)
60-65-year-olds	–	77 (5)
75+-year-olds	–	75 (6)
Stalling repetitions		
4-5-year-olds	464 (170)	107 (15)
9-year-olds	744 (209)	146 (14)
13-year-olds	195 (67)	219 (105)
20-30-year-olds	441 (96)	119 (9)
45-53-year-old	360 (77)	135 (8)
60-65-year-olds	585 (83)	127 (7)
75+-year-olds	441 (80)	150 (12)

Table 7: Significant differences between the age groups in the duration of editing phases

	4-5-year-olds	**9-year-olds**	**20-30-year-olds**
13-year-olds		$Z = -2.326$; $p = 0.020$	
20-30-year-olds		$Z = -3.041$; $p = 0.002$	
45-53-year-olds	$Z = -2.635$; $p = 0.008$	$Z = -3.875$; $p < 0.001$	$Z = -2.104$; $p = 0.035$
60-65-year-olds	$Z = -2.086$; $p = 0.037$	$Z = -3.383$; $p = 0.001$	
75+-year-olds		$Z = -3.226$; $p = 0.001$	

the repeated content words can refer to word finding problems. With repeating content words, children and elderly probably monitor if they said the appropriate word or they solve a greater difficulty in speech planning than young and middle-aged adults.

The second hypothesis was that the durational patterns of repetitions change across the lifespan. This was analysed together with the functions of repetitions—the third hypothesis. The differences of durational patterns reflect the differences of functions between the age groups. The fact that adolescents, young- and middle-aged adults produced more canonical repetitions than the other age groups, shows that they can monitor speech planning problems earlier, and the duration of R1 and p2 is enough for solving them.

Children and speakers over age 45 produced stalling repetitions more frequently than adolescents and young adults. This type refers to the assumption that they try to solve the speech planning problems later, during the pronunciation of R2. Often the duration of R2 is not enough for the solution. particular results were shown by the middle-aged group. Their results show that they make a transition between young adults and elderly speakers.

The analysis of durational patterns dependent on function shows that there are only slight differences between the age groups. Covert self-repair is similar in each age group. In case of canonical repetitions, the longer pauses in the speech of preschool children might be caused by slower speech rate. However, pauses are not different in the other groups despite the different speech and articulation rates. On the other hand, pauses might show that preschool children need more time to solve the speech planning problems. The differences in the ratio of R2 and R1 between the groups refer to the fact that R1 is much longer than R2 in the speech of young adults, and R1 is less long than R2 in the speech of pre-schoolers and elderly speakers. It seems that the function of R2 is not only a bridge between the original utterance (and R1) and the continuation, but also it gives time for further planning.

In case of stalling repetitions, there are no differences between the age groups in the duration of p2. This means that speakers do not intend to keep a long pause during solving the problem, so they repeat the last item of the original utterance (R1, the repetition is R2) to fill the gap. The ratio of R2 and R1 is significantly higher in elderly speech than in that of young adults. This correctly indicates the assumption that the elderly need more time for solving the speech planning or monitoring problems.

5. Conclusions

Our findings lead to the conclusions that (i) disfluent word-repetitions are good predictors for detecting some age-dependent changes in speech production process reflected by repetitions, (ii) children's use of repetitions differ from the adults' ones in a number of properties demonstrating a developing speech planning mechanism, (iii) elderly speakers' repetitions differ from young speakers' ones in function which shows their cognitive changes and different speech planning processes.

Acknowledgement

The author wishes to thank Ágnes Jordanidisz, Zsófia Koren-Dienes and Cheryl Winget for their help in preparing this paper. This research was supported by the Hungarian National Research, Development and Innovation Office of Hungary [project No. K-120234] and the Thematic Excellence program of the ELTE Eötvös Loránd University, Budapest, Hungary.

References

de Andrade, C. R. F., & de Oliveira Martins, V. (2007). Fluency variation in adolescents. *Clinical Linguistics & Phonetics*, *21*, 771–782.

de Andrade, C. R. F., & de Oliveira Martins, V. (2010). Speech fluency variation in elderly. *Pro-fono: revista de atualizacao cientifica*, *22*, 13–18.

Barresi, B. A., Nicholas, M., Connor, L. T., Obler, L. K., & Albert, M. L. (2000). Semantic degradation and lexical access in age-related naming failures. *Aging, Neuropsychology, and Cognition*, *7*, 169–178.

Benkenstein, R., & Simpson, A. P. (2003). Phonetic correlates of self-repair involving word repetition in German spontaneous speech. In R. Eklund (Ed.), *Proceedings of DiSS '03. Disfluency in Spontaneous Speech Workshop* (p. 81–84). Göteborg: Göteborg University.

Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer. URL: http://www.fon.hum.uva.nl/praat/download_win.html version 5.0.1).

Bóna, J. (2013). *A spontán beszéd sajátosságai az időskorban. Beszéd – Kutatás – Alkalmazás 2.*. Budapest: ELTE Eötvös Kiadó.

Bóna, J. (2015). Ismétlések mint megakadások fiatalok, idősödők és idősek beszédében. In M. Gósy (Ed.), *Diszharmóniás jelenségek a beszédben* (p. 149–169). Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences (RIL HAS).

Bóna, J., Imre, A., Markó, A., Váradi, V., & Gósy, M. (2014). GABI – Gyermeknyelvi Beszédadatbázis és Információtár. *Beszédkutatás*, (p. 246–252).

Bóna, J., & Vakula, T. (2018). Durational patterns and functions of disfluent word-repetitions: The effect of age and speech task. In M. Gósy, & T. E. Gráczi (Eds.), *Challenges in analysis and processing of spontaneous speech* (p. 169–184). Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences (RIL HAS).

Branigan, H., Lickley, R., & McKelvie, D. (1999). Non-linguistic influences on rates of disfluency in spontaneous speech. In *Proceedings of the 14th International Conference of Phonetic Sciences* (p. 387–389).

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults. *Journal of Memory and Language*, *30*, 542–579.

Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current directions in psychological science*, *13*, 21–24.

Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in cognitive sciences*, *10*, 131–138.

DeJoy, D. A., & Gregory, H. H. (1985). The relationship between age and frequency of disfluency in preschool children. *Journal of Fluency Disorders*, *10*, 107–122.

Duchin, S. W., & Mysak, E. D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of communication disorders*, *20*, 245–257.

Fletcher, J. (2010). The prosody of speech: Timing and rhythm. In *The Handbook of Phonetic Sciences, Second Edition* (p. 521–602).

Gósy, M. (2012). BEA – A multifunctional Hungarian spoken language database. *Phonetician*, *106*, 50–61. URL: http://www.isphs.org/Phonetician/Phonetician_105_106.pdf.

Gyarmathy, D. (2009). A beszélő bizonytalanságának jelzései: ismétlések és újraindítások. [Marks of uncertainty of speakers: whole-word repetitions and part-word repetitions. *Beszédkutatás*, (p. 196–216).

Haynes, W. O., & Hood, S. B. (1977). Language and disfluency variables in normal speaking children from discrete chronological age groups. *Journal of Fluency Disorders*, *2*, 57–74.

Heike, A. E. (1981). A content-processing view of hesitation phenomena. *Language and Speech*, *24*, 147–160.

Horváth, V. (2006). A spontán beszéd és a beszédfeldolgozás összefüggései gyerekeknél [Correlation between spontaneous speech and speech perception in children.]. *Beszédkutatás*, (p. 134–146).

Horváth, V. (2017a). Közlések grammatikai szerkesztettsége 6–9 éves gyermekek narratíváiban. [The grammatical structuredness of utterances in the narratives of children aged between 6 and 9]. *Anyanyelv-pedagógia*, *10*, 5–18.

Horváth, V. (2017b). Megakadásjelenségek és időzítési sajátosságaik 6–9 éves gyermekek spontán narratíváiban [Disfluencies and their durational patterns in spontaneous narratives of 6-9-year-old children]. In J. Bóna (Ed.), *Új utak a gyermeknyelvi kutatásokban* (p. 97–120). Budapest: ELTE Eötvös Kiadó.

Kowal, S., O'Connell, D. C., & Sabin, E. J. (1975). Development of temporal patterning and vocal hesitations in spontaneous narratives. *Journal of Psycholinguistic Research*, *4*, 195–207.

Leeper, L. H., & Culatta, R. (1995). Speech fluency: effect of age, gender and context. *Folia phoniatrica et logopaedica*, *47*, 1–14.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge (Massachusetts)–London (England: The MIT Press.

McGregor, K. K. (1997). The nature of word-finding errors of preschoolers with and without word-finding deficits. *Journal of Speech, Language, and Hearing Research*, *40*, 1232–1244.

Plauché, M., & Shriberg, E. (1999). Data-driven subclassification of disfluent repetitions based on prosodic features. In *Proc. International Congress of Phonetic Sciences* (p. 1513–1516). volume 2.

Ratner, N. B., & Sih, C. C. (1987). Effects of gradual increases in sentence length and complexity on children's dysfluency. *Journal of Speech and Hearing Disorders*, *52*, 278–287.

Shriberg, E. (1995). Acoustic properties of disfluent repetitions. *Proceedings of the international congress of phonetic sciences*, *4*, 384–387.

Shriberg, E. (1999). Phonetic consequences of speech disfluency. In *Proceedings of the International Congress of Phonetic Sciences. Vol. 1.* (p. 619–622).

Yairi, E., & Clifton, N. F. (1972). Disfluent speech behavior of preschool children, high school seniors, and geriatric persons. *Journal of Speech, Language, and Hearing Research*, *15*, 714–719.

A diskurzusjelölők a telefonos ügyfélszolgálati beszélgetésekben

Schirm Anita[1]

[1] *Szegedi Tudományegyetem Magyar Nyelvészeti Tanszék*

Abstract

Type, frequency and function of discourse markers (e.g. *hát* „well", *jó* „right", *ugye* „[tag]", *szóval* „so") present important information about the text types. In this paper I show what text type dependent and independent features of discourse markers can be observed in a corpus built from call center conversations. Moreover, I also exhibit how these elements reflect unintentionally the feelings of participants and how these elements improve or deteriorate the efficiency of the communication. On the part of the operator, recording the call requires exact relaying of the information and swift problem solving, and this is aided by reformulation markers, politeness phenomena, and mitigating discourse markers. The recording of the conversation and the semi-institutional situation often makes callers uneasy, which is manifested in their searching for words, and use of disfluency phenomena and mitigating elements. Due to question–answer adjacency pairs, the use of discourse markers employed as general answer markers are frequent in the speech of both types of speakers. However, role-specific usage appears in operators using *jó?* to confirm, and *hát* as a politeness element and mitigator. Operators are required to be objective in their style and to attempt to calm down callers besides providing the required information. Callers often start conversations with a negative attitude, and their strong emotionality, indignation, anger, and irony is clearly detectable in the discourse markers they use.

1. Bevezetés

Az ügyfélszolgálati, vagy más néven call centeres telefonbeszélgetéseket számos szempontból vizsgálták már (Vicsi & Sztahó, 2009; Bechet et al., 2012; Archer & Jagodziński, 2014; Kopparapu, 2015), azonban a bennük előforduló diskurzusjelölőket csak érintőlegesen említették a különböző leírások. Pedig a diskurzusjelölők multifunkcionalitásuk és textuális szerepük révén a szövegtípusok szemaforjainak tekinthetők, azaz a típusuk, a gyakoriságuk és a funkciójuk visszatükrözi a szövegtípus jellegzetességeit (Schirm, 2017). A diskurzusjelölők

Email address: schirmanita@gmail.com (Schirm Anita)

Beszédtudomány – Speech Science 2020. 237–258. DOI: 10.15775/Besztud.2020.237-258

elemzésével nyert információk tehát hozzájárulhatnak a call centeres telefonbeszélgetések szövegtípusának a jobb megismeréséhez, ami pedig elősegítheti az ügyfélszolgálatok kommunikációjának az eredményesebbé válását, továbbá a rohamosan fejlődő mesterségesintelligencia-kutatások szempontjából is haszonnal kecsegtethet.

A telefonbeszélgetések szerkezeti szabályszerűségeit már az 1960-as évektől kezdve vizsgálták. A kutatások eleinte a telefonhívások nyitására és zárására, az egyes fordulók hosszára, a kérdés-válasz szomszédsági párokra, a szóátvételek módjára és a javítási műveletekre korlátozódtak (Schegloff, 1968; Schegloff & Sacks, 1973). A későbbi vizsgálatok a szerkezeti szabályszerűségek mellett már a telefonbeszélgetések olyan pragmatikai jellegzetességeivel is foglalkoztak, mint például a hívások udvariassága, illetve a hatékony kommunikációs stratégiákat is igyekeztek feltárni, valamint az egyes kultúrák közti különbségekkel is foglalkoztak (Luke & Pavlidou, 2002). Kezdetben főként magánjellegű beszélgetéseket elemeztek, később azonban a kutatások az üzleti hívásokra és a call centeres beszélgetésekre is kiterjedtek (Orthaber & Márquez-Reiter, 2015). A minél sikeresebb ügyintézés érdekében manapság pedig már számos ügyfélszolgálaton használnak hangelemző programokat, amelyekkel vizsgálható többek közt a telefonáló felek hanghordozása, valamint az operátor és az ügyfelek beszédének érzelmi jellemzői is feltérképezhetők (Pallotta & Delmonte, 2013).

