

Nyelvtudományi Kutatóközpont

BESZÉDTUDOMÁNY

SPEECH SCIENCE

2021

Budapest

Hungarian Research Centre For Linguistics

BESZÉDTUDOMÁNY – SPEECH SCIENCE

2021 (2)

Szerkesztők/Editors:

Grácsi, Tekla Etelka

Gyarmathy, Dorottya

Horváth, Viktória

Krepsz, Valéria

Mády, Katalin

Nyelvtudományi Kutatóközpont, Eötvös Loránd Kutatási Hálózat (ELKH)

Hungarian Research Centre for Linguistics,

Eötvös Loránd Research Network (ELKH)

Budapest

Szerkesztők/Editors:

Grácz, Tekla Etelka

Gyarmathy, Dorottya

Horváth, Viktória

Krepsz, Valéria

Mády, Katalin

Szerkesztőbizottság/Editorial board:

Bunta, Ferenc (University of Houston)

Hámori, Ágnes (Hungarian Research Centre for Linguistics)

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

Huntley-Bahr, Ruth (University of South Florida)

Markó, Alexandra (Eötvös Loránd University & MTA–ELTE „Lendület”
Lingual Articulation Research Group)

Mildner, Vesna (University of Zagreb)

Olaszy, Gábor (Budapest University of Technology and Economics)

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

Sztahó, Dávid (Budapest University of Technology and Economics)

Trouvain, Jürgen (Saarland University)

White, Laurence (Newcastle University)

Technikai szerkesztés/Typesetting: Ligeti-Nagy, Noémi

Borítótér/Cover design: Gyarmathy, Dorottya ©

Korrektúra/Proofreading: Vakula, Tímea

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

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

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

1068 Budapest, Benczúr u. 33.

Szerkesztői előszó

Kedves Kollégák!

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

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

Üdvözlettel: a kötet szerkesztői

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

Editorial foreword

Dear Colleagues,

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

Sincerely, the Editors

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

Tartalomjegyzék/Table of contents

Arthur Frigyes Viktor – Csapó Tamás Gábor: Szájról olvasás automatizálása mély neurális hálózatok és mobilalkalmazás-kezelőfelület alkalmazásával	7
Grácz Tekla Etelka – Kohári Anna: Spiránsok mássalhangzó-kapcsolatokban és intervokális helyzetben	24
Neuberger Tilda: A rövid és hosszú zöngétlen explozívák észlelése felnőttéknél és kisiskolás gyermekeknél	65
Markó Alexandra – Huszár Anna – Krepsz Valéria – Grácz Tekla Etelka: Az alapfrekvencia jellemzőinek longitudinális összevetése felnőtt beszélők felolvasásában	99
Szabados Laura: Hangsúlyos szótagok vizsgálata a francia nyelvben a Melodic Analysis of Speech (MAS) módszer használatával: párizsi és québeci nyelvváltozatok összehasonlítása	135
Bóna Judit – Steklács János: Hangos olvasás közbeni hibázások, hibajavítások változása negyedik és ötödik osztály között	154
Németh Zsuzsanna: A nemlexikális <i>ö</i> hang mint diskurzusjelölő magyar nyelvű társalgásokban	173
Harmati-Pap Veronika – Vadász Noémi – Kas Bence – Tóth Ildikó: Anyai dajkanyelvi narratívák lexikai és szintaktikai jellemzőinek longitudinális vizsgálata	207
Trencsényi Réka – Czap László: Artikulációs fonetikai jellemzők verifikálása kvantitatív adatokkal	243
Tímea Fekete – Péter Mihajlik: End-to-End Recognition of Spontaneous Speech on the Hungarian BEA Database	261

Szájról olvasás automatizálása mély neurális hálózatok és mobilalkalmazás-kezelőfelület alkalmazásával

Arthur Frigyes Viktor¹, Csapó Tamás Gábor^{1,2}

¹*Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék*

²*MTA-ELTE „Lendület” Lingvális Artikuláció Kutatócsoport*

Abstract

Automatic lipreading is a technique to predict the spoken content using lip video input. The advantage of lip video compared to other articulatory techniques (e.g. ultrasound tongue imaging and MRI) is that it is easily available and affordable: most modern smartphones have a front camera. There are already a few solutions for lip-to-speech synthesis, but they mostly concentrate on offline training and inference. In this research, we propose a system built from three components: a backend for deep neural network training and inference, a webservice responsible for the communication between the server and the client, and a frontend as a form of a mobile application. We trained two approaches, both using convolutional and recurrent neural networks. In the first case, we record the mimic movements of the whole face and from this information, we deduce the phonetic information. In the latter case, only the mouth area is available as input data for the neural network. Our initial evaluation shows that the scenario is feasible: a top-5 classification accuracy of 74% is combined with feedback from the mobile application user, making sure that the speaking impaired might be able to communicate with this solution. The results of the articulatory-to-text conversion can contribute to the development of 'Silent Speech Interface' (SSI) systems. The essence of SSI is recording the articulation organs while the user of the device actually does not make a sound but yet the machine system is capable to synthesize speech based on the movement of the organs.

Keywords: vid2speech, lip-reading, lip video, DNN, speech technology

1. Bevezetés

A beszéd megértése képi információk alapján rendkívül nehéz feladat. Nehezen általánosítható, tekintve, hogy minden beszélő artikulációja kisebb-nagyobb mértékben eltér egymástól. Az artikuláció-akusztikum konverziós módszerek célja, hogy artikulációs mozgás alapján szintetizáljanak beszédet. Az artikulációs információ lehet például a nyelv mozgása ultrahanggal rögzítve (Csapó

Email addresses: hello@victorarthur.com (Arthur Frigyes Viktor),
csapot@tmit.bme.hu (Csapó Tamás Gábor)

Beszédtudomány – Speech Science 2021. 7–23. DOI: 10.15775/Besztud.2021.7-23

et al. [2017a, b], elektromágneses artikulográf [Cao et al. 2018], felszíni eletromiográfia [Diener & Schultz 2018], vagy mágnesesrezonancia-képalkotás [Csapó 2020].

1.1. Artikuláció-akusztikum átalakítás ajakvideó alapján

A konverzió egy másik lehetséges megoldása a beszédszintézis kizárólag egy arcról vagy ajakról készült videó képkockáiból [Rácz & Csapó 2020]. 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], Akbari et al. [2018] és Rácz & Csapó [2020]. Le Cornu & Milner [2015] a felvett ajakmozgás alapján mély neurális hálózatot tanítanak be a beszéd spektrális paramétereinek megbecsülésére, majd ebből egy vokóderrel beszédet szintetizálnak. Ephrat & Peleg [2017] konvolúciós hálózatokat (CNN) alkalmaznak, és a GRID audiovizuális adatbázison megmutatják, hogy a mély neuronháló betanítása során nem látt szavakat is képes a rendszer érthetően szintetizálni, 52%-os pontossággal. Akbari et al. [2018] egy újfajta spektrális reprezentációt (hallási spektrogram) alkalmaznak, és a gépi tanulást autoenkóder, konvolúciós, és rekurrens hálózatok (RNN) kombinációjával valósítják meg. Rácz & Csapó [2020] szintén CNN és RNN hálózatokat használnak a közvetlen 'lip-to-speech' beszédszintézisre.

Wand et al. [2016] és Sun et al. [2018] munkájukban a közvetett módszert mutatják be, azaz automatikus szájról olvasást végeznek. Wand et al. [2016] mutatták be az egyik első teljesen mély neuronháló alapú megoldást, melyben rekurrens hálózatot (LSTM, Long-Short Term Memory) alkalmaznak. Korábbi munkákhoz hasonlóan a GRID adatbázison tesztelték módszerüket, melynek eredménye 79.6%-os szófelismerési arány lett. Sun et al. [2018] a Lip-Interact rendszert mutatták be, melynek segítségével néma ajakmozgással lehet irányítani a mobiltelefon bizonyos funkcióit. Parancsszavas felismerést valósítanak meg (44 parancsot megkülönböztetve) a mobiltelefon első kameráját felhasználva. Az eredmények szerint a Lip-Interact elsősorban akkor hasznos, ha a

felhasználó egyik vagy mindkét keze foglalt (pl. vezetés közben), és ilyenkor a mobiltelefonos interakciók az automatikus szájról olvasás felhasználásával hatékonyabbá válnak.

A némabeszéd-interfész (Silent Speech Interface, SSI) az artikuláció-akusztikum konverziós módszerek egy olyan távlati alkalmazása, amelynek használatával némán beszélve, „tátogva” adhatunk ki hangot (Denby et al., 2010; Csapó et al., 2017b; Kimura et al., 2019; Gonzalez-Lopez et al., 2020). A némabeszéd-interfesszel 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.

1.2. A jelen kutatás célja

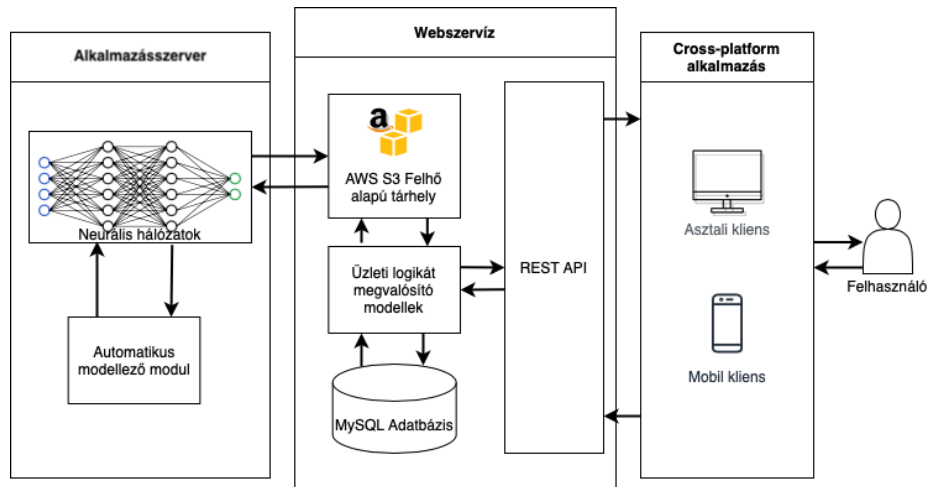
A kutatás célja egy megvalósíthatósági tanulmány: azt teszteltük, hogy egy mobil kliens alapú automatikus szájról olvasó rendszerhez milyen komponensek szükségesek. Választásunk a fenti lehetőségek közül egy 'közvetlen' módszer megvalósítására esett, mert így a felhasználónak lehetősége van az artikuláció-beszéd felismerés után a legvalószínűbb eredmények közül választani, és így mindenképp a helyes szöveg kerül a szövegfelolvasó bemenetére. Távlati célunk az alkalmazással, hogy hangtalan, illetve kis hangintenzitással beszélő páciensek számára egy alternatív kommunikációs eszközként használható megoldást nyújtson.