A számos nemzetközi leírás ellenére a call centeres telefonbeszélgetés szövegtípusának a jellemzése tudomásom szerint hiányzik a magyar szakirodalomból. A telefonbeszélgetésekről általánosságban ugyan jelent már meg magyarul is társalgáselemző tanulmány (Hámori, 2006), ám ahogy ebben is olvasható, a műfajnak több altípusa (pl. telefonos ügyfélszolgálati ügyintézés, telefonos csevegés, információkérés) is létezik, amelyekben a „telefonálás általános szabályai az adott szituáció jellegzetességeivel kombinálódnak" (i. m. 179), így mindre kissé más szabályok jellemzőek. A tanulmányban ezért call centeres párbeszédekből épített korpusz elemzésén keresztül azt mutatom be, hogy a diskurzusjelölők vizsgálatával mit tudhatunk meg a telefonos ügyfélszolgálati beszélgetés szövegtípusának a jellegzetességeiről. Továbbá arról is számot adok,

hogyan tükrözik vissza a diskurzusjelölők a telefonbeszélgetésben résztvevő felek érzelmeit, s hogyan strukturálják a hívásokat.

A kutatás fő kérdése, hogy melyek lesznek a call centeres beszélgetésekben a leggyakoribb diskurzusjelölők. Hipotéziseim szerint a telefonos ügyfélszolgálati beszélgetések kommunikációs helyzetének jellegzetességéből adódóan az operátorok és az ügyfelek más elemeket, más arányban és más célból részesítenek majd előnyben. Azt feltételeztem továbbá, hogy lesznek a diskurzusjelölőknek szövegtípus-specifikus szerepei: az ügyfelek elsődlegesen az elégedettségüknek adnak hangot a beszélői attitűdöket kifejező diskurzusjelölőkkel, míg az operátorok a sikeres ügyintézés és az ügyfelek elégedettsége érdekében főként az ellentétet tompító elemeket használják majd.

2. Anyag, módszer, kísérleti személyek

A kutatás anyagát a Voxindex korpusz 50 darab, magyar nyelvű call center beszélgetéséből származó, összesen 241 percnyi hangrészlet alkotta: volt köztük telefontársaság, infokommunikációs cég, bank, biztosító és áruházlánc által rögzített anyag is. A beszélgetések bizonyos paramétereiről (pl. a résztvevő operátorok, valamint ügyfelek életkoráról) adatvédelmi okok miatt nem állnak rendelkezésemre információk. Az elemzett részletek a teljes telefonhívások meghallgatása után úgy lettek kiválasztva, hogy minél több diskurzusjelölő legyen bennük. A diskurzusjelölők emocionális és expresszív funkciójuknál (Jucker, 1993), valamint szövegszervező szerepüknél (Lenk, 1998) fogva ugyanis kiválóan alkalmasak a beszélgetésben résztvevő felek érzelmeinek a detektálására, valamint a telefonbeszélgetés szerkezeti részeinek az elkülönítésére is, így az elemzésükkel magáról a szövegtípusról is fontos információkhoz lehet jutni. Ez a kiválasztási mód alkalmas a szemléltetésre, ám az így nyert adatok és következtetések a kutatás folytatásakor a későbbiekben még majd összevetendők az egész alműfajból származó eredményekkel.

A korpusz egészének az elemzése után kiválasztottam a leggyakoribb elemek legmarkánsabb szerepköreit, s a tanulmányban ezek bemutatására szorít-

kozom. A diskurzusjelölők szerepkörei közül először a korpusz élőnyelvi voltából adódó jellegzetességeit, azaz a hezitálást, a szókeresést és az önjavítást mutatom be. Majd a diskurzusjelölőknek a beszélgetés strukturálásában betöltött szerepére térek át, s a mondanivaló elkedzésében, továbbvitelében, a fordulók szervezésében és az egyes szerkezeti részek tagolásában élen járó elemeket ismertetem. Végül pedig a különféle beszélői attitűdöket kifejező diskurzusjelölőket tárgyalom. A diskurzusjelölők multifunkcionalitásából adódik, hogy az egyes szerepkörök nem mindig különíthetők el egymástól. A funkciók meghatározásánál felhasználom a korábbi kutatások (Kiefer, 1988; Németh T., 1998; Schirm, 2007; Alberti & Kleiber, 2014; Gyarmathy, 2015; Dér, 2017) eredményeit is. Megállapításaimat a korpuszból származó példákkal támasztom alá. Kvantitatív vizsgálatot egyelőre nem végeztem az anyagon, az induktív általánosítás módszerével dolgoztam, amely a diskurzuselemzésben teljesen bevettnek számít (l. pl. Schegloff, 2009).

Az egyes felvételeket egy betű- és számsorból álló kóddal láttam el (pl. Vxidx_Dis_001). A telefonbeszélgetések lejegyzése során a Jefferson-féle átírási konvenciót (Heritage, 1984: IX–XVI) követtem. Jelöltem az egymást átfedő és az egyidejű megnyilatkozásokat, a hezitációt (*ö*) és a hosszabb szüneteket. A beszélgetéseket anonimizáltam, vagyis az azonosításra alkalmas részeket töröltem belőlük. A névtelenített információkat a példákban < > jelek közé tett részek (pl. <NÉV>, <SZÁM>) jelzik. A párbeszédekben az ügyintéző fordulóit O (= operátor) vezeti be, az ügyfél forduló előtt pedig Ü áll.

A telefonos ügyfélszolgálati beszélgetések a félintézményes szövegtípusok (vö. Ilie, 1999) közé tartoznak, vagyis vannak bizonyos szabályok, amelyeket a kommunikáció mindkét résztvevőjének, az ügyintézőnek és az ügyfélnek is be kell tartania. E telefonbeszélgetések meghatározott forgatókönyv szerint épülnek föl (vö. Bechet et al., 2012, 1344.). Az első rész a beszélgetés nyitása, amely a köszönést és a bemutatkozást, valamint a hívás céljának a meghatározását tartalmazza. Utána jön az azonosítási szakasz, ahol az ügyintéző az azonosításhoz szükséges adatokat kéri el az ügyféltől. Ezt a szakaszt a konfliktusszituáció követi, itt történik meg a probléma kibontása. Majd – ideális esetben – a követ-

kező részben, a problémamegoldás során rendeződik a fennálló konfliktus, illetve az ügyfél megkapja a szükséges információkat, végül a beszélgetés elköszönéssel zárul. Ez a forgatókönyv és a beszélgetésbeli szerepek (ügyintéző vs. ügyfél) meghatározzák a beszédaktusokat, a fordulók hosszát és a szóátadás módját. A call centeres beszélgetésekben a kérdés-válasz szomszédsági párok dominálnak, ám a kérdező és a megkérdezett szerepek annak függvényében cserélődnek, hogy éppen a forgatókönyv melyik részénél tart a beszélgetés. Például adategyeztetésnél és értékesítési ajánlat közvetítésekor az ügyintéző a kérdező, az ügyfél pedig a válaszoló. Ám információkérésnél megfordulnak a beszédaktusok, s az ügyfél lesz a kérdező, az ügyintéző pedig a válaszoló.

A call centeres hívások egyes részeinek megvannak a jól bevett formulái is. A nyitás az operátor részéről általában a következőképpen hangzik: *Üdvözlöm. [NÉV] vagyok. Miben segíthetek?* A telefonbeszélgetés lezárásának is meghatározott menete van. Az operátor először megbizonyosodik arról, hogy az ügyfél nem szeretne mást intézni, nincs több kérdése (pl. *Még esetleg valamiben tudok segíteni?*), s csak ezután történik meg a zárás, amelyet az ügyfél szokott kezdeményezni (pl. *Hát nagyon szépen köszönöm a kedvességét!*), s erre szomszédsági párként jön a tényleges lezárás (pl. *Én köszönöm a hívást.*) és az elköszönés (*Viszonthallásra.*).

3. Eredmények

A telefonos ügyfélszolgálati beszélgetések szerkezetileg kötött rendszerében a diskurzusjelölőknek többféle szerepük van. Egyrészt a beszéd spontánságából adódóan jelennek meg, másrészt a beszélgetés strukturálásában vesznek részt, harmadrészt pedig beszélői attitűdöket fejeznek ki. A korpusz elemzése azt mutatta, hogy a telefonos ügyfélszolgálati beszélgetésekben mindhárom szerepkörben igen gyakoriak a diskurzusjelölők, ám a beszédhelyzet két szereplője másképp és másra használja ezeket az elemeket.

Mivel az ügyintézőknek rutinjuk van a telefonálásban, a beszélgetés nyitásában, zárásában, valamint az információadásnál előre megírt mondatok is a

rendelkezésükre állnak, ezért a beszélgetésekbeli hezitálás és szókeresés főként az ügyfeleknél figyelhető meg. Például:

(1) a. Ü: *Jó napot kívánok! Én [NÉV] vagyok, és* ***ö hát ilyen*** *átváltással kapcsolatos kérdéseim lennének.* (Vxidx_Dis_003)

b. Ü: *Szerintem beszéltünk mi már a héten,* ***ö*** *van nekem egy problémám,* ***ö hát*** *önöknél van már a levelem a posta szerint*

Az (1a) példában a gondolkodást a *hát* előtt megjelenő *ö*, illetve az utána elhangzó *ilyen* is megerősíti, s az (1b) példában is hezitáló *ö* található a *hát* előtt. A nemlexikális ö hangot a beszélők gyakran azért használják, hogy időt nyerjenek, ám ez az időnyerés a társalgásban sokszor nem egy szó, hanem egy nagyobb szintaktikai egység létrehozására vonatkozik (vö. Németh, 2020). Ez figyelhető meg a fentebbi (1a) példában is, ahol az ügyfél az első megnyilatkozásának az elején, a köszönés és a bemutatkozás után egy diskurzusjelölő-kollokációt (*hát ilyen*) használ, s ugyanígy az (1b)-nél is rögtön a hívás elején, a harmadik tagmondatban jelenik meg a *hát*. Mindkét esetben tisztában vannak az ügyfelek a hívásuk céljával, hiszen ők kezdeményezték a telefonálást, azonban nem tudják, hogyan is kezdjenek hozzá a probléma kibontásához. Ügyintézéskor a betelefonálók sokszor zavarban vannak, nem tudják, hogy pontosan hogyan fogalmazzák meg a mondandójukat, nem ismerik a bevett fordulatokat, a megfelelő szakkifejezéseket, sőt, olyanok is vannak, akik ún. „telefóbiában" szenvednek, azaz idegenkednek a telefontól mint közvetítő közegtől (Hopper, 1992, 5.). Vagyis az ügyfeleknél a beszélgetés legelején megjelenő diskurzusjelölők magából a kommunikációs helyzetből adódnak, s azt fejezik ki, hogy nehézségeik adódnak a mondanivaló megformálásával.

A diskurzusjelölőket a spontán beszéd jellegzetességéből (Gyarmathy, 2015) adódóan önjavításra is használják a call centeres telefonbeszélgetésekben. Ez a szerepkör a hezitálástól és a szókereséstől eltérően főként az operátoroknál volt megfigyelhető a korpuszban, akik a mondanivalójuk újrafogalmazására használ-

ták ezeket az elemeket, vagyis kijavították az előzőleg elmondottakat, ahogy azt az alábbi példa is mutatja:

(2) *Ü: A szerződésen ez rajta van, a számlaszám?*
O: *Nem, nem lesz rajta,* ***úgyhogy ö, hát ö nem is tudom****. Megnézheti, de* ***szerintem*** *nem lesz rajta.* (Vxidx_Dis_001)

Az idézett példában az ügyintéző az *úgyhogy, hát, nem is tudom, szerintem* diskurzusjelölőkkel és a többszöri *ö* elemmel a korábbi kijelentését pontosítja, finomítja. Mivel a telefonbeszélgetéseket minőségbiztosítási okokból rögzítik, ezért az ügyintézők nagyon megfontolják, hogy mit és hogyan mondanak, s megfigyelhető, hogy emiatt az információ továbbításakor diskurzusjelölőkkel korrigálják magukat.

A diskurzusjelölők a telefonos ügyfélszolgálati beszélgetések élőnyelvi jellegéből adódó szerepkörökön túl a beszélgetés strukturálásában is részt vesznek. Azaz a mondanivaló elkezdésére, továbbvitelére, lezárására használják őket, emellett témaváltásra, összegzésre, idézésre, történetmesélésre és példaadásra is használatosak, továbbá általános válaszjelölőként is állhatnak (Schirm, 2017). A telefonhívás nyitásakor gyakori, hogy az ügyfelek a köszönés és a bemutatkozás után a *na, na szóval, na most* elemekkel indítják a társalgást. Például:

(3) Ü: *Jó napot kívánok! Én [NÉV] vagyok.* ***Na, most*** *ö van itt énnekem egy olyan, hogy ügyintéző, lehet, hogy akkor őt kellene kérnem. [NÉV].*
O: *Igen, miben tudok önnek segíteni?*
Ü: ***Na, elmondom, hogy mi a helyzet****. Ö* ***a lényeg az****, hogy a lányoméknak van egy önöknél, tehát ilyen [NÉV]-es valamit megkötöttünk egy szerződést, . . .* (Vxidx_Dis_005)

Az idézett részletből is kitűnik, hogy az ügyfél nehezen indítja a mondanivalóját, bizonytalan abban, hogy kit is keres, s a probléma kibontásának is többször nekifut. Ezt a bizonytalanságot az *ö, egy olyan, hogy ügyintéző, lehet* részek is mutatják. Azonban a bizonytalanság ellenére a használt diskurzusjelölőkkel és panelekkel (*na, most; na, elmondom, hogy mi a helyzet; a lényeg*

az) a betelefonáló saját maga és az operátor számára is strukturálni próbálja a mondandóját. A *na* és a *na most* elemek más szövegtípusokban, például a tanári magyarázatokban is strukturáló szerepet töltenek be (Schirm, 2013): a témát tagolják, s a nagyobb új (al)téma előtt gyakoribbak.

A mondanivaló továbbvitelére a *hát*-ot, a *szóval*-t és a *tehát*-ot használják a legtöbbször, míg lezárásnál a *hát* konklúziószói megjelenése figyelhető meg. A telefonos ügyfélszolgálati beszélgetésekben a kérdés-válasz szomszédsági párok dominálnak, s a *hát* a beszélgetés mindkét szereplőjénél gyakran általános válaszjelölői szerepben áll (Németh T., 1998), ahogy az a lentebbi példákban megfigyelhető, a (4a)-ban az ügyfél kezdi *hát*-tal a válaszát, a (4b)-ben pedig az operátor.

(4) a. O: *Miért, mi a probléma?*
Ü: ***Hát*** *az a gond vele, hogy egyszerűen a volt [NÉV]-os rész egyszerűen nincs meg.* (Vxidx_Dis_007)

b. Ü: *És ezt hogy tudom megtenni?*
O: ***Hát*** *erről tudok önnek csekket indítani.* (Vxidx_Dis_028)

Az információkérő kérdésre adott válaszok elején megjelenő *hát*-nak más a funkciója a call centeres beszélgetésekben, mint a forduló belsejében lévő, *ö* hezitációs elemmel, illetve egyéb diskurzusjelölőkkel álló *hát*-oknak. A kérdésre adott választ indító *hát*-ok egy részéhez evidencialitás is kapcsolódik (Dér, 2017), ám ez a nyilvánvalóság általában csak a beszélő szemszögéből érvényesül. A (4b) alatti *Hát erről tudok önnek csekket indítani* válasz is csupán az operátor számára evidens, az ügyfél éppen azért tette fel a kérdését, mert nem tudja a választ és szüksége van az információra.

A *hát* által jelölt evidencialitás mellett a nyilvánvaló ismeretet a legtöbbször az *ugye* fejezi ki a call centeres párbeszédekben. Ám az *ugyé*-t teljesen másképp használja a beszélgetés két szereplője, az ügyfél és az operátor. A korpuszban a betelefonáló ügyfeleknél háromféle használati köre jelent meg az *ugyé*-nak. Az

egyik, amikor mesélés közben alkalmazták az ügyfelek, s a saját szemszögükből nyilvánvaló dolgot fejezték ki vele. Például:

(5) Ü: *Hát én, amíg betegállományba voltam, nekem ez összesen volt százhatvankét nap, me* ***ugye*** *utána kerültem nyugdíjba.* (Vxidx_Dis_014)

Az operátornak fogalma sem volt arról, hogy az ügyfél mikor került nyugdíjba, a betelefonáló viszont annyira belehelyezkedett a saját nézőpontjába, hogy ezt észre sem vette, s úgy mesélt, mintha az ügyintéző is jól ismerné az ő egész életét.

Az *ugyé*-nak az ügyfelek megnyilatkozásaiban megfigyelhető másik használati köre a bizonytalanság kifejezése volt. A bizonytalanságot kódoló *ugyé*-k kérdő formájúak voltak, s vagy a forduló belsejében (6a), vagy pedig a forduló végén, mintegy utókérdésként jelentek meg (6b).

(6) a. Ü: *Ez a vé, en,* ***ugye*** *ez kell nekünk?* (Vxidx_Dis_004)

b. Ü: *Tehát ez a [NÉV] tizennyolcas lesz,* ***ugye****?* (Vxidx_Dis_004)

Az *ugyé*-k ilyenkor azt jelezték, hogy az ügyfél nem biztos valamiben, s megerősítést vár az ügyintézőtől. Az ilyen kérdések gyakran hangzottak el az ügyfelek részéről az azonosítási szakaszban, amikor valamilyen szerződésszámot, azonosítót, jelszót kellett megadniuk. A bizonytalanság mellett azonban az ügyfelek *ugyé*-i a mondat kijelentő mondati magjának igazsága melletti elkötelezettségét, azaz a betelefonálóknak az általuk mondottakkal azonos polaritású válasz iránti elfogultságát is kifejezték (vö. Alberti & Kleiber, 2014).

Azonban az *ugye* nem mindig az egyszerű bizonytalanságot, s a puszta megerősítés igényét kódolta a korpuszban, a visszajelzés kérése olykor fenyegetésszerű hangnembe csapott át, ahogy az az alábbi példában is megfigyelhető. Az ügyfél ebben a hívásban arra panaszkodott, hogy hiába rendezte a számláját, mégis kapott csekket, s ráadásul ki is kapcsolták nála az internetet.

(7) Ü: *De nekem a januárit is kikapcsolták, pedig be volt fizetve, szóval érti? Azér, én lehet, hogy el is megyek az önök cégétől. De azért nekem nagyon*

precízen mondja meg, hogy hogy február végéig akkor nekem rendezve van, ***ugye****?* (Vxidx_Dis_027)

Az idézetben az ügyfél támadó attitűdjét az *ugye* fenyegető hangsúlyán túl az üzenet tartalma, a *de* ellentétes kötőszók halmozása, s a részlet *szóval érti* diskurzusjelölő-kollokációja is alátámasztja.

Az ügyfelektől eltérően az ügyintézők viszont általában nem kérdésekben használták az *ugyé*-t, hanem magyarázatokban, ami nem meglepő, hiszen ennek a diskurzusjelölőnek a magfunkciója éppen a magyarázat adása (Abuczki, 2014). Ilyen esetekben az *ugye* a nyilvánvalóságot és a bizonyosságot közvetíti. Olyan információk előtt használják az operátorok az *ugyé*-t, amit az ügyfeleknek is kellene tudniuk, mert például benne van az általános szerződési feltételekben (8a), vagy a beszélgetés folyamán korábban már elmondták (8b).

(8) a. O: *Így van, így van, így van. Mer* ***hát ugye*** *a szolgá általános szerződési feltételek szerint a szolgáltatás-hozzáférési pontunk, az a a vo m telefonig van.* (Vxidx_Dis_033)

b. O: *De a decemberi, az, mondtam, hogy az* ***ugye*** *a tizenkettedik hó harminc ö tizenkettedik hó elsejétől tizenkettedik hó harmincegyedikéig vonatkozik.*
Ü: *Igen, igen.*
O: *Azt mondtam önnek.* (Vxidx_Dis_027)

Az *ugyé*-nak ez a fajta használati köre összefügg a mondat aktuális tagolásával: a már ismert információt jelzik vele az operátorok. Azonban ez az ismertség az ő saját tudáskeretükhöz képest értelmeződik, hiszen az ügyfelek éppen híján vannak annak az ismeretnek, amit az ügyintézők a magyarázatukban kifejtenek.

A telefonos ügyfélszolgálati beszélgetésekben a spontán társalgásoknál jóval gyakrabban jelenik meg a *jó* elem, s ennek is többféle használati körét lehet megfigyelni. Egyfelől tagolásra, az egyes szerkezeti részek elválasztására használják a beszélgető felek. Másrészt nyugtázzák is általa a korábban elhangzottakat, s visszajelzést adnak a partnernek. Emellett elégedettséget, pozitív értékelést is

kifejez a szó. Továbbá magyarázatot is zárhat, kérdésként pedig fatikus funkciója van, s nyugtázó választ várnak rá. E sokféle szerepkört jól illusztrálja az alábbi beszélgetés, amelyben egy beszedés nem teljesülése miatt reklamált az ügyfél. Az ügyintéző szerint fedezethiány volt a probléma oka, az ügyfél elmondása alapján azonban volt elegendő összeg a számlán, végül abban maradnak, hogy az ügyfél megpróbálja átutalni az összeget. Ekkor hangzik el a következő részlet:

(9) Ü: ***Jó****, felírtam, és akkor megnézem a számlán.*
O: ***Jó****, tessék megnézni,* ***jó****? És akkor erre átutalja. Érdeklődjön akkor azért utána is, hogy minden rendbe van-e,* ***jó****? Hogy átjött-e,* ***jó****? Azt kérném öntől.* ***Jó****?*
Ü: ***Jó****, rendben.* (Vxidx_Dis_001)

Az idézett példa első *jó*-jával az ügyfél nyugtázza az információt, s erre az operátor azonnali visszajelzést, szintén nyugtázó választ ad. Ám mivel problémás esetről, reklamációról van szó, a teendők újbóli megismétlése után a biztonság kedvéért a további félreértések elkerülése végett az operátor négyszer is felteszi a *jó?* kérdést, amelyekre visszajelzést is vár, s végül az ügyféltől meg is kapja (*jó, rendben*). E sokszoros diskurzusjelölő-használatnak az az oka, hogy az operátornak munkaköri kötelességéből adódóan ellenőriznie kell, hogy az ügyfél számára világosak-e az elhangzottak, s mivel a telefonbeszélgetéseket rögzítik, így egy esetleges későbbi reklamációt megelőzendő még fontosabb, hogy nyoma is legyen annak, hogy az ügyfél megkapta a szükséges tájékoztatást, és meg is értette azt.

A spontán beszéd jellegzetességéből adódó, valamint a beszélgetés strukturálásában résztvevő diskurzusjelölőkön kívül nagy számban vannak olyanok is a call centeres hívásokban, amelyek valamilyen beszélői attitűdöt fejeznek ki. A beszélői attitűd jelzésében a *hát* elem jár az élen (a különféle funkciókról l. Schirm, 2007; Dér, 2017), ám ismét megfigyelhető, hogy más beszélői attitűd kifejezésére és máshogy használják a *hát*-ot az operátorok és az ügyfelek. Az operátorok egyrészt udvariassági stratégiaként alkalmazzák, s olyankor hasz-

nálják, amikor tompítani szeretnék az ügyfélnek mondottakat, főként amikor a betelefonálók számára kellemetlen kötelességet fogalmaznak meg. Ilyen szerepben jelenik meg a *hát* például a *kellene* modális elem előtt, ahogy azt a (10) alatti példa is mutatja, ahol a befizetés szükségességére hívja fel az ügyintéző az ügyfél figyelmét.

(10) Ü: *És akkor most ezt, ha nem fizetem ki, nem lesz ebből gond?*
O: ***Hát ö igazság szerint*** *be kellene fizetni, és ezek az összegek, hogyha túlfizetés lenne a számláján, ezt jóváírják önnek.* (Vxidx_Dis_048)

A (10)-es példában a felszólítás beszédaktusának az élét próbálja meg az ügyintéző a *hát*-tal elvenni. Szintén udvariasságból, de a saját arculatuk fenyegetésének a mérséklésére használják az operátorok akkor a *hát*-ot, amikor az ügyfél reklamációja jogos. Az alábbi példában az ügyfél egy meg nem érkezett csekk miatt reklamál:

(11) O: *Mikor kötötte ön ezt a szerződést?*
Ü: *Ö hát január első hetébe írtam alá a papírokat (....) az <NÉV>-nél, Dunaújvárosba. Akkor lefényképezte a fiatalember, és akkor kérdeztem tőle, hogy ö ugye január volt már, tehát jó, hogy ö a cascómat, azt mikor, hogy tudom fizetni majd önöknél. Azt mondta, hogy ez valami késleltetett fizetés, és hogy majd meg fog érkezni a csekk.*
O: *Igen, igen, még nem látom az éles rendszerbe, egy kis türelmet kérek szépen!* (...) ***Hát*** *még sajnos nem látom előtétben sem. Tehát még nem látom a szerződést.* (Vxidx_Dis_044)

Az ügyintézés szituációjában a cég számára kellemetlen, ha el kell ismerniük, hogy ők hibáztak. A panaszkezelés menetéhez hozzátartozik, hogy csökkenteni kell az arcrombolás mértékét, s ezt az enyhítést az opetáror a *hát*-tal, valamint az utána megjelenő sajnos elemmel fejezi ki. A *hát* és a *sajnos* elem kombinációjával az operátorok a szabadkozásukat és a sajnálatukat tudják az ügyfelek felé kommunikálni, ahogy az az alábbi példában is látható:

(12) Ü: *És, hogyha most ez év ö ez év október elsejéig kérnám az adó, illetve a díjmentesítést. Akkor még mit kapnék vissza belőle?*
O: ***Hát**, ezt így telefonon keresztül sajnos nem tudjuk megmondani így előre.* (Vxidx_Dis_009)

A (12)-es példa egyúttal azt is mutatja, hogy a *hát* erősen multifunkcionális elem, nem feltétlenül egyetlen szerepe van egy adott használatban (vö. Dér, 2017). Az idézett példában a *hát* kérdésre adott válasz elején szerepel, tehát egyrészt általános válaszjelölői szerepben áll, másrészt az utána lévő szünet és a mondanivaló folytatása, valamint a *sajnos* elem is jelzik, hogy az egyszerű válaszjelölésen túl még beszélői attitűdöt, szabadkozást is kódol.