2. Módszerek

A kutatás során megvalósítottuk a teljes architektúrát, a háttérszerverrel és mobil kliens alkalmazással együtt.

2.1. A rendszer főbb komponensei

A prototípus rendszer három fő komponensből és azok alárendelt rendszereiből tevődik össze, melyek az 1. ábrán láthatók.



1. ábra. Rendszer architektúra

Alkalmazáserver Az első komponens az alkalmazáserver, mely közvetlenül az AWS S3 tárterületre felkerült új, a felhasználóra specifikus videó anyagot tanító adatként használva elvégzi a mély neuronháló tanítását, azaz frissíti a neurális hálózat súlymátrixát. Az így elkészült predikciós modell alapján később a mobilalkalmazás beszédpredikcióra képes. A legenerált modelleket az alkalmazáserver felmásolja a tárterületre, ahonnan azt a későbbi felhasználás során elérheti.

Webszerviz Az alkalmazáserver és a mobilalkalmazás között helyezkedik el egy webszerviz, melynek főbb feladatai a felhasználói autentikáció, adatok rögzítése, hívások kezelése lokális adatbázisban, valamint a tanulási adatok mentése és feltöltése felhő alapú tárterületre.¹

¹A kiértékeléshez felhasznált adatok AWS S3 szerveren kerültek tárolásra, míg a webszerviz DigitalOcean szerverein került kiszolgálásra, az SSL tanúsítványt a LetsEncrypt bocsájtotta ki.

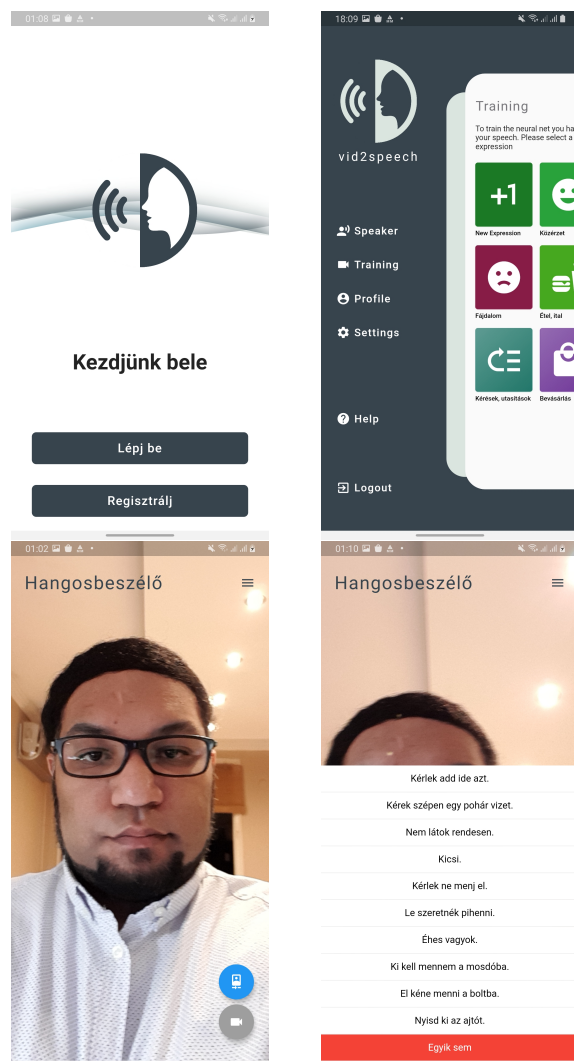
Cross-platform mobilalkalmazás A felhasználó közvetlenül a mobilalkalmazással van kapcsolatban, mely a Flutter keretrendszer által biztosított platformfüggetlenség előnyeit élvezi. Ennek okán iOS és Android mobil operációs rendszereken egyaránt elérhető a szolgáltatás. A betanítási fázisban a felhasználónak előre megadott mondatokat kell némán felolvasnia, melyeket utána a webszervizen keresztül az alkalmazásszerverre töltünk fel; melyből a gépi tanulási módszerrel egy predikciós modell készül.

A mobilalkalmazáson belül a predikciós mód esetén az ajakvideó alapú felismerés után a felhasználó számára megjelenítjük a legvalószínűbb találatokat, melyekből választhat. A kiválasztott szöveget átadjuk a rendszerben lévő szövegfelolvasónak (Androidon lehetséges alternatívák például: gTTS, ProfiVox és ProfiVox-HMM). A jelen kutatás során a szövegfelolvasó modulok előnyeit/hátrányait nem vizsgáltuk. A fejlesztés során Samsung Galaxy Note 9, valamint Huawei P10 mobil eszközön futó Android 9.0 és 10.0 operációs rendszereken történt a kiértékelés és az adatrögzítés. A [2](#) ábra a mobilalkalmazás angol és magyar változataiból jelenít meg képernyőképeket.

A leírt három részből álló struktúra lehetővé teszi a nagy számítási kapacitást igénylő gépi tanulási folyamat elkülönítését: az ilyen műveletek aszinkron módon történnek meg az alkalmazásszerveren. Az itt lezajló kiértékelés eredménye minimális késleltetéssel a mobil eszközön megjelenik, ezzel a kliens oldalt mentesítve a nagyobb erőforrás igényű feladatok elvégzésétől. Továbbá megvalósítja a "Single-point-of-change" elvét, tehát elegendő egy helyen változtatunk a rendszert és annak hatását vizsgálhatjuk a kimeneten, a többi komponens megváltoztatása nélkül. Ilyen módon a rendszer alkalmas új neurális hálózatok hatásának tesztelésére.

2.2. Tanító adatok rögzítése a mobil kliens segítségével

Mind a tanító adatokat, mind pedig a predikció során használt videóanyagot a mobilalkalmazás rögzíti.



2. ábra. Képernyőképek a mobilalkalmazásról (angol és magyar nyelvű verziók). Bal felső: Bejelentkezési képernyő. Jobb felső: Navigációs és tanítási képernyő. Bal alsó: Néma ajakmozgás rögzítése. Jobb alsó: A szájról szövegre konvertálás eredményének megjelenítése (több eredmény egy listában).

A tanítási adatok szöveges tartalmához 88 kifejezést választottunk ki. A 88 példamondat a StrokeAid² elnevezésű segédprogramból került átemelésre, aminek célja, hogy segítséget nyújtson a stroke-on átesett páciensek számára. A StrokeAid programban található kifejezések segítségével a páciens gyorsan tud reagálni egy beszélgetésben.

Tanítási módban a rögzítés során a felhasználókat arra kérjük, hogy ötször ismételjék el ezeket a mondatokat, így minden példamondathoz öt felvétel keletkezik (azaz összesen 440 videó). Az arcról készült videót 720x1280 képpontos felbontással és 25 képkocka/másodperces sebességgel rögzíti a mobilalkalmazás az okostelefon elülső kamerájával, és minden mondat rögzítése után az adatokat a webszervízen keresztül elküldjük az alkalmazásszerverre. A megvalósíthatósági tanulmányunk során egy férfi beszélővel (a cikk első szerzője) teszteltük a tanítást, aki felmondta a 440 kifejezést.

2.3. Videóadatok feldolgozása

Miután megtörtént a tanító adatok gyűjtése a mobilalkalmazással, a feldolgozás hátralévő részéért az alkalmazásszerver felelős. Az arc egyes jellemző pontjainak meghatározásához három eljárást teszteltünk: 2d106det MobileNet (Deng et al., 2019), Google Firebase ML Kit FireVision³, és 'shape_predictor_68_face_landmarks' modell, mely a DLib programcsomag része⁴. Miután a sebesség és kompatibilitási teszteket elvégeztük, a választásunk a DLib módszerre esett, mely 68 jellemző pontot jelöl ki az arcon. Ennek megfelelően a neurális hálózat bemeneteként a videókból csak a száj környéki régiót mutató, 299x299 pixelre átméretezett részeket használtuk fel. Néhány minta képkocka látható a 3 ábrán.

²<https://play.google.com/store/apps/details?id=com.onlab.monddki>

³<https://firebase.google.com/docs/ml-kit/detect-faces>

⁴<https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/>



3. ábra. Néhány a videókból kivágott ajak-régió minta, amelyek a DNN bemenetére kerültek.

2.4. DNN tanítása az alkalmazáserveren

A rögzített 440 videó 60%-át tanítási és 40%-át validációs halmazra osztottuk. A jellemzők kinyerésére konvolúciós hálózatot, a végső osztályzáshoz pedig rekurrens hálózatot használtunk. Automatikus hiperparaméter optimalizálást végeztünk, melynek során a mély neuronháló különböző paramétereit több tartományban állítottuk, majd a betanított rendszereket kiértékeljük. A tanítás során a korai leállítást 10 epoch türelemmel alkalmaztuk. A hálózatot osztályozási módban tanítottuk, kategorikus kereszt-entrópia költségfüggvény és ADAM optimalizáló használatával (tanulási ráta: 10^{-5}).

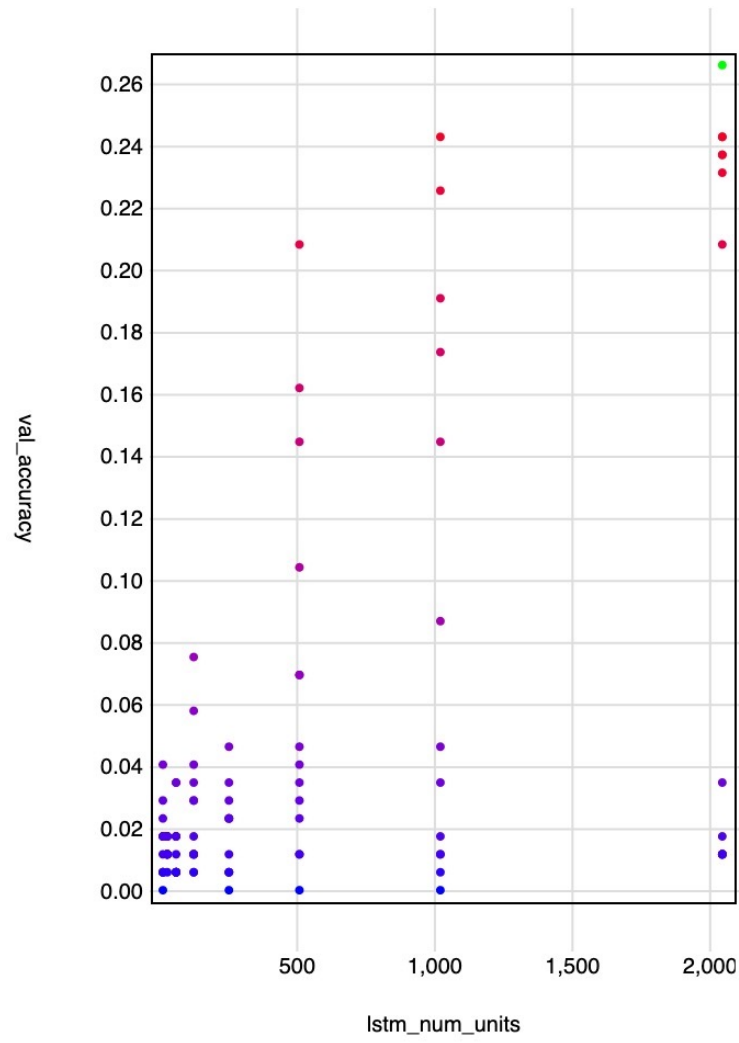
3. Eredmények és diszkusszió

3.1. Mély neuronhálós tanítás

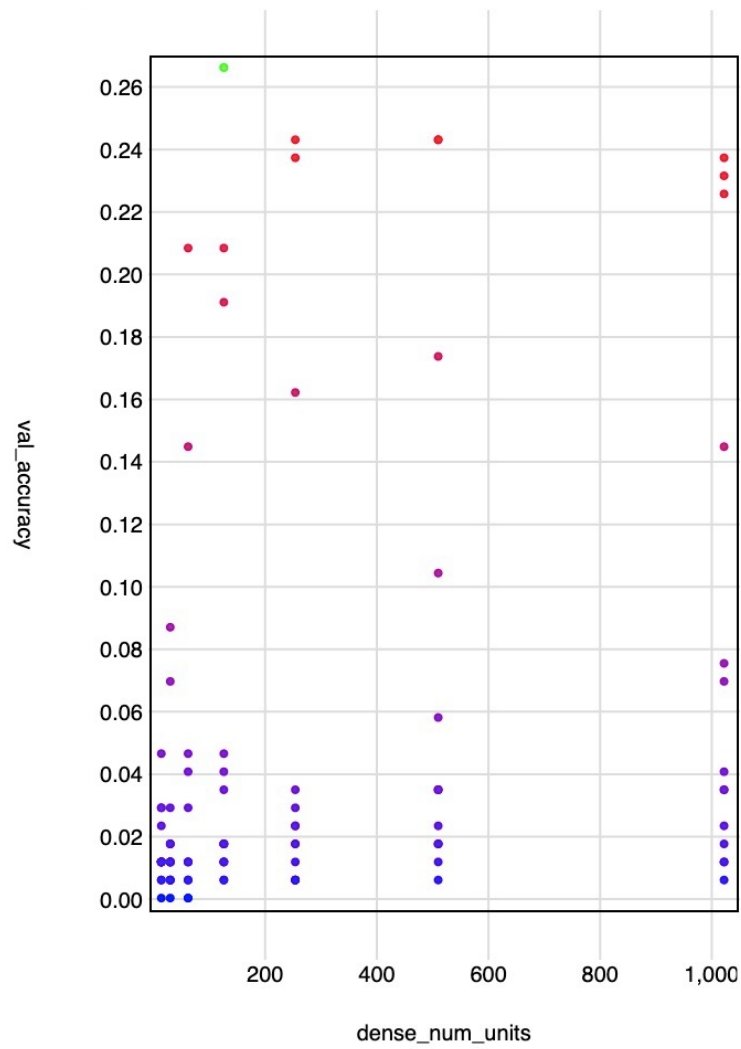
A mély neuronhálók hiperparaméter optimalizálása során kapott eredményeket a [4][6] ábrák mutatják. A [4] ábrán az LSTM rétegben (a hálózat rekurrens része, mely a szekvenciális adatok feldolgozásáért felel) lévő neuronok száma, az [5] ábrán az előreccsatolt rétegben lévő neuronok száma, míg a [6] ábrán a hálózatban használt Dropout hatása látható. Az optimális hálózati struktúra a következő: InceptionV3 a jellemző kinyeréséhez, amelyet egyetlen LSTM réteg követ, 2048 neuronnal, majd 10% Dropout és egy teljesen kapcsolt réteg, amelynek végén 128 neuron található.

3.2. Tesztelés új felvételekkel

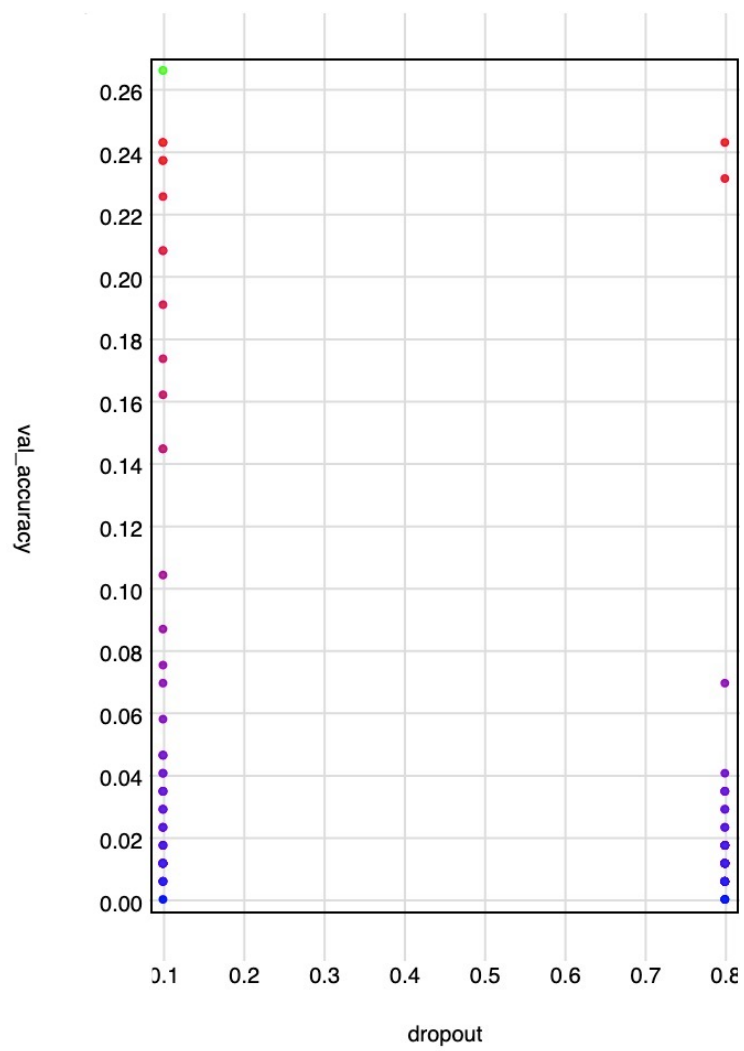
Tesztelés céljából új videofelvételeket rögzítettünk (a 88 magyar mondat mindegyikéhez egy-egy bemondás), biztosítva, hogy tanító és a kiértékelés során használt adatok ne kerüljenek ismételt felhasználásra, azaz függetlenek legyenek.



4. ábra. Top-1 validációs pontosság az LSTM rétegben lévő neuronok számának függvényében.



5. ábra. Top-1 validációs pontosság az előrecsatolt rétegben lévő neuronok számának függvényében.



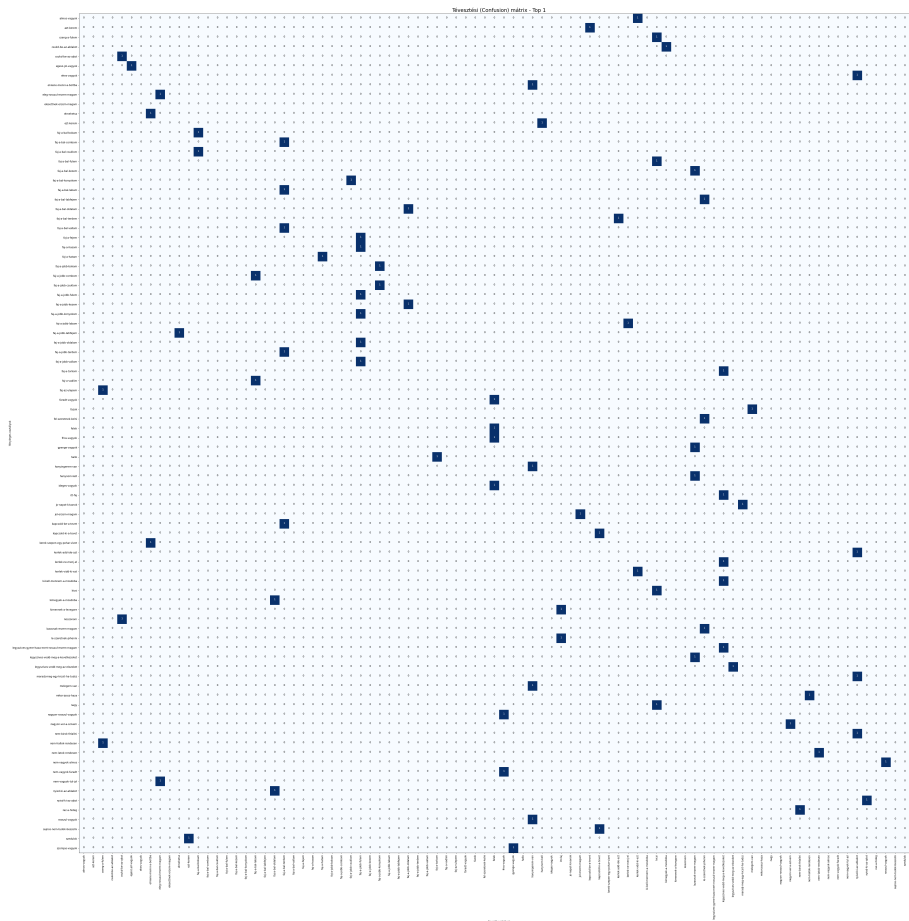
6. ábra. Top-1 validációs pontosság hálózatban használt DropOut függvényében.