Az ügyintézők a konfliktus kirobbanásakor a diskurzusjelölőket az ellentét tompítására használják. A call centeres beszélgetésekben ugyanis az operátorok részére előírás, hogy ne keveredjenek vitába az ügyfelekkel, s az is a feladatuk, hogy megnyugtassák az elégedetlenkedő betelefonálókat, de emellett a valóságnak megfelelő információval kell szolgálniuk, s finoman azt is jelezniük kell, ha az ügyfeleknek nincs igazuk. A diskurzusjelölők ilyen helyzetekben is segítenek, hiszen finomítják a mondanivalót. Az alábbi részlet egy olyan telefonhívásból származik, ahol az ügyfél amiatt reklamál, hogy a lakásukban lévő egyik telefon működéséhez elosztót kell felszereltetniük, amit pénzért vállalnak csak a szerelők. Az operátor erre finoman az ügyfél tudtára adja, hogy a szerződési feltételek szerint a szolgáltatás csak egy telefonra vonatkozik, s ekkor hangzik el az alábbi részlet:

(13) O: *Az, hogy, hogy önöknek két telefon kell, az **ugye** az nem a **úgymond** a mi hibánk.* (Vxidx_Dis_033)

Az ügyfél kiszállási díjra vonatkozó reklamációját az ügyintéző úgy hárítja el, hogy nyilvánvalóvá teszi, hogy nem a telefontársaság tehet arról, hogy az ügyfélnek két telefonra van szüksége. A *„nem a mi hibánk”* kijelentésben implicit benne van annak sugallása, hogy az ügyfél saját magának köszönheti az extra költséget, ám ezt a sugallt jelentést az operátor az *ugye* és az *úgymond*

diskurzusjelölőkkel igyekszik finomítani, s az utóbbi elemmel próbál kissé el is határolódni a választott kifejezéstől.

A konfliktus mérséklésének egy másik esete az operátorok részéről az, amikor úgy reagálnak az ügyfél támadására, hogy bizonyos diskurzusjelölőkkel (pl. *hát, jó*) látszólag igazat adnak nekik, ám a megnyilatkozás folytatása valójában jelzi a hárításukat. Például mikor az ügyfél amiatt reklamált, hogy rendezte a számláját, mégis kapott csekket, az operátor hiába magyarázta fordulókon keresztül ennek az okát, s próbálta tisztázni a helyzetet, az ügyfél nem értette meg, hanem továbbra is ugyanazt hajtogatta, ahonnan a telefonhívás elején elindultak:

(14) Ü: *De én miért ne miért fizetek én előre? Szóval én nem értem. Mondom, hogy rendeztem.*
O: *Hát jó uram, a számlázást kellene megpróbálni felhívni, miért előre küldték.* (Vxidx_Dis_027)

Az operátor, miután már sokadjára ugyanazokat a köröket futotta végig az ügyféllel, végül ráhagyta (*hát jó uram*) a mondandóját, s más osztályhoz irányította.

Míg az ügyintézők a probléma kirobbanása után a *hát*-okkal igyekeznek mérsékelni a konfliktust, addig az ügyfelek gyakran használják erősítésre és nyomatékosításra ezt az elemet, s a beszélői attitűdök közül legtöbbször felháborodást fejeznek ki általa (pl: *Azt a mindenit! Mer én erről nem tudok, csak azér.* ***Hát*** *ez marha jó!* (Vxidx_Dis_036)). Az idegesség fokozódását a diskurzusjelölők számának szaporodása is jelzi, ahogy az az alábbi beszélgetésben látható. Az ügyfél itt azzal a problémával telefonált, hogy nincs nála internet. Az ügyintéző türelmesen és részletesen elmagyarázta a hibaelhárítás menetét (modem resetelése, vezérlőpult megnyitása, kapcsolódás elindítása), ám az ügyfélnek még így sem sikerült az internetkapcsolatot létrehozni. Ekkor hangzott el az alábbi részlet:

(15) O: *Valamit akkor elnyomott ott.*

Ü: ***Hát*** *ez!* ***Hát*** *ez nem igaz.* ***Hát*** *szóval* ***hát*** *ez borzasztó. Hogy* ***hát*** *másnak simán visszajött meg bejött, itt egész Szerencsen így van, csak (. . .)*

O: *Aki meg tudja, hogy hogy kell resetelni, illetve új kapcsolatot csinálni, az megoldja magának általában.*

Ü: ***Hát akkor*** *most mit csináljak?*

O: *Jó, mondom, annyi az egész ...*

Ü: ***Hát*** *ez meg mi a búbánat? Uram isten.* ***Hát*** *én megőrülök ettől.*

(Vxidx_Dis_022)

Az operátor *Valamit akkor elnyomott ott* megnyilatkozására az ügyfél kifakadt. A fordulójában megjelenő *hát*-ok a csalódottságát és a felháborodottságát fejezik ki. Az operátor a *hát*-okkal teli panaszáradatra és támadásra védekező-en reagál, s elmondja, hogy ezt a problémát az ügyfelek általában saját maguk megoldják. Erre a válaszra az ügyfélnél a kezdeti dühöt a tanácstalanság váltja fel (*Hát akkor most mit csináljak?*). Ám az újbóli telefonos segítségnyújtás sem éri el a hatását, s az ügyfél idézett utolsó fordulójában megjelenő első *hát* (*Hát ez meg mi a búbánat?*) a beszélő csodálkozással teli felháborodását fejezi ki, míg a *Hát én megőrülök ettől* rész *hát*-ja konklúziószói értelemben áll.

Ahogy az idézett részletben is látható, a telefonhívásban a *hát*-ok számának a megnövekedése a minél nagyobb érzelmi bevonódást jelzi. A sok diskurzus-jelölő által kifejezett érzelmi telítettség nemcsak a call centeres beszélgetések sajátossága, hanem általánosabb jellegzetesség lehet, hasonlót találtak ugyan-is strukturált szociolingvisztikai interjúkban a BUSZI II. anyagán (Bartha & Hámori, 2010), valamint a SZÖSZI-ben is (Schirm, 2019). Az érzelmileg telí-tettebb megnyilatkozások az ügyfelek részéről gyakoriak, s főként a telefonos ügyfélszolgálati beszélgetések elején, valamint a végén figyelhetők meg. A nyitó részben az elégedetlenség a gyakoribb, a záró részben pedig sikeres ügyintézés esetén az öröm és az elégedettség érződik ki az ügyfelek fordulóiból, sikertelen

panaszkezelés esetén viszont a düh és a csalódottság uralkodik el az ügyfelek megnyilatkozásaiban.

Sikertelen ügyintézéskor, illetve az ügyfél számára nem kielégítő válasz esetén a felháborodás mellett olykor ironizálni is kezdenek az ügyfelek. A beszélői attitűdöt ilyenkor is jól tükrözik a diskurzusjelölők: megszaporodik a számuk és többféle, a szubjektív viszonyulást kifejezni képes elem (pl. *hát, ja, áhá, szerintem*) is megjelenik a társalgásban. Például a korpusz egyik beszélgetésében egy nem teljesült kötvényátváltás miatt telefonált az ügyfél, s kiderült, hogy nem jó formanyomtatványon küldte be a kérelmét. A hibát az ügyfél szerint a cég munkatársa követte el, a releváns fordulót idézve: *Hát nem én, ne haragudjon, az önök (nevetve) munkatársa. Hát én, mit értek én hozzá?* Ilyen előzmények után hangzott el a következő rész:

(16) Ü: ***Na**, várjon! Most mik vannak itt? Mm semmi jó nincsen most. <NÉV> részvény? <NÉV> részvény, az mit jelent?*
O: *Ö én nem ismerem a részvényeket.*
Ü: *Ja, hogy nem tudja. Áhá, jó. Hát akkor jobb, hogy ha most meg se kérdem, hogy ez az egész szarság hogy áll, mert szerintem jobb, ha nem is tudom (kínosan nevetve), annyit esett azóta. Na, jó. Hát így jártam.* (Vxidx_Dis_003)

Miután a beszélgetés elején kiderült, hogy korábban a cég ügyintézője rontotta el az adminisztrációt, most pedig egy másik operátor nem tudott megfelelő választ adni az ügyfél kérdésére, az ügyfélnek elfogyott a türelme. Diskurzusjelölőkkel (*ja, áhá, jó*) nyugtázta a helyzetet, ám ez a nyugtázás csak látszólagos volt, valójában felháborodott az ügyintéző tudatlansága miatt, s diskurzusjelölőkkel (*hát akkor, szerintem, na, jó, hát így*) mintegy saját magának kommentálta a helyzetet, ám az üzenet valójában a társaság felé közvetített kritika volt. Az elégedetlenséget és a gúnyolódást a diskurzusjelölőkön kívül az ügyfél megváltozott szóhasználata (*ez az egész szarság*) és kínos nevetése is jelzi. A call centeres beszélgetésekben ugyanis a kommunikációs helyzet szerepeiből adódó elvárás, hogy az ügyintéző tudjon érdemi információt adni az ügyfél kér-

désére, illetve igyekezzen megoldani a fennálló problémát. A szerepelvárás nem teljesülése esetén az ügyfelek a fentihez hasonló módokon adnak hangot az elégedetlenségüknek.

Szintén egyszerre több diskurzusjelölőt használnak a betelefonálók akkor, amikor valamilyen számukra kényes témára terelődik a szó. Ilyenkor a diskurzusjelölők eufemisztikus szerepben állnak, finomítják a kimondott szavakat, tompítják a mondanivalót. Például:

(17) Ü: ***Ö igen, mert ö hát ez egy*** *eléggé furcsa szituáció volt. Mert ö itt [SZÁM] harmadik hónapban ö ö megkeresett két hölgy, aki, aki* ***hát, meg kell, hogy mondjam,*** *hogy hogy ö* ***majdhogynem*** *eherőszakkal (zavarában nevetve) beszervezett, de ez lenne a legkevesebb. Sok ilyen erőszakkal ö történt dolgot éltünk má át. De ugyanakkor pedig hetedik hónap akárhányadikán kaptunk egy levelet, vagy hatodik hónap végén, hogy hogy nem tudják ezt a ö szerződést ö ö ö,* ***hogy mondjam csak****, elfogadni, és felbontják, egyoldalúan.* (Vxidx_Dis_045)

Az idézett részlet egy olyan beszélgetésből származik, amelyben az ügyfél amiatt reklamált, mert megszüntették a szerződését. Számára az egész téma kínos, magát a helyzetet is furcsának tartja, amelyet már a történet elmesélésének a diskurzusjelölők általi indítása (*ö igen, mert ö hát ez egy*) és a szóhasználat (*eléggé furcsa szituáció*) is jelez. A történet mesélése közben a legkínosabbnak vélt részek előtt a *hát, meg kell, hogy mondjam, majdhogynem, hogy mondjam csak* diskurzusjelölőket, valamint az *ö* nemlexikális elemet használja tompításra. Az ügyfél érzését a beszédét kísérő kínos nevetés is mutatja, amellyel a zavarát próbálja leplezni. A megfogalmazás eufemisztikus voltához az is hozzájárulhatott, hogy az ügyfél tisztában volt vele, hogy a telefonbeszélgetést rögzítik.

A call centeres telefonbeszélgetések problémamegoldás részébe az ügyintézők által mellékszekvenciaként gyakran valamilyen értékesítési ajánlat is beékelődik. Ezeket az ügyintézők mintegy mellékes, véletlenül eszükbe jutott információként igyekeznek tálalni, pedig a témaváltás igenis megtervezett. A témaváltást az ügyintézők rendszerint az *egyébként* diskurzusjelölővel oldják meg, amely elem

az ajánlat spontánságának a látszatát kelti, azt sugallja, mintha az üzleti ajánlat nem is lenne annyira fontos. Például:

(18) O: *Annyit még engedjen meg, hogy tájékoztassam* ***egyébként*** *önt arról, hogy a telefonszáma, az jogosult lenne egy kedvezményre. (. . .) Szeretné ezt esetleg kipróbálni?*

Ü: ***Háttöömm****, hö, jó, legyen!* ***De hát öm igazából*** *nem használom, otthon netezek a nagy gépen.* ***Tehát*** *én . . .* ***Mhm öm****, köszönöm, nem kérném, mer úgyse látom a telefonon ezeket a dolgokat,* ***úgyhogy*** *. . . ehhez nekem túl apró a kijelző.* (Vxidx_Dis_021)

A beszélgetésbe szőtt ajánlattételeket azonban hárító diskurzusjelölőkkel (*de hát, igazából, tehát*) többnyire vissza is utasítják az ügyfelek, ahogy az a (16)-os példában látható. Az értékesítési ajánlatra reagálva az ügyfelek a diskurzusjelölőket udvariassági stratégiaként használják, hiszen kevésbé bántó valamit tompítva, diskurzusjelölőkkel elhárítani, mint nyíltan visszautasítani.

4. Következtetések

A Voxindex korpusz elemzése azt mutatta, hogy a diskurzusjelölők használata visszatükrözi az ügyfélszolgálati beszélgetések kommunikációs szituációjának a főbb jellemzőit. A félintézményesség és a kötött forgatókönyv elsődlegesen a beszélgetés strukturálásában használt diskurzusjelölőknél érhető tetten. Az ügyfelekre a telefonbeszélgetés nyitásánál a *na, na szóval, na most* elemek jellemzőek, míg a mondanivaló továbbvitelére és lezárására főként a *hát* diskurzusjelölőt használják. Az operátorok az ügyintézés során az egyes szerkezeti részek elválasztásakor gyakran alkalmazzák a *jó* elemet. A beszélgetés rögzítésének a ténye az operátor részéről a minél pontosabb, minél egyértelműbb információátadást és gördülékenyebb problémamegoldást írja elő, s ebben az újrafogalmazás-jelölőknek, udvariassági elemeknek és tompító diskurzusjelölőknek (pl. *hát, szerintem, úgymond*) van fontos szerepük. A beszélgetés rögzítése és a félintézményes szituáció azonban az ügyfeleket gyakran feszélyezi, náluk ez

a szókereső és megakadásjelenségként álló, valamint a tompító elemekben (pl. *ö, hát, ilyen*) mutatkozik meg. A kérdés-válasz szomszédsági párok miatt az általános válaszjelölőként használt diskurzusjelölők mindkét félnél gyakoriak, s az általános válaszjelölőként álló *hát*-ok egy részéhez evidencialitás is kapcsolódik. Szintén mindkét szereplőnél gyakori az *ugye* diskurzusjelölő. Az *ugye* az ügyfeleknél a probléma kibontásakor jelölheti a beszélő szemszögéből fennálló nyilvánvalóságot, de kifejezhet bizonytalanságot, és fenyegetést is kódolhat, míg az operátoroknál főként magyarázatban fordul elő, s a már ismert, tudott dolgot jelöli. Szerepspecifikus, s az operátorhoz kötődik a nyugtázó *jó?*, valamint az udvariassági elemként és a tompító jelölőként használt *hát*. Az operátor beszélgetési stílusánál előírás a tárgyilagosság, s neki a megfelelő információnyújtás mellett az ügyfelet is meg kell próbálnia lenyugtatnia. Ugyanis a betelefonáló fél legtöbbször eleve negatív attitűddel indít a beszélgetéskor, s az erőteljes emocionalitást, a felháborodást, a dühöt és a gúnyolódást az általa használt diskurzusjelölők is érzékeltetik. A telefonbeszélgetésekben az egyre nagyobb érzelmi bevonódást – főként az ügyfelek részéről – a diskurzusjelölők számának a növekedése is jelzi.

A kutatást érdemes volna még nagyobb mintán folytatni, s megnézni, hogy vannak-e diskurzusjelölőkhöz kötődő vagy azoktól független, eltérő diskurzusstratégiák a különböző életkorú és nemű operátorok, valamint ügyfelek közt.

Köszönetnyilvánítás

A tanulmány a Bolyai János kutatási ösztöndíj támogatásával készült.

A kutatás a GINOP-2.2.1-15-2017-00071 „Voxindex" - Magyar, call centerek számára készült hangbányászati rendszer nemzetköziesítése intonált kifejezések azonosítására alkalmas technológia kifejlesztésével projekt keretén belül valósult meg.

Hivatkozások

Abuczki, Á. (2014). *A Core/Periphery Approach to the Functional Spectrum of Discourse Markers in Multimodal Context. PhD-értekezés. Kézirat*. Debrecen: Debreceni Egyetem BTK Nyelvtudományok Doktori Iskola. URL: https://dea.lib.unideb.hu/dea/bitstream/handle/2437/210101/disszertacio_abuczki_agnes_t.pdf?sequence=13&isAllowed=y.

Alberti, G., & Kleiber, J. (2014). ReALIS. Discourse representation with a radically new ontology. In L. Veselovská, & M. Janebová (Eds.), *Complex Visibles Out There. Proceedings of the Olomouc Linguistics Colloquium 2014: Language Use and Linguistic Structure* (p. 513–528). Olomouc: Palacký University.

Archer, D., & Jagodziński, P. (2014). Call centre interaction: A case of sanctioned face attack? *Journal of Pragmatics*, *76*, 46–66.

Bartha, C., & Hámori, A. (2010). Stílus a szociolingvisztikában, stílus a diskurzusban. Nyelvi variabilitás és társas jelentések konstruálása a szociolingvisztika „harmadik hullámában". *Magyar Nyelvőr*, *134*, 298–321.

Bechet, F., Maza, B., Bigouroux, N., Bazillon, T., El-Bèze, M., De Mori, R., & Arbillot, E. (2012). DECODA: a call-centre human-human spoken conversation corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation. European Language Resources Association* (p. 1343–1347).

Dér, Cs. I. (2017). A *hát* multifunkcionalitása a beszédműfajok és a diskurzusjelölőtársulások függvényében. *Beszédkutatás*, *25*, 169–184.

Gyarmathy, D. (2015). Diszharmóniás jelenségek, megakadások a beszédben. In M. Gósy (Ed.), *Diszharmóniás jelenségek a beszédben* (p. 9–49). Budapest: MTA Nyelvtudományi Intézet.

Hámori, Á. (2006). A társalgási műfajokról. In G. Tolcsvai Nagy (Ed.), *Szöveg és típus* (p. 157–181). Budapest: Tinta Könyvkiadó.

Hopper, R. (1992). *Telephone conversation*. Bloomington–Indianapolis: Indiana University Press.

Ilie, C. (1999). Question-response argumentation in talk shows. *Journal of Pragmatics*, *31*, 975–999.

Jucker, A. (1993). The discourse marker *well*: A relevance-theoretical account. *Journal of Pragmatics*, *19*, 435–452.

Kiefer, F. (1988). Modal particles as discourse markers in questions. *Acta Linguistica Hungarica*, *38*, 107–125.

Kopparapu, S. (2015). *Non-Linguistic Analysis of Call Center Conversations*. Springer.

Lenk, U. (1998). Discourse markers and global coherence in conversation. *Journal of Pragmatics*, *30*, 245–257.

Luke, K. K., & Pavlidou, T. S. (Eds.) (2002). *Telephone Calls, Unity and diversity in conversational structure across language and cultures*. Amsterdam: John Benjamins.

Németh, Z. (2020). A nemlexikális *öö* hang interakciós szerepének elemzése magyar nyelvű társalgásokban. *Jelentés és Nyelvhasználat*, *7*, 23–50.

Németh T., E. (1998). A *hát*, *így*, *tehát*, *mert* kötőszók pragmatikai funkciójának vizsgálata. *Magyar Nyelv*, *94*, 324–331.

Orthaber, S., & Márquez-Reiter, R. (2015). 'thanks for nothing': Impoliteness in service calls. In R. Şükriye, & Y. Aksan (Eds.), *Exploring (im)politeness in specialized and general corpora, Converging methodologies and analytic procedures* (p. 11–39). Newcastle upon Tyne: Cambridge Scholars Publishing.

Pallotta, V., & Delmonte, R. (2013). Interaction mining: the new frontier of customer interaction analytics. In C. Lai, G. Semeraro, & E. Vargiu (Eds.), *New challenges in distributed information filtering and retrieval* (p. 91–111). Berlin–Heidelberg: Springer.

Schegloff, E. A. (1968). Sequencing in conversational openings. *Kommunikáció II*, *70*, 1075–1095.

Schegloff, E. A. (2009). One perspective on Conversation Analysis: Comparative Perspectives. In J. Sidnell (Ed.), *Conversation Analysis: Comparative Perspectives* (p. 357–406). Cambridge: Cambridge University Press.

Schegloff, E. A., & Sacks, H. (1973). Opening up Closings. *Semiotica*, *7*, 289–327.

Schirm, A. (2007). A *hát* diskurzusjelölő története. *Nyelvtudomány*, *III–IV*, 185–201.

Schirm, A. (2013). A diskurzusjelölők a tanári magyarázatokban. In E. Szöllősy, L. Prax, & A. Hoss (Eds.), *Találkozások az anyanyelvi nevelésben* (p. 259–268). Pécs: PTE Nyelvtudományi Doktori Iskola.

Schirm, A. (2017). A diskurzusjelölők és a szövegtípusok viszonyáról. *Magyar Nyelv*, *113*, 330–341.

Schirm, A. (2019). A diskurzusjelölő-használat életkori sajátosságai a nyelvi interjú szövegtípusában. *Beszédkutatás*, *27*, 187–205.

Vicsi, K., & Sztahó, D. (2009). Ügyfél érzelmi állapotának detektálása telefonos ügyfélszolgálati dialógusban. In A. Tanács, D. Szauter, & V. Veronika (Eds.), *VI. Magyar Számítógépes Nyelvészeti Konferencia* (p. 217–225). Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport.

Discourse markers and connectives in interpreted Hungarian discourse: A corpus-based investigation of discourse properties and their interdependence

Andrea Götz[1]

[1] *Károli Gáspár University*

Abstract

This paper investigates the frequency of discourse markers and connectives in interpreted Hungarian discourse. Even though the relationship between these items and other linguistic phenomena, such as hesition, is well known, no study to date has set out to explore it in relation to Hungarian interpreted discourse. This study examines the link between the frequency of these items, delivery speed, and filled pauses in a corpus of European Parliamentary speeches interpreted from English to Hungarian. According to the results, discourse markers and connectives are more frequent in interpreted than original discourse, and their frequency positively and significantly correlates with delivery speed, while filled pauses do not show such a straightforward relationship.

1. Introduction

Discourse markers and connective items (DMCs) have been essential to the investigation of translated discourse due to their role in discourse cohesion. They have also been at the centre of attention in the rapidly developing field of corpus-based interpreting studies. Corpus-based interpreting studies has experienced a growth spurt in recent years, going from a "cottage industry" to a "booming research field" (Bendazzoli, 2018; Bendazzoli et al., 2018), which resulted in numerous studies on a wide variety of topics. A number of findings have emerged with regard to DMCs in simultaneous interpreting:

- DMC frequency increases in interpreting relative to source speeches (Defrancq et al., 2015),

Email address: gotz.andrea@kre.hu (Andrea Götz)

Beszédtudomány – Speech Science 2020. 259–284. DOI: 10.15775/Besztud.2020.259-284

- DMC frequency is higher in interpreted than in translated texts (Defrancq et al., 2015),
- DMC frequency can be higher in interpreted than in non-interpreted speeches (Defrancq, 2018),
- DMCs can be used differently in interpreted than in non-interpreted texts (Defrancq, 2016).

To explain these results, research mostly points to the unique conditions (e.g. time pressure) under which simultaneous interpreters work that inevitably shape interpreted discourse output. DMC frequency differences between interpreted and non-interpreted discourse are understood to come about because interpreters "drastically reshape the discourse structure of the source text" (Defrancq et al., 2015). However, as the properties of spoken and interpreted discourse alike are the result of the complex interplay of many factors, it stands to reason that DMCs should be studied in relation to other discourse factors too.

In fact, there is a growing trend in the study of interpreted discourse to investigate the interrelationship of discourse properties (e.g. Collard & Defrancq, 2017; Defrancq & Plevoets, 2018; Plevoets & Defrancq, 2018; Collard & Defrancq, 2019, 2020).

Filled pauses, for example, are known to respond to several factors in interpreted discourse. Filled pause frequency of interpreters increases with factors boosting information load, such as higher delivery speed (Plevoets & Defrancq, 2016), and decreases with factors lowering cognitive load, such as formulaicity, which seems to "free up" cognitive bandwidth (Plevoets & Defrancq, 2018).

As interpreted discourse has been found to contain more hesitation than original discourse in a number of languages (e.g. Plevoets & Defrancq, 2016; Götz, 2018, 2019b; Collard & Defrancq, 2020), and given that DMCs are interlinked with hesitation in structuring spoken discourse (Crible, 2018), forming functionally distinct patterns (Crible, 2017) that perform specific discourse functions (e.g. transition, giving floor, etc.) (Crible et al., 2017), it would be

all the more important to understand how filled pauses and DMCs interact in interpreted speech.

Delivery speeds have also been linked to DMCs (Götz, 2019a; Magnifico & Defrancq, 2020). In an exhaustive study of multiple factors (e.g. target language, the speaker's and the interpreter's gender), only delivery speed was found to have a statistically significant impact on DMC frequency (Magnifico & Defrancq, 2020). This result is somewhat surprising since a number of "striking gender differences" have been pinpointed in interpreted discourse (Magnifico & Defrancq, 2020, 6.) that vary with the particular target language.

Gender seems to influence, for example, politeness (Magnifico & Defrancq, 2016) and hedge use (Magnifico & Defrancq, 2017). On the other hand, target language also remains profoundly influential even when gender differences emerge (Magnifico & Defrancq, 2017). Nevertheless, gender has been also clearly ruled out as a decisive factor in an extensive study on hesitation in interpreted speech, finding only limited gender differences (Collard & Defrancq, 2020).

On the whole, however, men are usually observed to hesitate more (Collard & Defrancq, 2017; Götz, 2018), DMC frequency is variably higher among female (Götz, 2019a) or male interpreters (Magnifico & Defrancq, 2020), with delivery speed showing similar variation (cf. Russo, 2018; Götz, 2019a).

With all this complexity, one aspect of interpreted discourse has so far escaped attention: individual differences. It is well understood in Hungarian discourse marker research that age and gender can influence discourse marker choice (Markó & Dér, 2011; Vukov Raffai, 2016; Schirm, 2019), but it is also clear that individual-specific patterns exist in DMC use (e.g. Dér & Markó, 2010; Vukov Raffai, 2016; Schirm, 2019), as well as hesitation marker use (Horváth, 2014), and these can override group-level tendencies. This means that both groups and individuals need to be examined.

This study represents a preliminary investigation into DMC frequency, delivery speed, and filled pauses, including the interrelationships of these properties in English to Hungarian simultaneously interpreted discourse. In addition, the properties of individual discourse output is contrasted with that of groups.

The remainder of the paper is structured as follows. First, the paper presents a brief overview of DMCs in interpreted discourse, proceeding with the data and methods of this study, which is followed by the results and the conclusion. According to the results, DMCs are more frequent in interpreted than in original discourse, and while DMC frequency correlates positively with delivery speed, filled pauses and DMC frequency do not show such a clear relationship.

2. Discourse markers and connectives in corpus-based interpreting studies

A number of misconceptions persist about the role of discourse markers and connectives in interpreted discourse. As DMCs do not contribute to propositional meaning, they are often seen as non-essential, and thus "vulnerable in the interpretation process" (Defrancq et al., 2015, 198.).

Empirical research, however, disproved this received wisdom, finding a high frequency of these items in interpreted discourse (Defrancq et al., 2015), indicating that they play a profound role in re-creating cohesion in interpreting. But beyond linking segments of discourse, DMCs have various role, and therefore can cause considerable problems for interpreters in the "pragmatic aspects of discourse" (Hale, 1999, 57.). DMCs, for example, can influence how speakers are perceived and judged, which can have far-reaching consequences in a legal or political context.