Az osztályozás eredményeként a Top-1 tévesztési mátrix a [7] ábrán látható, míg a [8] ábra a Top-5 pontosságot mutatja. Összességében a végső modell 53% top-1 és 74% top-5 pontosságot ért el. Összehasonlíthatjuk ezt az emberi szájról olvasás teljesítményével, amely körülbelül 30% (Altieri et al. 2011). A tévesztési mátrixokban az optimális eset az lenne, ha a mátrix átlójában szerepelne a legtöbb eredmény. Top-1 pontosság esetén ([7] ábra) ezt nem sikerült elérni: sok esetben a néma videókat rosszul osztályozta a rendszer. Másrészt a top-5 tévesztési mátrix ([8] ábra) több elemet tartalmaz az átló körül, ami azt mutatja, hogy a hálózat elfogadható teljesítménnyel találta meg az alany által kimondott szöveget. A szakirodalmi áttekintés során a GRID adatbázison a többi rendszer nagyságrendileg hasonló eredményt ért el (Ephrat & Peleg 2017: 52%, Wand et al. 2016: 79.6%), bár ott a nyelv és a szótárméret is jelentősen különbözött a mi kísérleteinktől.

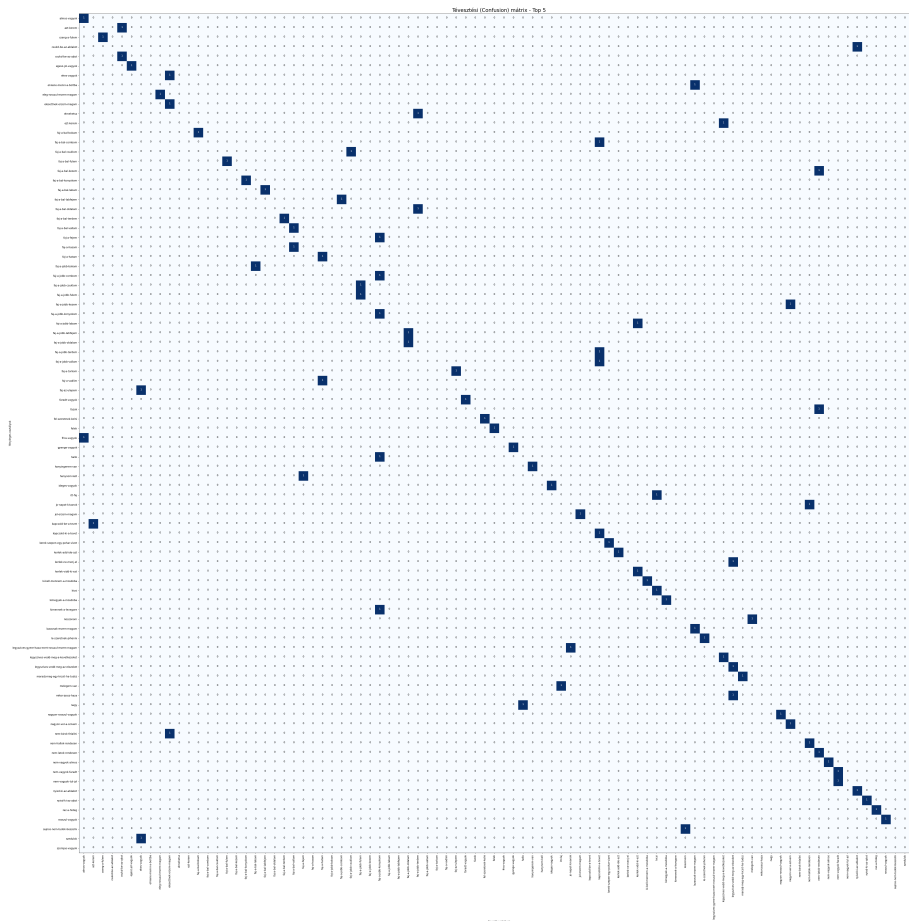
A gyakorlati megvalósításban az alkalmazáserveren futtatott predikció után a legvalószínűbb felismerési eredményeket visszaküldi a mobil kliensnek. Itt a felhasználó kiválaszthatja, hogy melyik volt a ténylegesen kimondott mondat (lásd [2] ábra, jobb alsó kép), mielőtt elküldené a rendszer szövegfelolvasó moduljába. Ez a lépés biztosítja a megfelelő mondat hangos felolvasását valós kommunikáció esetén.

4. Összegzés

A kutatás során automatikus szájról olvasásra betanított mély neurális hálózatokat terveztünk, majd egy end-to-end alkalmazás architektúrát fejlesztettünk. A rendszer mobilalkalmazást használ a felhasználóval történő interakcióra, míg a gépi tanulási lépések egy háttérszerveren történnek.



7. ábra. Top-1 tévesztési mátrix. (teljes méretben: https://github.com/victorarthur/vid2speech_images)



8. ábra. Top-5 tévesztési mátrix. (teljes méretben: https://github.com/victorarthur/vid2speech_images)

A kezdeti kiértékelésünk azt mutatja, hogy az eljárás megvalósítható, és az alkalmazást a potenciális végfelhasználók használni tudják majd. A némabeszéd-interfészek elsődleges célfelhasználói a beszédszervi sérüléssel élő emberek (Denby et al., 2010; Gonzalez-Lopez et al., 2020). Ezenkívül az automatikus szájról olvasás hasznos lehet, ha figyelembe vesszük az adatvédelmi aggályokat: egyesek nem érzik jól magukat, ha hangosan kell beszélniük okostelefonjukkal, amikor mások a közelben vannak.

A jövőbeli munkánk során a rendszert a célfelhasználói csoport több tagjával is tesztelni tervezzük. Más, összetettebb hálózatokat is szándékozunk használni, figyelembe véve a valós idejű kommunikációhoz szükséges gyors válaszsebességet.

Köszönetnyilvánítás

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

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.
- Altieri, N. A., Pisoni, D. B., & Townsend, J. T. (2011). Some normative data on lip-reading skills (L). *The Journal of the Acoustical Society of America*, 130, 1–4. doi:10.1121/1.3593376
- Cao, B., Kim, M., Wang, J. R., Van Santen, J., Mau, T., & Wang, J. (2018). Articulation-to-Speech Synthesis Using Articulatory Flesh Point Sensors' Orientation Information. In *Proc. Interspeech* (pp. 3152–3156). Hyderabad, India. doi:10.21437/Interspeech.2018-2484
- Csapó, T. G. (2020). Speaker dependent articulatory-to-acoustic mapping using real-time MRI of the vocal tract. In *Proc. Interspeech* (pp. 2722–2726). Shanghai, China. doi:10.21437/Interspeech.2020-0015

- Csapó, T. G., Grósz, T., Gosztolya, G., Tóth, L., & Markó, A. (2017a). DNN-Based Ultrasound-to-Speech Conversion for a Silent Speech Interface. In *Proc. Interspeech* (pp. 3672–3676). Stockholm, Sweden. doi:[10.21437/Interspeech.2017-939](https://doi.org/10.21437/Interspeech.2017-939).
- Csapó, T. G., Grósz, T., Tóth, L., & Markó, A. (2017b). Beszédszintézis ultrahangos artikulációs felvételekből mély neuronhálók segítségével. In *MSZNY 2017* (pp. 181–192).
- Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., & Brumberg, J. S. (2010). Silent speech interfaces. *Speech Communication*, 52, 270–287. doi:[10.1016/j.specom.2009.08.002](https://doi.org/10.1016/j.specom.2009.08.002).
- Deng, J., Roussos, A., Chrysos, G., Ververas, E., Kotsia, I., Shen, J., & Zafeiriou, S. (2019). The Menpo benchmark for multi-pose 2D and 3D facial landmark localisation and tracking. *International Journal of Computer Vision*, 127, 599–624.
- Diener, L., & Schultz, T. (2018). Investigating Objective Intelligibility in Real-Time EMG-to-Speech Conversion. In *Proc. Interspeech* (pp. 3162–3166). Hyderabad, India. doi:[10.21437/Interspeech.2018-2080](https://doi.org/10.21437/Interspeech.2018-2080).
- Ephrat, A., & Peleg, S. (2017). Vid2speech: Speech Reconstruction from Silent Video. In *Proc. ICASSP* (pp. 5095–5099). New Orleans, LA, USA. [arXiv:1701.00495](https://arxiv.org/abs/1701.00495).
- Gonzalez-Lopez, J. A., Gomez-Alanis, A., Martin Donas, J. M., Perez-Cordoba, J. L., & Gomez, A. M. (2020). Silent Speech Interfaces for Speech Restoration: A Review. *IEEE Access*, 8, 177995–178021. doi:[10.1109/access.2020.3026579](https://doi.org/10.1109/access.2020.3026579). [arXiv:2009.02110](https://arxiv.org/abs/2009.02110).
- 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](https://doi.org/10.1145/3290605.3300376).