As a rule, discourse markers in interpreted speech are attributed to the speaker and not to the interpreter (Blakemore & Gallai, 2014). Since DMCs in utterance comprehension serve as clues to the cognitive processes of speakers, guiding the interpretation processes of the hearer (Blakemore, 2002), in interpreting, they are perceived to reflect thought processes of the speaker, and not those of the interpreter (Blakemore & Gallai, 2014).

This means that in case DMCs interpreters use are stigmatized, these items can damage the image of the speaker, even if the particular DMCs do not originate from the speaker. Accordingly, "polished" versions of interpreted tes-

timonies, meaning that they had been edited to remove hedges and discourse markers, are evaluated as significantly more competent, credible, and intelligent than unedited versions which contain discourse markers (Hale, 2010).

Another important aspect of DMC use is the particular institutional context in which the discourse itself is created. Most interpreting corpora and studies on EU languages use European Parliamentary (EP) speeches (e.g. European Parliament Translation and Interpreting Corpus (EPTIC) in Bernardini et al. (2016); European Parliament Interpreting Corpus Ghent (EPICG) in Plevoets & Defrancq (2018)). Since EP data looms so large, at least in European corpus-based interpreting studies, it is important to consider the impact of the EP's institutional context, as well as the limitations of these data sets.

The discourse of EP interpreters could converge in some aspects. Case in point, on the basis of DMC frequencies (*well, now, so*) in French, Spanish, and Italian to English interpreting, EP interpreters could be described as forming a discourse community, while EP interpreters and MEPs together could not (Defrancq, 2018). Such patterns could be caused by interpreters adhering to certain institutional norms of discourse (cf. Magnifico & Defrancq, 2020).

In summary, the frequency of DMCs has been observed to increase in interpreting compared to source texts (Defrancq et al., 2015), some DMCs have also been found to be more frequent in interpreted than in non-interpreted, original discourse (Defrancq, 2018), and while gender differences can appear, these are statistically not significant in the studies so far available (cf. Magnifico & Defrancq, 2020). Their frequency, however, positively and significantly correlates with delivery speed (Magnifico & Defrancq, 2020).

3. Research design

3.1. Research goals and hypotheses

The purpose of this study is to examine the frequency of DMCs in interpreted Hungarian discourse and test whether this frequency is related to other factors,

such as delivery speeds or the frequency and duration of filled pauses. The study consequently investigates the following hypotheses:

1. DMCs are more frequent in interpreted than in original Hungarian discourse.
2. DMCs are more frequent in Hungarian target speeches than in their English source speeches.
3. Interpreted discourse contains more filled pauses than non-interpreted.
4. DMC frequency is positively correlated with delivery speed.
5. Both filled pause duration negatively correlated with DMC frequency.
6. The discourse output of female and male interpreters shows significant differences.

3.2. Corpora

This study uses three corpora: the Interpreted Hungarian Corpus (IHC), the Original Hungarian Corpus (OHC), and corpus of English source speeches (ESC). The texts of these corpora are sourced from the Hungarian European Parliamentary intermodal corpus (HEPIC) (Götz, 2017). The HEPIC is composed of EP speeches delivered between 2008 and 2012 on plenary sitting days, presently comprising about 230,000 words.

Table 1 shows the properties of the corpora, while Table 2 presents them broken down according to the sex of the speakers. Both corpora contain over one hour's worth of speeches produced by five female and five male speakers. Table 3 displays the properties of the discourse of the individual speakers. For the full data set of the corpora see Appendix 2.

3.3. Methods

This study investigates the frequency of discourse markers (raw frequency, normalized frequency per minute), delivery speeds (number of words per minute), and the frequency and duration of filled pauses. Statistical significance is probed using t-tests, correlation by Pearson's correlation coefficient, sex differences are

Table 1: Properties of the interpreted and the original corpora

	IHC	OHC
Speech time	1 hour 18 min 40 sec	1 hour 15 min 56 sec
English speech time	1 hour 16 min 58 sec	
No. of speeches	50	40
No. of words	8064	9174
No. of English words	12,183	
No. of speaker	5 female, 5 male	5 female, 5 male

Table 2: Properties of the interpreted and original corpora by the sex of the speakers

	IHC	OHC
Speech time (min) (f)	41 min 30 sec	36 min 27 sec
Speech time (min) (m)	37 min 10 sec	39 min 29 sec
No. of speeches (f-m)	27-23	20-20
No. of words (f)	4240	4469
No. of words (m)	3824	4705

explored with the Mann-Whitney test, and outliers are identified with Grubb's test.

The data of this study are derived from two sources: individual speeches, and the discourse output of individual interpreters. The former is utilized in the descriptive analysis of DMC frequency, rate of delivery speeds, and the frequency and length of filled pauses, while the two are combined to probe correlations.

The frequency DMCs set is examined which contain both traditional conjunctions and discourse markers. These sets are based on the most frequent items and are matched between Hungarian and English. Both DMC sets can found in Appendix 1.

Filled pauses are defined here as vocalisations not contributing to the propositional. Although lengthening vowels and consonants could be classified as filled pauses, this study opts for a more restricted definition by only including schwa-

Table 3: Properties of the interpreted and original by individual speakers

IHC	Speech time	No. of speeches	No. of words
Female interpreter #1	4 min 24 sec	3	488
Female interpreter #2	1 min 52 sec	2	192
Female interpreter #3	6 min 6 sec	5	703
Female interpreter #4	13 min 16 sec	7	1304
Female interpreter #5	15 min 49 sec	10	1553
Male interpreter #1	3 min 44 sec	3	427
Male interpreter #2	20 min 47 sec	13	2083
Male interpreter #3	2 min 9 sec	2	216
Male interpreter #4	6 min 55 sec	3	740
Male interpreter #5	3 min 32 sec	2	358
OHC	**Speech time**	**No. of speeches**	**No. of words**
Female speaker #1	6 min 50 sec	5	879
Female speaker #2	7 min 34 sec	4	899
Female speaker #3	10 min 42 sec	5	1150
Female speaker #4	6 min 13 sec	3	685
Female speaker #5	5 min 6 sec	3	856
Male speaker #1	8 min 7 sec	4	904
Male speaker #2	6 min 14 sec	3	776
Male speaker #3	13 min 2 sec	7	1612
Male speaker #4	5 min 40 sec	3	720
Male speaker #5	6 min 24 sec	3	693

like neutral vowel resembling [ø] or [ə] and its combinations with [m], and [ɒ] type filled pauses. For the annotation of filled pauses Exmaralda Partitur Editor was used.

4. Results and discussion

4.1. Frequency of discourse markers and connective items

Table 4 shows the raw and normalized frequency of DMCs, breaking the data down by the sex of the speakers as well. In the case of the English source corpus (ESC), "female" and "male" refer to the input that Hungarian female and male interpreters were exposed to, and not the sex of the English speakers.

Table 4: The frequency of DMCs

DMCs	IHC	OHC	ESC
Raw frequency	1355	1069	1093
Raw frequency (f)	735	545	573
Raw frequency (m)	620	524	520
Frequency /1 minute	17.2	14.1	14.2
Frequency /1 minute (f)	17.7	15.0	14.1
Frequency /1 minute (m)	16.7	13.3	14.3

Frequency of DMCs is significantly higher in the interpreted Hungarian speeches than in their English sources speeches ($t = -2.379$, $p = 0.01$), or in non-interpreted Hungarian speeches ($t = 3.122$, $p = 0.001$).

Higher frequency in comparison to source speeches can, of course, be influenced by cross-linguistic tendencies, therefore it should not be interpreted as an increase purely due to the effect of interpreting. However, the fact that DMC frequencies are similar across the original Hungarian and English source speeches, but higher in interpreted discourse, might indicate that interpreted discourse makes use of these items differently.

Although female speakers in interpreted and original Hungarian discourse use more DMCs than men, this variation is not statistically significant (interpreters: $z = -0.954$, $p = 0.342$, original speakers: $z = 0.122$, $p = 0.904$).

4.2. Delivery speed

Table 5 shows the delivery speeds of Hungarian and English original speakers and interpreters. Interpreted and original Hungarian delivery rates differ significantly overall (t = -4.793, p =< 0.000). From the three groups, interpreters have the lowest delivery speeds. Of course, simultaneous interpreting cannot be expected to produce similar delivery speeds to original speakers who read out loud their speeches.

Table 5: Delivery speeds

	IHC	**OHC**	**ESC**
Words/min	102.5	120.8	158.3
Words/min (f)	102.2	122.6	156.9
Words/min (m)	102.9	119.2	159.9

Interpreters produce approximately 18 words fewer per minute than original speakers. Female and male interpreters show very similar delivery speeds with no statistically significant variation ($z = 0.058$, $p = 0.476$). Female interpreters were exposed to an overall slightly slower English delivery speed than male interpreters.

4.3. Filled pause duration

Table 6 shows the normalised duration (seconds per minute) of filled pauses in interpreted and original Hungarian discourse.

The normalised duration of filled pauses is significantly longer in the IHC than in the OHC ($t = 8.603$, p =< 0.000). The duration of filled pauses differs significantly between female and male interpreters ($z = -2.141$, $p = 0.032$).

4.4. Filled pause frequency

Table 7 shows the absolute and normalized frequency of filled pauses in the corpora.

The normalised frequency of filled pauses is significantly higher in the interpreted corpus (t = 9.233, p =< 0.000) than in the OHC. Unlike duration,

Table 6: The duration of filled pauses

	IHC	OHC
Absolute duration of filled pauses (sec)	231.4	30.2
Absolute duration of filled pauses (sec) (f)	110.4	10.1
Absolute duration of filled pauses (sec) (m)	121.0	20.1
Normalised duration of filled pauses (sec)	2.9	0.4
Normalised duration of filled pauses (sec) (f)	2.7	0.3
Normalised duration of filled pauses (sec) (m)	3.3	0.5

Table 7: Frequency of filled pauses

	IHC	OHC
Absolute frequency of filled pauses	723	99
Absolute frequency of filled pauses (f)	377	36
Absolute frequency of filled pauses (m)	346	63
Normalised frequency of filled pauses	9.2	1.3
Normalised frequency of filled pauses (f)	9.1	1.0
Normalised frequency of filled pauses (m)	9.3	1.6

frequency is slightly higher for female interpreters than for males, though not significantly ($z = -0.954$, $p = 0.342$).

4.5. Correlation

4.5.1. DMC frequency and delivery speed

Figure 1 presents the correlation between DMC frequency and delivery speed in each speech interpreted by female and male interpreters. DMC frequency and delivery speed show a statistically significant, positive moderate correlation. This suggests that there is a tendency for higher delivery speeds to correlate with higher DMC frequencies. Table 8 presents the results of the correlation tests.

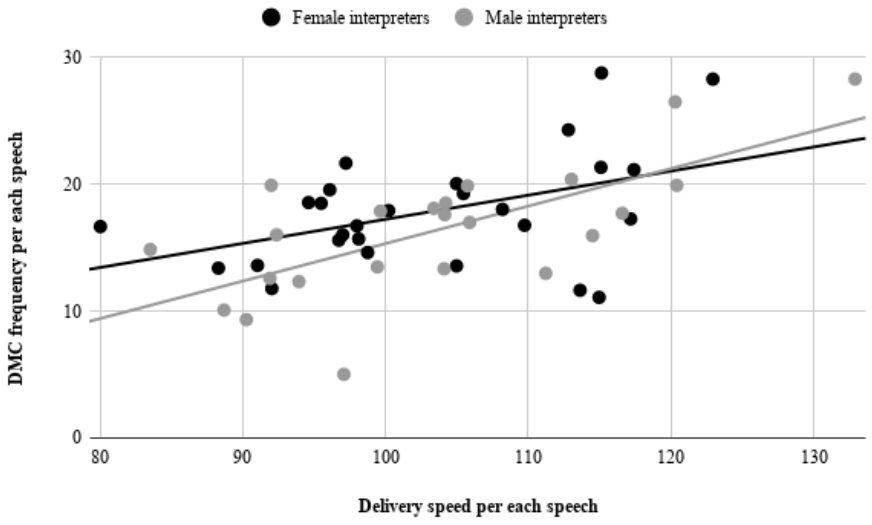


Figure 1: The correlation between delivery speed and DMC frequency in each speech interpreted by female (black) and male (grey) interpreters

Table 8: Pearson's correlation test of DMC frequency and delivery speed

	r	**p**
IHC	0.6	0.0
OHC	0.6	0.0
ESC	0.6	0.0

However, since all corpora show this relationship with very similar, statistically significant r values, this tendency does not exclusively characterize interpreted speech but rather represents a more universal tendency.

For female interpreters, the relationship is weaker ($r = 0.448$, $p = 0.019$) but significant. Among male interpreters, the connection is stronger ($r = 0.691$, $p = 0.000$) and similarly significant.

Since DMCs in English source speeches could have an effect on DMC frequency in interpreted Hungarian discourse, it is necessary to provide a compar-

ison. Figure 2 demonstrates the correlation between the frequency of English DMCs in the source speeches and their corresponding Hungarian values.