- Le Cornu, T., & Milner, B. (2015). Reconstructing intelligible audio speech from visual speech features. In *Proc. Interspeech* (pp. 3355–3359). Dresden, Germany.
- Rácz, B., & Csapó, T. G. (2020). Ajakvideó alapú beszéd-szintézis konvolúciós és rekurrens mély neurális hálózatokkal. *Beszédtudomány – Speech Science*, 1, 57–72.
- 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. doi:[10.1145/3242587.3242599](https://doi.org/10.1145/3242587.3242599).
- Wand, M., Koutník, J., & Schmidhuber, J. (2016). Lipreading with long short-term memory. In *Proc. ICASSP* (pp. 6115–6119). Shanghai, China.

Spiránsok mássalhangzó-kapcsolatokban és intervokális helyzetben

Grácsi Tekla Etelka¹, Kohári Anna¹

¹*Nyelvtudományi Kutatóközpont*

Abstract

Voiceless alveolar and postalveolar sibilants were investigated comparatively in clusters followed by /t/ and intervocalic positions. The main question of the study was whether there is a difference in the coarticulatory patterns of these speech sounds between female and male speakers of Hungarian. 3 intervocalic /s/, 3 intervocalic /ʃ/, 3 V/st/, and 3 V/ft/ clusters were analysed in sentences read by 72 Hungarian speakers. The duration and spectral features (centre of gravity (CoG), standard deviation (sd), skewness, and kurtosis) of the sibilants and the manner of articulation of /t/ were measured. We found that the speakers' gender and the position of /t/ had a significant effect on the ratio of alternative and plosive-like /t/-realisations. The gender affected the duration, CoG, and skewness of the sibilants alongside one or two other factors, namely the consonants (/s/ or /ʃ/) and phonetic context. The influence of phonetic context on the duration manifested at a different magnitude for the two genders. The magnitude of the difference in the CoG of the two sibilants and the effect of the context on this value were dependent on the gender. The skewness was found to be influenced by a three-wise interaction of these factors. The results indicate that there is a difference in the coarticulatory patterns of the analysed two sibilants between the genders that, on the one hand, strengthens that the gender-related CoG-differences of /s/ do not merely result from their physiological differences in Hungarian, similarly to, e.g., English, German and Polish. On the other hand, the coarticulatory patterns also show gender-specific behaviour.

Keywords: szibilánsok, spektrális súlypont, koartikuláció

1. Bevezetés

A szibilánsok akusztikai jellemzőiben több nyelvben megfigyelték, hogy a nők és a férfiak ejtése között eltérés található. Az adatok nem kizárólag anatómiai különbségekre (az akadály előtti üreg méretére) vezethetőek vissza. A koartikulációs vizsgálatok – bár nem elsődleges céljuk volt a nemi különbség-

Email addresses: graczi.tekla.etelka@nytud.hu (Grácsi Tekla Etelka),
kohari.anna@nytud.hu (Kohári Anna)

gek feltárása – mégis mutattak eltérést a nők és férfiak ejtése között. A jelen tanulmányban azt a kérdést tettük fel, hogy a két zöngétlen szibiláns (/s/ és /ʃ/) intervokális és /t/-től követett pozícióban milyen eltéréseket mutat, és ezek között felfedezhető-e különbség a magyar felnőtt nők és férfiak ejtése között.

1.1. A szibilánsok képzése

A réshangok képzése során szűkület jön létre valamely artikulációs szervpár tagjai között a toldalékcsoportban, és a levegő rendezetlen, turbulens kiáramlása a keskeny résen keresztül sajátos zörejt eredményez (Shadle, 1990; Stevens, 2000). A szűkület létrejöhet például a fogmedernél, ezt nevezzük alveoláris réshangnak, illetve megvalósulhat a fogmeder mögötti területen is, amelyet posztalveoláris réshangnak hívunk. Jelen tanulmányban az alveoláris /s/ és a posztalveoláris /ʃ/ zöngétlen réshangokat vizsgáljuk meg a magyar beszédben különféle hangkörnyezetekben; intervokális helyzetben és az alveoláris zöngétlen felpattanó zárhang előtt (/t/). A spiránsok szerkezetét tipikusan a következő négy akusztikai paraméterrel szokás jellemezni: spektrális eloszlás súlypontja (Centre of Gravity, CoG), spektrális szóródás, spektrális csúcsosság és spektrális ferdeség (Forrest et al., 1988; Koenig et al., 2013). A spektrális eloszlás súlypont, röviden CoG-érték a mássalhangzó frekvenciaértékének amplitúdóval történő súlyozott átlagát jelenti. A többi akusztikai paraméter azt mutatja meg, hogy a teljes spektrális eloszlás hogyan viszonyul a spektrális eloszlás súlypontjához, azaz a CoG-hez a mérési időponton/időtartományban. A spektrális szóródás a spektrumösszetevő frekvenciák súlyponttól mutatott átlagos eltérését adja meg, tehát adott pillanatbeli CoG értékehez viszonyítva határozza meg, hogyan szóródnak a különböző frekvencia-összetevők. A spektrális ferdeség ugyanakkor a frekvencia-összetevők eloszlásának szimmetriáját mutatja meg. Amennyiben a spektrális ferdeség 0, akkor a CoG alatti és feletti terület kiterjedése ugyanakkora, amennyiben ez az érték pozitív, akkor viszont a súlypont feletti terület kiterjedése nagyobb. A spektrális csúcsosság pedig azt adja meg, hogy a spektrális szerkezet milyen mértékben tér el egy CoG-központú Gauss-görbétől. Minél laposabb a görbe, annál alacsonyabb a csúcsosság értéke.

1.2. A zöngétlen alveoláris és posztalveoláris spiráns akusztikai különbségei

A különböző spiránsok jól elkülöníthetők egymástól az akadály helye alapján, amely eltérések az akusztikai szerkezetükben is tetten érhetők (Stevens, 2000; magyarra pl.: Olaszy, 2007; Grácz & Krepsz, 2018). Az alveoláris réshang képzésekor a szűkület előtti artikulációs csatorna rövidebb, mint a posztalveoláris réshang esetében. Ennek következtében az alveoláris spiráns spektruma magasabb CoG értékeket mutat, mint a posztalveolárisé (Perkell et al., 2004). A CoG értékek mellett más spektrális mutatók esetében is eltéréseket tapasztaltak a két réshang között. Amerikai angolban azt találták, hogy az alveoláris réshang esetében szintén magasabb volt a spektrális csúcosság, a spektrális szórás és ferdeség viszont alacsonyabbnak bizonyult a posztalveoláris réshanghoz képest (Jongman et al., 2000). Hozzá kell tennünk azonban, hogy a különböző vizsgálatok eredményei nem teljesen konzisztensek minden spektrális mutató esetében. Míg a szakirodalomban a CoG értékei következetesen magasabbak zöngétlen alveoláris réshang esetében, mint posztalveolárisok esetében (Nittro-uer et al., 1989; Shadle & Mair, 1996; Jongman et al., 2000; Perkell et al., 2004; Nissen & Fox, 2005; Koenig et al., 2013; Jannedy & Weirich, 2016), addig a spektrális ferdeség vagy csúcosság tekintetében a fentebb leírt trendekkel ellentétes eredményeket is kaptak (Shadle & Mair, 1996; Nissen & Fox, 2005). Koenig et al. (2013) szerint az ellentmondások feltehetően a mérés módszertan különbségeire vezethetők vissza, mint például hogy hány helyen mérték a más-salhangzóban, és hogy átlagolták vagy nem a mért értékeket stb. Ez azonban a CoG-értékekre kapott tendenciákban is okozhatna eltérést, így felmerül a kérdés, hogy az egyes mérési módszertanok pontosan milyen módon vezethetnek eltérő értékekhez pont ezekben az eredményekben, de az sem egyértelmű, hogy esetleg lehet-e más magyarázat a különbségek hátterében.

1.3. Nemek közötti eltérések

A spiránsok akusztikai szerkezetére számos szociolingvisztikai tényező lehet hatással, ilyen például a beszélő lakhelye, dialektusa is (Van der Harst et al.,

[2007], [Jannedy & Weirich, 2016]). A jelen tanulmány kérdései szempontjából a nemek közötti eltéréseken van a fókusz.

Mind az alveoláris, mind a posztalveoláris réshang esetében megfigyelték, hogy a CoG-érték magasabb nőknél, mint férfiaknál ([Jongman et al., 2000], [Perkell et al., 2004], [Koenig et al., 2013]), a két réshang közti akusztikai távolság viszont ugyanakkora a két beszélői csoporton belül. A nőknél és a férfiaknál is egyértelműen elkülönülnek a különböző helyen képzett réshangok a CoG alapján. További nemek közti eltérések, hogy a nőknél a réshangok nagyobb spektrális szórását mutatták ki a férfiakhoz képest. A spektrális ferdeség alacsonyabb, a csúcosság pedig magasabb volt a nők ejtésében, mint a férfiakéban ([Jongman et al., 2000], [Fuchs & Toda, 2010]). A nemek közti réshangbeli eltérések mögött megemlíthetjük a biológiai eltéréseket, mivel a nőknek kisebb a szájürege, mint a férfiaknak. Másrészt viszont felmerültek az akusztikai vizsgálatok során is olyan mintázatok, amelyek alapján feltételezhető, hogy nem csak az akadály előtti toldalékcsoport mérete játszik szerepet az eltérésben. Lengyelben ez az akusztikai eltérés olyan mértékű, hogy a nők ejtésében egy „gyerekesnek”, „nem helyesnek” vélt – stigmatizált – /s/-változatot tapasztaltak és mértek, amely ugyanakkor a lengyel szibilánsok közti akusztikai szembenállást erősíti, ezért feltételezik, hogy a lengyel esetében ez egy olyan hangváltozási folyamat kezdete, amely a lengyel szibilánsok akusztikai megkülönböztetésére szolgál ([Czaplicki et al., 2016]).

Az artikulációs vizsgálatok megerősítették, hogy a két nem közti eltérés nem csak a fiziológiai különbségekből ered. [Fuchs & Toda (2010)] vizsgálatában azt találta, hogy részben az akadály előtti üreg mérete meghatározza a két nem között talált eltéréseket, de ennek a területnek a méretjellemzőire kontrollálva a statisztikai modelleket, továbbra is szignifikáns különbséget kaptak a nők és a férfiak eredményei között. Ez azt jelenti, hogy további meghatározó faktor is szerepet játszik a nemek közti eltérésben, csupán az akadály előtti terület méretbeli különbségével nem írható le a CoG-beli eltérés. A palatum hossza nem tért el jelentősen a nemek között. Kiemelik, hogy a metszőfogak hosszát nem vizsgálták, így erre nem kontrollálták az eredményeket. Az artikulációs adatok mellett

ugyanakkor az akusztikai eredményekben is találtak olyan részleteket, amelyek a biológiai eltéréseken túl is látható eltérés a nemek között. Jongman és kollégái ([2000]) azt találták, hogy míg a férfiak ejtésében a képzési hely mentén hátrafelé haladva egyre alacsonyabbak voltak a spektrális csúcsok értékei, a nők ejtésében a labiodentálisok és a dentálisok között fordított tendencia mutatkozott.

Percepció tesztek alapján azt találták angolban (pl. Linville [1998], Smyth et al. [2003]), dánban (Pharao et al. [2014]) és magyarban (Rácz & Schepács [2013]), hogy férfiak beszédében a réshangok magasabb CoG-vel történő ejtése femininebb, homoszexuális orientációjú percepciót eredményezett.

A nemek mellett a beszélő dialektusa, illetve lakhelye is hatással lehet a spiránsok akusztikai szerkezetére (Van der Harst et al. [2007], Jannedy & Weirich [2016]), akár oly módon is, hogy az alveoláris és a posztalveoláris réshang közti különbségek csökkennek. Német felolvasott beszédben például azt találták, hogy a CoG és a spektrális ferdeség egyértelműen eltér a két réshang esetében, ugyanakkor a spektrális csúcsosság és szórás a beszélők lakhelye szerint változott. Míg a Jenában lakó beszélők esetében csak a csúcsosság, a Buxtehudeban lakó beszélők esetében csak a spektrális szórás különbözött a két hangzó összevetésekor (Jannedy & Weirich [2016]).

1.4. Az alveoláris és posztalveoláris réshangok koartikulációja magánhangzókkal és mássalhangzókkal

A koartikulációs hatások tekintetében az alveoláris és posztalveoláris réshang rezisztenciája és agresszivitása igen erős (Recasens [2014]). A nagy koartikulációs rezisztencia azt jelenti, hogy a hangkörnyezet kevésbé van hatással képzésükre, mint más hangok produkciójára. Az agresszivitás pedig azt, hogy nagyobb mértékben befolyásolják a környező hangok ejtését, mint más beszédhangok. Ennek oka, hogy ejtésük során a nyelvhat precíz beállítására van szükség ahhoz, hogy nagy intenzitású spirantikus zörej jöjjön létre. A követő hang labializáltságával szemben azonban kevésbé rezisztensek lehetnek, hiszen az alveoláris és a posztalveoláris réshang megvalósítása során történő ajakmozgás kevésbé korlátozott (Koenig et al. [2013], Recasens [2014]). Labiális magánhangzó vagy mássalhangzó

előtti réshang CoG értékét alacsonyabbnak mérték, mint illabiális környezetben. Ennek oka, hogy a kerekítő ajakmozgás meghosszabbítja az elülső üreget. Továbbá a spektrális ferdeséget magasabbnak mérték, ha labiális hang követte a réshangot, mint amikor nem labiális (Munson, 2004; Koenig et al., 2013).

Recasens és Rodríguez (2016) katalán magánhangzók és alveoláristól palatális képzési helyig terjedő mássalhangzók artikulációs vizsgálatában felállított egy koartikulációs érzékenységi sorrendet, mely szerint az /f/-nél az /s/ valamivel érzékenyebb – és mivel a két jellemző negatívan korrelál, kevésbé agresszív – volt magánhangzós környezetek között összevetve. A vizsgálatban nyelvkontúrok pozícióit elemezték különböző VCV-hangsorokban. Az /s/ esetében az /f/-hez képest a nyelv elülsőbb területével hozzuk létre az akadályt, ezért az /s/ esetében nyelvnak az akadályhoz képest hátsóbb területeinek pozícióját, mozgását jobban befolyásolta a kontextus, mint az /f/ esetében. A mássalhangzókkal alkotott kapcsolatokban elemezve a mássalhangzókat nagyobb mértékű koartikulációs hatást találtak, mint magánhangzókkal alkotott kapcsolatokban (Recasens & Pallarès, 2001). Az /s/ és az /f/ összevetését nem külön, hanem az elülső alveolárisok (/s, l, r/) és az alveolopalatálisok (/ʃ, ʎ, ɲ/) közötti összevetésekből vonhatjuk el. Az alábbi mintázatok rajzolódtak ki: Az alveolárisok esetében az akadály jellemzőire kisebb hatással volt a hangkörnyezet, mint a posztalveolárisokra. A követő dorzális kontakt hatására azonban érzékeny volt az /s/, palatalizációra hajlamosabb, míg az /f/ az /s, l, r/ hatására depalatalizációra volt hajlamos.

Az alveoláris és posztalveoláris réshang CC-koartikulációját főként az angol nyelv /st/ és /ft/ szekvenciáiban elemezték sokan. Meg kell jegyezni, hogy bár ennek egyik oka az, hogy az /stɪ/ szekvenciában sok nyelvváltozatban megfigyelhető az alveoláris posztalveolárisrá válása, melyben nyelvváltozási folyamatot is feltételeznek, ugyanakkor a kettős szibiláns + zárhang kapcsolatokat is elemezték, részben mint kontrollkörnyezetet, ezért érdemes ezeket tárgyalni az /stɪ/ szekvenciák figyelembevételével.

Angolban tehát szókezdő pozícióban megfigyelték az /stɪ/ fonémakapcsolatokban az /s/ [ʃ]-realizációját (öf. l. Janda & Joseph, 2003), Janda és Joseph

(2003) pedig kibővítette ezeket a megfigyeléseket, miszerint további, nem csak /t/-t tartalmazó és nem csak szókezdő fonémakapcsolatokban jelenik meg. A legtöbb tanulmány szerint az /s/ CoG értékeinek módosulását a követő /ɹ/ hanggal való koartikuláció indítja el (vö. Phillips 2020). Figyelembe kell venni azt, hogy az angolban ez szókezdő pozícióban /s/ + /ɹ/ hangkapcsolat nem fordul elő, míg /fɹ/ igen (Algeo, 1978), illetve, hogy emellett a /tɹ/ hangkapcsolatok szó elején sok angol változatban [tʃ]-szerű realizációval valósulnak meg (öf.: Magloughlin, 2018). Ennek artikulációs hátterét (egy angol változatban) Magloughlin (2018) tárta fel: az ajkak és a nyelv elülső területének pozíciója a /tɹ/- és /dɹ/-hangkapcsolatokban a [tʃ] és [dʒ] affrikátákra, míg a nyelvgyök az [ɹ]-re jellemző pozíciót veszi fel a turbulens zörej időbeni közepén. Az affrikátává válás tehát nem teljes artikulációs szempontból. Dolgozatából az is kiderül, hogy ezen realizációk /tʃ/, /dʒ/ észlelete nagyban függ a további hangkörnyezettől. Ohala és Solé (2010) az ehhez hasonló jelenségeket úgy elemzi, hogy a felpattanó zárhangok egy magas nyelvi pozícióval rendelkező hanghoz történő felnyílása során aerodinamikai okokból hosszabb zörejelem valósul meg, ami a percepcióban értelmezhető affrikataként. Összességében tehát a /tɹ/-beli, /tʃ/-közeli /t/-realizáción át az /ɹ/ /s/-re gyakorolt hatása is feltételezhető az /stɹ/-kapcsolatokban (Shapiro, 1995). Továbbá fontos megjegyeznünk, hogy Smith és munkatársai (2019) megpróbálta kapcsolatba hozni a felpattanó zárhangok /ɹ/ hangok előtti affrikációját az /stɹ/ hangkapcsolatban megjelenő /ʃ/-közelítő realizációval. Eredményeik szerint viszont azok a beszélők, akik affrikáltan ejtették a /tɹ/ hangkapcsolatban a felpattanó zárhangot, azoknál nem volt valószínűbb, hogy a réshangot alacsonyabb CoG-val valósították volna meg /stɹ/-ben (Smith et al., 2019).

Mindez azt jelenti, hogy ebben a hangkapcsolatban tehát nem csak fonetikai jellemzők mentén kell értelmezni az /s/ realizációját és a két szibiláns összefüggéseit.

Ezen kutatások ugyanakkor egyrészt kimutattak nemek közötti eltéréseket is (ausztrál angol: Stevens & Harrington 2016), másrészt a kételemű, szibiláns + /t/ kapcsolatok elemzését is elindították.

Amerikai, ausztrál, skót angol és német nyelvben is kimutatták, hogy /st/ hangkapcsolatban a réshang CoG értéke alacsonyabb, mint amikor magánhangzó követte (Baker et al. 2011; Stevens et al. 2015; Stuart-Smith et al. 2019). Baker és kollégái (2011) akusztikai elemzésében két csoportra osztotta a beszélőket: akik a két szerző percepciók élménye alapján az /s/-t [ʃ]-ként képzik az /st.ɪ/-hangkapcsolatokban és akik nem. Az eredmények szerint minden /s/ + felpattanó zárhang esetében alacsonyabb volt az /s/ megvalósulásának CoG-értéke, mint az /s/ + magánhangzós kontextusban, míg az /ɪ/ jelenléte még alacsonyabb értékeket eredményezett mindhárom explozívával alkotott kapcsolatban. Az /st.ɪ/, /sp.ɪ/ és /sk.ɪ/ összevetésében pedig az /st.ɪ/ esetében (és így az összes kontextusban is ebben a kontextusban) kapták a legalacsonyabb CoG-t. Az [ʃ]-ejtőként kategorizált beszélők ejtésében az /s/ spektrális csúcsa jóval közelebb esett a magánhangzótól követett /ʃ/-éhoz az /st.ɪ/-hangkapcsolatokban, mint a #_V pozícióban, míg a másik csoportban változatos mintázatot kaptak, de az ő ejtésükben is minden esetben alacsonyabb volt a CoG a /t/-s kapcsolatokban, mint a #_V pozícióbeli /s/ esetében.