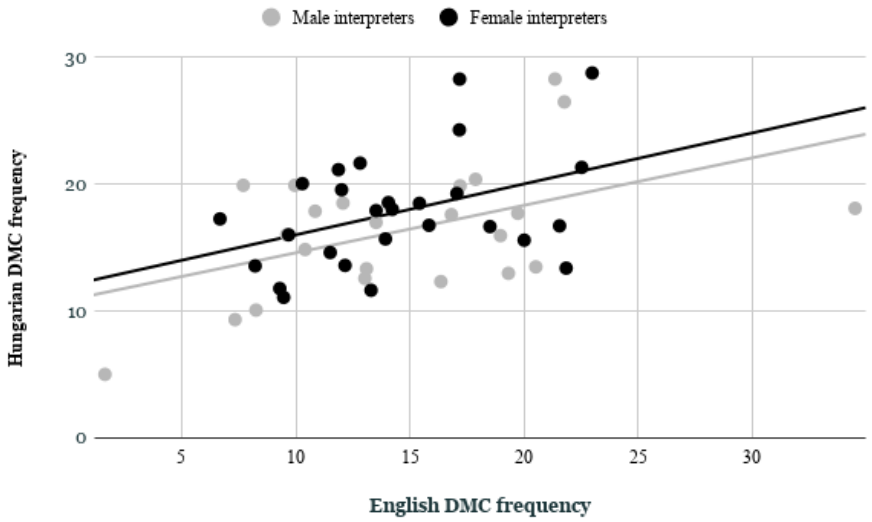


Figure 2: Correlation between English DMC frequency and Hungarian DMC frequency in each speech interpreted by female (black) and male (grey) interpreters

English and Hungarian DMC frequency correlate positively, and though the relationship is weak, it is statistically significant ($r = 0.444$, $p = 0.001$). This relationship is weaker among women ($r = 0.407$, $p = 0.035$) than men ($r = 0.491$, $p = 0.017$) but it is significant in both groups.

Figure 3 shows this correlation in the discourse output of each individual interpreter calculated from their total interpreting output.

When it comes to individual interpreters, there is a positive weak, statistically not significant correlation between DMC frequency and delivery speed ($r = 0.085$, $p = 0.815$), which is stronger for women ($r = 0.087$, $p = 0.889$) and weaker for men ($r = 0.053$, $p = 0.933$). Grubb's test detected no significant outliers in the group, either in terms of delivery speed or DMC frequency.

This might indicate that while there is an overall tendency for higher delivery speeds to correlate with higher DMC frequencies, there is variation and

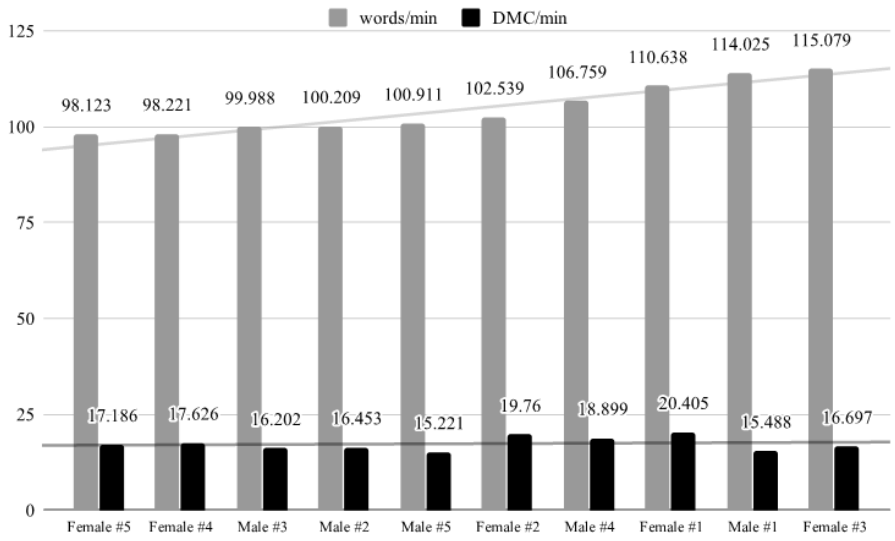


Figure 3: DMC frequency (black) and delivery speed (grey) in the discourse of individual interpreters

divergence from this tendency within the interpreting output of individual interpreters as the performance of interpreters can vary from occasion to occasion.

4.5.2. DMC frequency and filled pause duration

Figure 4 demonstrates the relationship between normalized filled pause duration and DMC frequency. DMC frequency and normalized filled pause duration show a weak, negative, statistically not significant correlation. Table 9 shows the results of the correlation tests.

Table 9: Pearson's correlation of DMC frequency and filled pause duration

	r	**p**
IHC	-0.2	0.2
OHC	0.1	0.6

Filled pause duration and DMC frequency do not correlate significantly or strongly either for either group, but the correlation is positive among women

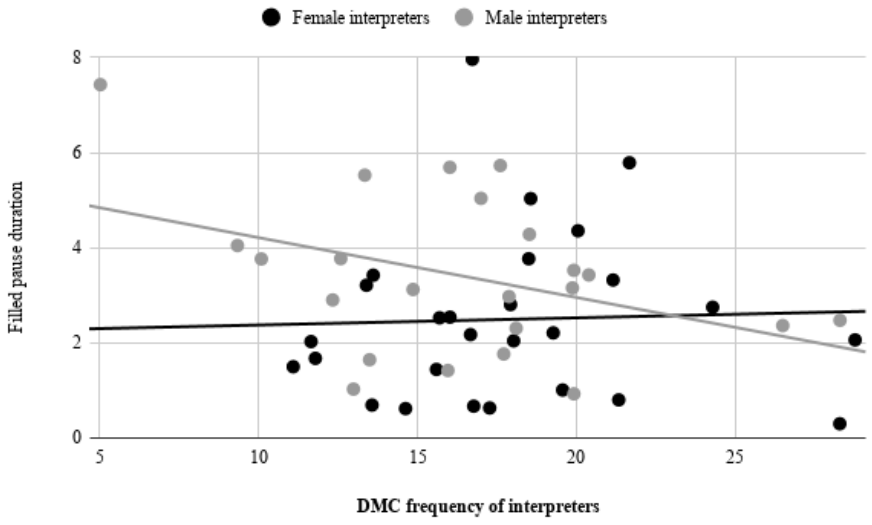


Figure 4: The correlation between DMC frequency and normalized filled pause duration in each speech interpreted by female (black) and male (grey) interpreters

($r = 0.038$, $p = 0.850$) and negative among men ($r = -0.393$, $p = 0.863$). Figure 5 shows the correlation of DMC frequency and normalized filled pause duration in the discourse of individual interpreters.

Looking at the discourse production of individual interpreters, the correlation between DMC frequency and pause duration is weak and positive ($r = 0.134$, $p = 0.712$), both for women ($r = 0.182$, $p = 0.770$) and men ($r = 0.380$, $p = 0.528$). No significant outliers were detected.

4.5.3. DMC frequency and filled pause frequency

Figure 6 shows the correlation between the normalized frequency of DMCs and filled pauses. Filled pause frequency and DMC frequency correlate negatively and weakly, not forming a statistically significant relationship. Table 10 presents the results correlation tests.

For female interpreters, the relationship is positive and weak ($r = 0.099$, $p = 0.623$), while for men, the correlation is negative and weak ($r = -0.227$,

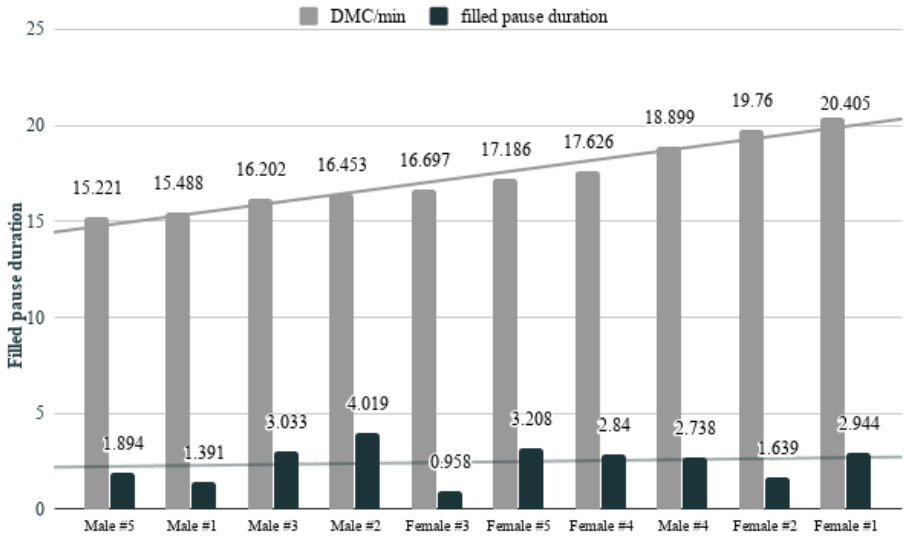


Figure 5: DMC frequency (grey) and normalized filled pause duration (black) in the discourse of individual interpreters

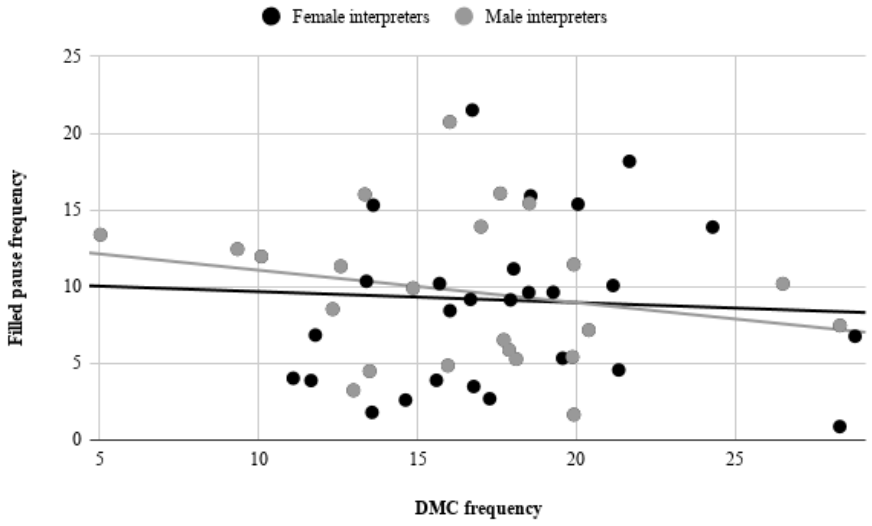


Figure 6: The correlation between DMC frequency and normalized filled pause frequency in each speech interpreted by female (black) and male (grey) interpreters

Table 10: Pearson's correlation of DMC frequency and filled pause duration

	r	p
IHC	-0.1	0.6
OHC	0.1	0.5

$p = 0.298$). The contrast between these trends underlines the need for an individual investigation.

Figure 7 presents the correlation between DMC frequency and filled pause frequency in the discourse of the individual interpreters. Pause frequency correlates positively with DMC frequency ($r = 0.214$, $p = 0.553$), more strongly for women ($r = 0.233$, $p = 0.706$) than men ($r = 0.061$, $p = 0.922$). Despite the noticeable variation in filled pause frequency, Grubb's test did not identify significant outliers.

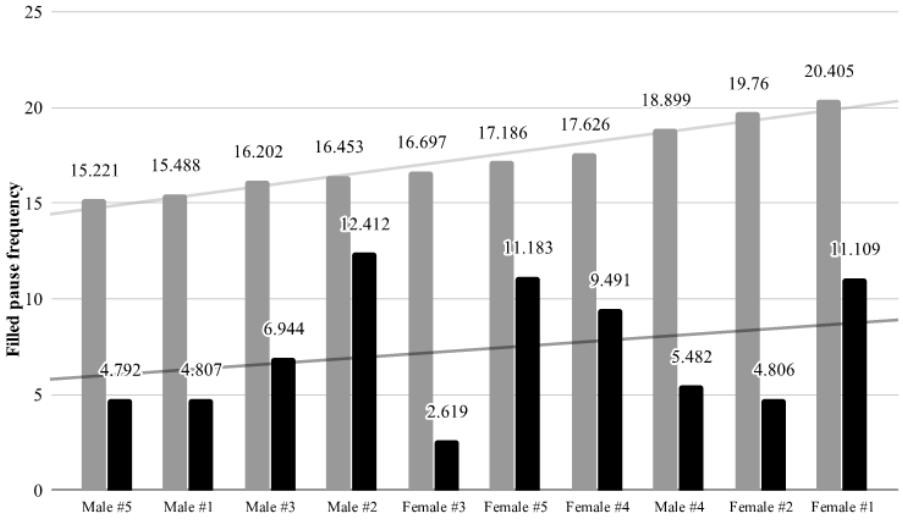


Figure 7: DMC frequency (grey) and normalized filled pause frequency (black) in the discourse of individual interpreters

Filled pause frequency, as opposed to duration, seems to show greater variation among individuals. The difference between the group-based and individual

results therefore could stem from the individual speech styles and varying performance of the particular interpreters. However, since they are unevenly represented in the corpus, these results can only be taken as preliminary, outlining directions for future research.

5. Conclusion

This study investigated the frequency of DMCs in interpreted Hungarian EP speeches, probing the relationship between DMC frequency, delivery speed, filled pause duration and frequency.

First, it examined whether DMCs are more frequent in interpreted than in original discourse. Since DMC frequency is highest in Hungarian interpreted discourse compared to both English source speeches and Hungarian original speeches, hypotheses 1 and 2 are both confirmed.

Despite these positive findings, a number of important caveats must be pointed out. It should be stressed that language mediation could have a range of effects. Interpreters can add items but can also omit shorter and longer sections from the original speeches. It is therefore not obvious whether higher DMC frequencies indicate a general tendency in interpreted discourse to use more DMCs, or a higher frequency is simply the result of interpreters omitting sections, while maintaining cohesion. This would mean that interpreters omit sections, conveying only essential information, but without compromising discourse cohesion between the omitted and the interpreted sections, thus transferring and inserting DMCs into abridged texts, creating shorter but equally or more cohesive texts. This could explain why interpreted texts are shorter but contain more DMCs.