Ausztrál angolban is elemezték az /s/ realizációit felpattanó zárhangok (/p t k/) előtt, és hasonlóan a fenti eredményekhez, mindhárom explozíva alacsonyabb CoG-t eredményezett, a legalacsonyabbat, a /t/ - feltehetően az /ʃ/ irányába való eltolódás következtében. Az eltérések az intervokális és /s/C-kapcsolatok között azonban beszélőspecifikusak voltak (Stevens & Harrington 2016).

Skót angolban az /s/ kettős mássalhangzó-kapcsolatokban ugyanannyira alacsony CoG értékeket mutatott, mint hármás mássalhangzó-kapcsolatokban (Stuart-Smith et al. 2019). A nők ejtésében a réshang kevésbé közelített az /ʃ/-realizációhoz a férfiakéhoz képest különböző hangkörnyezetekben, kivéve az /st.ɪ/ mássalhangzó-kapcsolatot, amikor a nők CoG értékei jobban közelítettek a posztalveolárishoz, mint a férfiakéi (Stuart-Smith et al. 2019). Hozzá kell tennünk, hogy ausztrál angolban a nemek között nem találtak ilyen eltéréseket (Stevens & Harrington 2016).

1.5. A spiránssok akusztikai vizsgálatai a magyar beszédben

A nemzetközi szakirodalomhoz hasonlóan a magyar beszédben is eltérést találtak a zöngétlen alveoláris és a posztalveoláris réshang spektrális tulajdonságai között. Az /s/ réshang CoG értékei a várható módon minden vizsgálatban magasabb értékeket mutattak, mint az /ʃ/ réshang esetében (Beke & Gyarmathy, 2010; Bóna & Beke, 2013; Grácsi & Krepsz, 2018). A CoG mellett a spektrális ferdeség mutatott még életkortól, nemtől függetlenül eltérést a réshang képzéshelye szerint (Grácsi & Krepsz, 2018). A nőket és férfiakat együttesen vizsgáló elemzésben a spektrális csúcsosság is jelentős szerepet játszik az alveoláris és posztalveoláris réshangok elkülönítésében (Beke & Gyarmathy, 2010). A réshangok spektrális szerkezetére ugyanakkor a beszélő neme és életkora is hatással volt a magyar beszédben is, hasonlóan az angol nyelvben kapott eredményekhez (Bóna & Beke, 2013; Grácsi & Krepsz, 2018). Azt találták magyar spontán beszédben előforduló réshangok vizsgálatakor, hogy a CoG-érték magasabb nőknél, mint férfiaknál (Beke & Gyarmathy, 2010; Grácsi & Krepsz, 2018) hasonlóan a nemzetközi eredményekhez. Továbbá a spektrális szerkezet szóródása az alveoláris esetében a felnőttek ejtésében volt magasabb a tinédzserekéhez képest, míg a posztalveoláris szóródásértékei a tinédzserek esetében voltak magasabbak a felnőttekhez képest (Grácsi & Krepsz, 2018). Az alveoláris spiráns spektrális szerkezetének ferdeségére a nem oly módon volt hatással, hogy az értéke a férfiak-fiúk esetében a nullához közelített, míg a nők-lányok esetében negatív értékeket találtak. A posztalveoláris spiráns spektrális szerkezetének ferdeségére nem a nem, hanem az életkor volt hatással, a felnőttek esetében ugyanis magasabb értékeket mértek a tinédzserekéhez képest. A nemek még az alveoláris spiráns csúcsossága szempontjából mutattak eltérést, a női beszélők ugyanis centralizáltabban ejtették ezt a hangot, mint a férfiak (Grácsi & Krepsz, 2018). Az elemzések azt is megmutatták, hogy a spiránst követő magánhangzó minősége befolyásolja a spektrális szerkezetet. Az /s/ COG-je alacsonyabb volt követő palatális magánhangzó környezetében, mint veláris magánhangzó előtt, míg /ʃ/ esetében a követő palatális magánhangzó környezetében mértek magasabb CoG értéket (Beke & Gyarmathy, 2010). Az alveoláris és posztalveoláris

spiránsok időtartamában nem találtak lényeges eltérést intervokális környezetben (Beke & Gyarmathy, 2010; Grácz & Krepsz, 2018). Míg az egyik vizsgálat szerint a nem semmilyen hatással nem volt a réshangok időtartamára (Grácz & Krepsz, 2018), addig egy másik elemzés szerint a rövid, zöngétlen posztalveoláris spiránsokat a nők szignifikánsan hosszabban ejtették, mint a férfiak (Beke & Gyarmathy, 2010).

A magyar nyelvben az /st/ és /ft/ hangkapcsolat igen gyakori, előbbi a 3., utóbbi a 7. leggyakoribb mássalhangzó-kapcsolat egy kb. 2 millió szóból álló szövegkorpusz-vizsgálat szerint (Olaszy, 2007). Legjobb tudomásunk szerint azonban ezeknek az igen gyakori mássalhangzó-kapcsolatoknak spektrális akusztikai szerkezetét még nem vizsgálták. A mássalhangzó-kapcsolatok időtartamait viszont felolvasott szavakban és mondatokban részletesebben is vizsgálták. Az /st/ hangkapcsolat teljes időtartama átlagosan 197 ms, míg az /ft/ hangkapcsolaté 198 ms volt. Az alveoláris réshangot átlagosan 78 ms, míg a posztalveolárist átlagosan 81 ms hosszúnak mérték, tehát hasonló időtartamban realizálódtak (Olaszy, 2007).

1.6. Hipotézisek

A jelen tanulmány konkrét kérdései arra vonatkoznak, hogy a két zöngétlen szibiláns milyen akusztikai eltéréseket mutat intervokális és V_/t/ közötti pozícióban, és hogy ezekben tapasztalható-e eltérés a nemek között. A hipotéziseink az alábbiak voltak. (H1) Feltételeztük, hogy a két mássalhangzó a V_/t/ pozícióban az intervokális pozíciótól eltérő akusztikai mintázatot mutat, és hogy (H2) az /s/ és /f/ közötti akusztikai távolság csökken ebben a pozícióban. (H3) Feltételeztük, hogy a nők és a férfiak ejtése között nem csak a két spiráns CoG-értékében találhatunk eltérést, hanem a /t/-vel való koartikuláció eltérő mintázattal jelenik meg a két nem esetében. Bár ennek a tanulmánynak a központi témája a két szibiláns, a koartikulációs eredmények megértése végett a /t/ realizációinak képzési módját is górcső alá vettük. A /t/ realizációjában magyar nyelvben korábban azt találták, hogy C_V pozícióban gyakran fordul elő, hogy réshangként jelenik meg (beszélőnként a vizsgált szekvenciák 5,6–27,8%-

ában), illetve réses szivárgás van jelen a zárszakasz alatt (beszélőnként a vizsgált szekvenciák 0–36,4%-ában) (Neuberger & Grácz, 2013). Ez alapján feltételeztük, hogy (H4) a /t/ ejtésében a tipikus zár + felpattanás realizációktól eltérő képzési mód is megjelenik.

2. Kísérleti személyek, anyag, módszer

Kutatásunkhoz 72 középkorú magyar anyanyelvi beszélő (40 nő és 32 férfi) hanganyaga állt rendelkezésre automatikusan hangszinten felannotálva a BEA-adatbázisból (Neuberger et al., 2014). A kísérleti személyek életkora 20 és 46 év között mozgott, átlagosan 27,3 évesek voltak (szórás $\pm 7,1$ év). A mondat- és szövegfelolvasásokból kiválasztottunk 6-6 olyan szópárt, amelyekben az /f/ és /s/ hangok hasonló hangkörnyezetben fordultak elő. Mivel mind a megelőző, mind a követő hangzó hatással lehet a spiránsok akusztikai szerkezetére (Smorenburg & Heeren, 2020), ezért úgy választottuk meg a szavakat, hogy a közvetlen hangkörnyezet azonos legyen az /f/ és /s/ esetében. A vizsgált szópárok a következők voltak: *része* - *megbetegedéseket*, *tulipánágysók* - *gyöngyhalászosok*, *része* - *termékekétől*, *egyrészt* - *megbetegedést*, *esti* - *keresztül*, *estére* - *szilveszterkor*. A *része* szó kétszer fordult elő a felolvasandó szövegben, tehát minden beszélőnél két külön realizációt vizsgáltunk. Úgy választottuk ki a szópárokat, hogy a spiránsokat intervokális helyzetben és mássalhangzó-kapcsolatban is vizsgálni tudjuk. 3 esetben a réshangok intervokális (V_V), 3 esetben pedig V_/t/ hangkörnyezetben fordultak elő. Felmerül, hogy egy szópár esetében homoszillabikus, szóvégi mássalhangzó-kapcsolatot vizsgáltunk, míg a többi esetben heteroszillabikus, szóbelseji klasztereket. Egy esetben a szópár között eltér, hogy első szótagi, tehát szóhangsúlyt viselő /f/ áll szemben többedik szótagbeli /s/ ejtésével. Továbbá a szóvégi pozícióban az /f/ és a /t/ között morfémahatár húzódik, az /s/ és a /t/ között nem. Mindezek kontrollálására az adatbázis nem adott lehetőséget, ezért az értékek jellemző tartományait, eloszlását figyeltük meg. Ezek nem mutattak olyan eltérést, ami alapján a jelen vizsgálatot

befolyásolnák. Ugyanakkor a koartikulációs jellemzők további vizsgálata során ezeket a szempontokat célzott felvételekben, kontrolláltan tervezzük elemezni.