Then this paper investigated if interpreted discourse contains more filled pauses. Both in terms of duration and frequency, filled pauses have been found to be more prevalent in interpreted than in original Hungarian discourse. As filled pauses are both over seven times as long and as frequent in interpreted than in original Hungarian discourse, this hypothesis is confirmed. In accordance

with other studies, male interpreters hesitate more than women (cf. Collard & Defrancq, 2017).

This paper also sought to establish whether certain discourse properties form interdependent relationships, namely DMC frequency with delivery speed, and DMC frequency with filled pauses. This study has found a positive moderate, statistically significant correlation between delivery speed and DMC frequency in interpreted Hungarian discourse, lending support to hypothesis 4. However, the fact that a very similar level of correlation is revealed in original Hungarian, as well as English speeches, indicates that this relationship is not exclusive to interpreted discourse.

By contrast, the relationship between filled pauses and DMC frequency is not nearly as straightforward. It differs between female and male interpreters, although overall both filled pause duration and frequency correlate weekly and negatively in the IHC. Despite these findings supporting hypothesis 5, weak correlation and considerable individual variation in the data caution against confirming this relationship.

Finally, this paper also tested potential gender differences among interpreters. The only statistically significant variation is found in filled pause duration: male interpreters produce longer filled pauses. In this study, female interpreters used DMCs slightly more frequently, while the differences on the other measures were negligible. Correlation tests exposed more divergence between the sexes. While DMC frequency and delivery speeds correlated significantly in both groups, filled pause duration and frequency correlated positively with DMC frequency for women, and negatively for men. As these trends are weak and not significant, they are most likely caused by individual variation. On the basis of these results, hypothesis 6 is rejected.

As a final note, the role of individual variation deserves more intense research attention. Due to the size of the corpora used here, no broad generalisations can be reached. However, the results of this study do underline the need to account for individual differences when examining interpreted discourse.

Acknowledgements

Supported by the ÚNKP-17-3 New National Excellence Program of the Ministry of Human Capacities.

References

Bendazzoli, C. (2018). Corpus-Based Interpreting Studies: Past, Present and Future Developments of a (wired) Cottage Industry. In M. Russo, C. Bendazzoli, & C. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies* (pp. 1–19). Singapore: Springer.

Bendazzoli, C., Russo, M., & Defrancq, B. (2018). Corpus-based interpreting studies : a booming research field. *InTRAlinea Online Translation Journal*, (pp. 1:1–1:2).

Bernardini, S., Ferraresi, A., & Miličević, M. (2016). From EPIC to EPTIC — Exploring simplification in interpreting and translation from an intermodal perspective. *Target*, *28*, 61–86. doi:10.1075/target.28.1.03ber.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge University Press.

Blakemore, D., & Gallai, F. (2014). Discourse markers in free indirect style and interpreting. *Journal of Pragmatics*, *60*, 106–120.

Collard, C., & Defrancq, B. (2017). *Sex Differences in Simultaneous Interpreting: A Corpus-Based Study*. [Poster]. Conférence Internationale permanente d'Instituts Universitaires de Traducteurs et Interprètes (CIUTI)'s Forum, Geneva. http://hdl.handle.net/1854/LU-8518872.

Collard, C., & Defrancq, B. (2019). Predictors of ear-voice span, a corpus-based study with special reference to sex. *Perspectives*, *27*, 431–454. doi:10.1080/0907676X.2018.1553199.

Collard, C., & Defrancq, B. (2020). Disfluencies in simultaneous interpreting, a corpus-based study with special reference to sex. In L. Vandevoorde, J. Daems, & B. Defrancq (Eds.), *New Empirical Perspectives on Translation and Interpreting* (p. 264–300). Routledge. URL: https://doi.org/10.4324/9780429030376-12. doi:10.4324/9780429030376-12.

Crible, L. (2017). Discourse markers and (dis)fluency in English and French: Variation and combination in the DisFrEn corpus. *International Journal of Corpus Linguistics*, *22*, 242–269. doi:10.1075/ijcl.22.2.04cri.

Crible, L. (2018). *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers*. John Benjamins Publishing Company.

Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency. *Languages in Contrast*, *17*, 69–95. doi:10.1075/lic.17.1.04cri.

Defrancq, B. (2016). Well, interpreters... a corpus-based study of a pragmatic particle used by simultaneous interpreters. In G. Pastor, & M. Seghiri (Eds.), *Corpus-based Approaches to Translation and Interpreting* (p. 105–128). Peter Lang D. URL: https://doi.org/10.3726/b10354. doi:10.3726/b10354.

Defrancq, B. (2018). The European Parliament as a discourse community: Its role in comparable analyses of data drawn from parallel interpreting corpora. *The Interpreters' Newsletter*, *23*, 115–132. URL: https://doi.org/10.13137/2421-714x/22401. doi:10.13137/2421-714x/22401.

Defrancq, B., & Plevoets, K. (2018). Over-uh-Load, Filled Pauses in compounds as a Signal of cognitive Load. In M. Russo, C. Bendazzoli, & B. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies* (pp. 43–64). Springer Singapore. doi:10.1007/978-981-10-6199-8_7.

Defrancq, B., Plevoets, K., & Magnifico, C. (2015). Connective Items in Interpreting and Translation: Where Do They Come From? In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2015* (p. 195–222).

Springer International Publishing volume 3. doi:10.1007/978-3-319-17948-3_9.

Dér, Cs. I., & Markó, A. (2010). A Pilot Study of Hungarian Discourse Markers. *Language and Speech*, *53*, 135–180. doi:10.1177/0023830909357162.

Götz, A. (2017). Az első magyar intermodális korpusz bemutatása. Kutatási lehetőségek. *Argumentum*, *13*, 126–139.

Götz, A. (2018). Sex differences in interpreted hungarian discourse [poster]. In *2nd International Conference on Sociolinguistics (ICS.2), Budapest*. URL: http://rgdoi.net/10.13140/RG.2.2.28928.48645.

Götz, A. (2019a). A corpus-based, individualised investigation of Hungarian of interpreted discourse.

Götz, A. (2019b). Diskurzusjelölők és kötőelemek gyakorisága írott és beszélt, mediált és nem mediált diskurzusokban. In K. Laczkó, & S. Tátrai (Eds.), *Kontextualizáció és metapragmatikai tudatosság* (p. 291–313). ELTE Eötvös József Collegium.

Hale, S. (1999). Interpreters' treatment of discourse markers in courtroom questions. *Forensic Linguistics*, *6*, 57–82. doi:10.1558/sll.1999.6.1.57.

Hale, S. B. (2010). *The discourse of court interpreting: Discourse practices of the law, the witness, and the interpreter*. John Benjamins Pub. (Pbk. ed., with corrections).

Horváth, V. (2014). *Hezitációs jelenségek a magyar beszédben*. ELTE Eötvös Kiadó.

Magnifico, C., & Defrancq, B. (2016). Impoliteness in interpreting: A question of gender? *Translation and Interpreting*, *8*, 26–45. doi:10.12807/ti.108202.2016.a0.

Magnifico, C., & Defrancq, B. (2017). Hedges in conference interpreting: The role of gender. *Interpreting*, *19*, 21–46. doi:10.1075/intp.19.1.02mag.

Magnifico, C., & Defrancq, B. (2020). Norms and gender in simultaneous interpreting: A study of connective markers. *Translation & Interpreting*, *12*, 1–17. doi:10.12807/ti.112201.2020.a01.

Markó, A., & Dér, Cs. I. (2011). Diskurzusjelölők használatának életkori sajátosságai. In J. Navracsics, & Zs. Lengyel (Eds.), *Lexikai folyamatok egy- és kétnyelvű közegben. Pszicholingvisztikai tanulmányok II* (p. 49–61). Tinta Könyvkiadó. URL: http://real.mtak.hu/26155/.

Plevoets, K., & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting: A corpus-based regression analysis. *Translation and Interpreting Studies*, *11*, 202–224. URL: https://doi.org/10.1075/tis.11.2.04ple. doi:10.1075/tis.11.2.04ple.

Plevoets, K., & Defrancq, B. (2018). The cognitive load of interpreters in the European Parliament: A corpus-based study of predictors for the disfluency uh(m). *Interpreting*, *20*, 1–28. URL: https://doi.org/10.1075/intp.00001.ple. doi:10.1075/intp.00001.ple.

Russo, M. (2018). Speaking patterns and gender in the european parliament interpreting corpus: A quantitative study as a premise for qualitative investigations. In M. Russo, C. Bendazzoli, & B. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies* (p. 115–131). Springer Singapore. doi:10.1007/978-981-10-6199-8_7.

Schirm, A. (2019). A diskurzusjelölő-használat életkori sajátosságai a nyelvi interjú szövegtípusában. *Beszédkutatás*, . URL: https://doi.org/10.15775/beszkut.2019.187-205. doi:10.15775/beszkut.2019.187-205.

Vukov Raffai, E. (2016). A diskurzusjelölő-választások életkori sajátosságai az így, ilyen, hát, mondjuk, ugye esetében. *Magyar Nyelvőr*, *140*, 483–497.

Appendix 1

Hungarian DMCs: *ahogy* 'as, like', *akkor* 'then', *ám* 'although', *azért* 'because of that', *azonban* 'however', *aztán* 'then', *bár* 'although', *csak* 'just, only', *de* 'but', *-e* 'if, whether', *egyébként* 'by the way', *éppen* 'just', *és* 'and', *ezért* 'because of this', *ha* 'if', *hanem* 'but', *hát* 'well', *hiszen* 'since', *hogy* 'that conj.', *hogyha* 'if', *így* 'so', *illetve* 'and', *is* 'too, also', *itt* 'here', *már* 'already', *mármint* 'meaning', *még* 'yet', *mégis* 'still, yet, nevertheless', *mert* 'because', *most* 'now', *nemtom* 'dunno', *noha* 'although, while', *nos* 'well', *pedig* 'yet', *például* 'for example', *s* 'and', *sőt* 'what is more', *talán* 'maybe', *tehát* 'because', *tényleg* 'really', *tudniillik* 'namely', *úgy* 'so', *ugyan* 'although', *ugyanakkor* 'at the same time, nonetheless', *ugyanis* 'since', *ugye* 'is it not?, right?', *úgyhogy* 'so', *vagy* 'or', *vagyis* 'namely', *vajon* 'I wonder', *valamint* 'as well', *viszont* 'but'

English DMCs: *actually, after, albeit, already, also, although, and, anyway, because, before, but, considering, either, even, finally, here, however, if, indeed, instead, maybe, meanwhile, nevertheless, now, oh, okay, once, only, or, otherwise, secondly, since, then, therefore, though, till, too, unless, until, well, when, whenever, where, whereas, while, whilst, yeah, yet*

Appendix 2

Interpreted Hungarian Corpus

Speaker and text	Speech time	Word count	DMCs	Filled p. (sec)	Filled p. no.
F (1)	119.08	233	32	6.6	20
F (1)	76.43	125	16	3.22	13
F (1)	69.14	130	13	3.17	16
F (2)	35.44	68	12	1.22	4
F (2)	76.91	124	15	1.85	5
F (3)	59.5	114	11	1.49	4
F (3)	66.3	116	9	0.77	2
F (3)	67.84	139	22	0.34	1
F (3)	111.12	217	22	1.17	5
F (3)	61.78	117	9	2.09	4
F (4)	170.61	285	38	7.98	26
F (4)	98.52	145	17	5.28	17
F (4)	71.13	115	11	3.01	10
F (4)	188.21	305	48	18.16	57
F (4)	85.84	157	19	0.96	5
F (4)	114.83	189	19	1.19	5
F (4)	67.43	108	17	1.14	6
F (5)	105.82	186	23	3.9	17
F (5)	70.52	107	11	4.02	18
F (5)	162.1	258	39	10.19	26
F (5)	89.73	157	22	6.52	23
F (5)	75.3	123	18	10	27
F (5)	78.72	151	23	1.05	6
F (5)	69.87	126	17	2.38	13
F (5)	157.75	242	30	4.41	18
F (5)	67.85	107	20	5.69	18
F (5)	71.97	96	11	2.61	11
Subtotal (f)	2489.74	4240	544	110.41	377

Speaker and text	Speech time	Word count	DMCs	Filled p. (sec)	Filled p. no.
M (1)	86.46	165	13	2.05	7
M (1)	73.9	137	12	1.27	4
M (1)	64.33	125	17	1.89	7
M (2)	84.76	118	12	4.41	14
M (2)	57.83	87	7	3.9	12
M (2)	152.61	338	46	6.3	19
M (2)	63.65	98	12	6.04	22
M (2)	78.35	136	18	7.48	21
M (2)	67.43	117	7	6.21	18
M (2)	58.86	118	18	2.32	10
M (2)	77.72	135	14	5.55	20
M (2)	95.31	146	18	5.99	18
M (2)	35.84	58	2	4.44	8
M (2)	77.62	137	13	6.52	18
M (2)	255.7	378	26	16.05	51
M (2)	141.52	217	29	8.32	27
M (3)	66.39	117	20	3.49	6
M (3)	63.23	99	10	3.06	9
M (4)	72.26	145	19	1.12	2
M (4)	234.81	390	47	11.64	23
M (4)	108.82	205	26	6.22	13
M (5)	133.36	221	25	3.66	10
M (5)	79.5	137	21	3.06	7
Subtotal (m)	2230.26	3824	432	120.99	346