A spiránsok és az őket követő hangok határainak megkaptuk az automatikus annotációját (Varga et al., 2015), majd kézzel ellenőriztük az oszcillogram, a spektrogram és auditív információk alapján a Praat 6.1 szoftverrel (Boersma & Weenink, 2018). A spiránsok határait a rés indulásától a rés befejeztéig jelöltük.

A mássalhangzó-kapcsolatok többféleképpen realizálódtak. A megjelenő beszédhang főként a /t/ esetében volt változatos. 12 esetben azonban nem lehetett a két mássalhangzót elválasztani vagy nem jelent meg /t/-realizáció. A két beszédhang között a határt akkor húztuk meg, ha a spiránsra jellemző zörej intenzitása erőteljesen csökkent, így esetleg egy nem tökéletes zár vagy egy kevésbé intenzív réshang követte. Több esetben tehát nem volt zárszakasz. A szét nem választható ejtések az időtartambeli és spektrális elemzésekből kizártuk. A vizsgált mássalhangzó-kapcsolatok /t/-realizációinak elemzését ezekkel együtt ismertetjük. További 22 esetben a beszélők félreolvasták a célszavakat és/vagy szünetet tartottak a szóban, ezért ezeket a megvalósulásokat is figyelmen kívül hagytuk (akkor is, ha a célzott szóbeli pozícióval azonos hangkörnyezetben jelent meg bennük a célhang). Így a 864 lehetséges szóból 34-et zártunk ki félreolvasás vagy nem szegmentálható hanghatár miatt. Összesen 830 spiránsrealizációt vizsgáltunk meg. Az említett realizációk figyelmen kívül hagyása nem befolyásolta az elemzéseinket, mivel az általunk használt statisztikai módszer, a lineáris kevert modell kisszámú hiányzó adat esetén is megbízhatóan alkalmazható (Winter 2019).

A spiránsok spektrális méréseit egy erre a célra létrehozott Praat-szkripttel végeztük, amely a spektrális jellemzőket a mássalhangzó időtartamának 21 azonos időközönkénti pontján mérte annak 0–100% tartományában. A jelen tanulmányban a mássalhangzó középső 33%-án, azaz a 6–13. ponton mért adatok átlagát mutatjuk be. A méréseket 500 Hz–12000 Hz tartományban 0,005 ms ablaknagysággal 0,002 s-os tolással, 20 Hz-es frekvencialépéssel és Gauss-ablakkal hajtottuk végre. A szkripttel megmértük a spiránsok spektrális eloszlás súly-

pontját (Centre of Gravity: CoG), a spektrális szórását (SD), a spektrális ferdeségét és csúcosságát.

Kiszámítottuk a két mássalhangzó spektrális távolságát normalizált Euklidészi-távolsággal. A négy spektrális jellemzőt (súlypont, szórás, ferdeség, csúcosság) beszélőnként Z-transzformáltuk. Ezekre az értékekre számoltuk ki beszélőnként a két mássalhangzó távolságát a két kontextusban külön-külön.

A statisztikai elemzéseket a R 3.4.3 szoftverben (R Core Team, 2020) hajtottuk végre. A spiránsok időtartamának és spektrális szerkezetének elemzéséhez lineáris kevert modelleket állítottunk fel. Külön-külön modellben vizsgáltuk függő változóként a spiránsok spektrális mérőszámait (CoG, spektrális szórás, csúcosság és ferdeség) és időtartamát. A modellek fix hatása a mássalhangzó, a követő hangkörnyezet (CV vs. CC) és a beszélők neme voltak, míg a beszélőket és a szót (azaz itemet) tekintettük random hatásnak. A random hatásokra egyéni metszéspontokat (intercept) és a beszélőkre meredekséget (random slope) is számoltunk. A statisztikai modelleket az lme4 csomag (Bates et al., 2015) segítségével hoztuk létre, és a p -értékek meghatározásához az lmerTest csomagban (Kuznetsova et al., 2017) található Satterthwaite-approximációt használtuk. A modellfelállítás során a lehető legbővebb modellből indultunk ki, és szűkítettük addig, amíg a modell nem mutatott szignifikáns eltérést a legbővebb modelltől (hasonlóan Winter, 2019). Minden modell esetében a három faktor [mássalhangzó (/s/ vs. /ʃ/), kontextus (V_V vs. V_/t/), nem] interakciójától indultunk a kisebb modellek felé. Azt, hogy melyik változó esetében melyik modell bizonyult a legmegfelelőbbnek, az *Eredmények* fejezetben ismertetjük azok tárgyalásakor. A random meredekséget minden esetben a mássalhangzóra és a követő beszédhangra is állítottunk, de ez nem minden esetben volt megfelelő, mivel túlillesztést okozott. Ezért több esetben csökkenteni kellett a modell random hatásain, és csak a mássalhangzóra illesztett meredekséget megtartani. Ezeket az egyes méréseknél közöljük. A marginális és a kondicionális hatásnagyságot a MuMIn (Barton, 2020) és afex (Singmann et al., 2021) csomagok segítségével számoltunk.

Elemeztük a /t/-realizációk eloszlását a két mássalhangzó (/ʃ/ vs. /s/), a két beszélői nem (nő vs. férfi) és a szóbeli pozíció alapján (szóbelseji intervokális pozíció: pl. *szilveszterkor* vs. szóvégen áll, pl.: *egyrészt*). Ezt χ^2 -próbával végeztük.

A normalizált Euklidészi távolság esetében ismételt méréses ANOVA-t alkalmaztunk (Lawrence, 2016), mivel a Shapiro–Wilk-próba szerint normál eloszlásúnak tekinthető adatokat kaptunk.

3. Eredmények

3.1. A mássalhangzó-kapcsolatok /t/-realizációi

Bár a jelen tanulmány központi kérdése, hogy a két spiráns hogyan jellemezhető a /t/ előtti pozícióban az intervokális megvalósulásaihoz képest, ahhoz, hogy a spiránsok elemzését megtehessük, érdemes elemezni, hogy azok hogyan hatnak a /t/-re. Ezt azért érdemes elemezni, mert egyrészt világossá válik, hogy néhány esetben miért nem szegmentálható a két beszédhang, vagy miért nem kimondható egyértelműen a /t/ törlése, másrészt pedig a spiránsban található hangátmenet értelmezésében szerepet játszhat a követő fonéma realizációja. Statisztikai elemzést nem végeztünk a követő /t/ realizációjának függvényében, mivel gyakori, hogy az itt bemutatandó három főkategória közül csak kétféle fordul elő egy beszélő ejtésében.

A mássalhangzó-kapcsolatokban a két elemzett spiráns mindig réshangként valósult meg, azok esetében főleg az időtartam variabilitása szembetűnő. A /t/ ellenben gyakran nem tipikus zárszakasz + felpattanás jellegű megvalósulással jelent meg. A zárszakasz időtartama alatt jellemző volt a réshangból fennmaradó szivárgás, azaz a nem teljes zárképzés. Ennek arányát nehéz megállapítani, mert néhol intenzívebb, néhol kisebb intenzitású a fennmaradó réses lenyomat, ezért nem elemeztük. Ugyancsak több esetben kérdéses, hogy a kialakított kategóriák átmeneténél hol lehet meghúzni a határt. A /t/-realizációkat kategorizáltuk, de az átmeneti kategóriák miatt az I táblázatban úgy foglaltuk össze az eredményeket, hogy megadtunk főkategóriát, azaz azt, hogy milyen fő

képzésmód-csoporthoz tartozhat az adott megvalósulás, de alkategóriákat is különk, amelyek alapján az átmeneti jellegű megvalósulások is láthatóak. Fő kategóriaként **felpattanó zárhang**, **affrikáta** és **réshang** megvalósulást vettünk fel, továbbá ebben az elemzésben az arányok bemutatása végett még közöljük azokat a realizációkat, amelyeknél nem tudtuk elkülöníteni a két beszédhangot, tehát a /t/ ejtése elmaradt vagy annyira hasonult a megelőző réshanghoz, hogy nem voltak külön szegmentálhatóak. Egy ilyen, nem szegmentálható mintát mutat be az 1. ábra első regisztrátuma.

Az explozív típusú realizációkhoz azokat soroltuk, amelyben elkülöníthető volt a zárszakasz és az intervokális helyzetű /t/-megvalósulásra emlékeztető, rövid felnyílás. Egy esetben nem volt felnyílás, ezt ugyanide soroltuk, mivel affrikátára nem jellemző a felnyílás elmaradása az akusztikai regisztrátumon. Az 2. ábra 2. regisztrátuma tipikus felpattanó zárhang jellegű realizációra mutat példát. Előfordultak olyan realizációk, amelyekben a zárszakasz ideje alatt annyira intenzív volt a szivárgás, de felpattanás követte, hogy felmerült, hogy esetleg inkább gyenge intenzitású réshangként is lehet őket kategorizálni, amelyet felpattanás követ. Erre mutat példát az 3. ábra 3. lenyomata. Kevés esetben a zárszakasz jól detektálható volt, de több felpattanásszerű esemény történt alatta. A felpattanó zárhangok esetében gyakran előfordul, hogy a zárfelnyílás több felpattanással jellemezhető (magyarra vö. Grácz & Kohári 2014), de a jelen esetben hosszabb felpattanás nélküli „kimaradások” tagolták ezeket szét, nem egy felnyílási folyamat részének voltak tekinthetők.

Affrikált realizációnak azt tekintettük, amely esetében ugyan a zárszakasz detektálható, de a felnyílás hosszabb időtartamú (4. ábra 4. lenyomat), esetleg főként réssel (5. ábra 5. lenyomat) jellemezhető. Ezekben a realizációkban is fordult elő (ritkán) felpattanás a zár alatt.

Érdemes megjegyezni, hogy a zár alatti felpattanások megjelenésekor nem találtunk réses szivárgást egy zárhangtípusban sem.

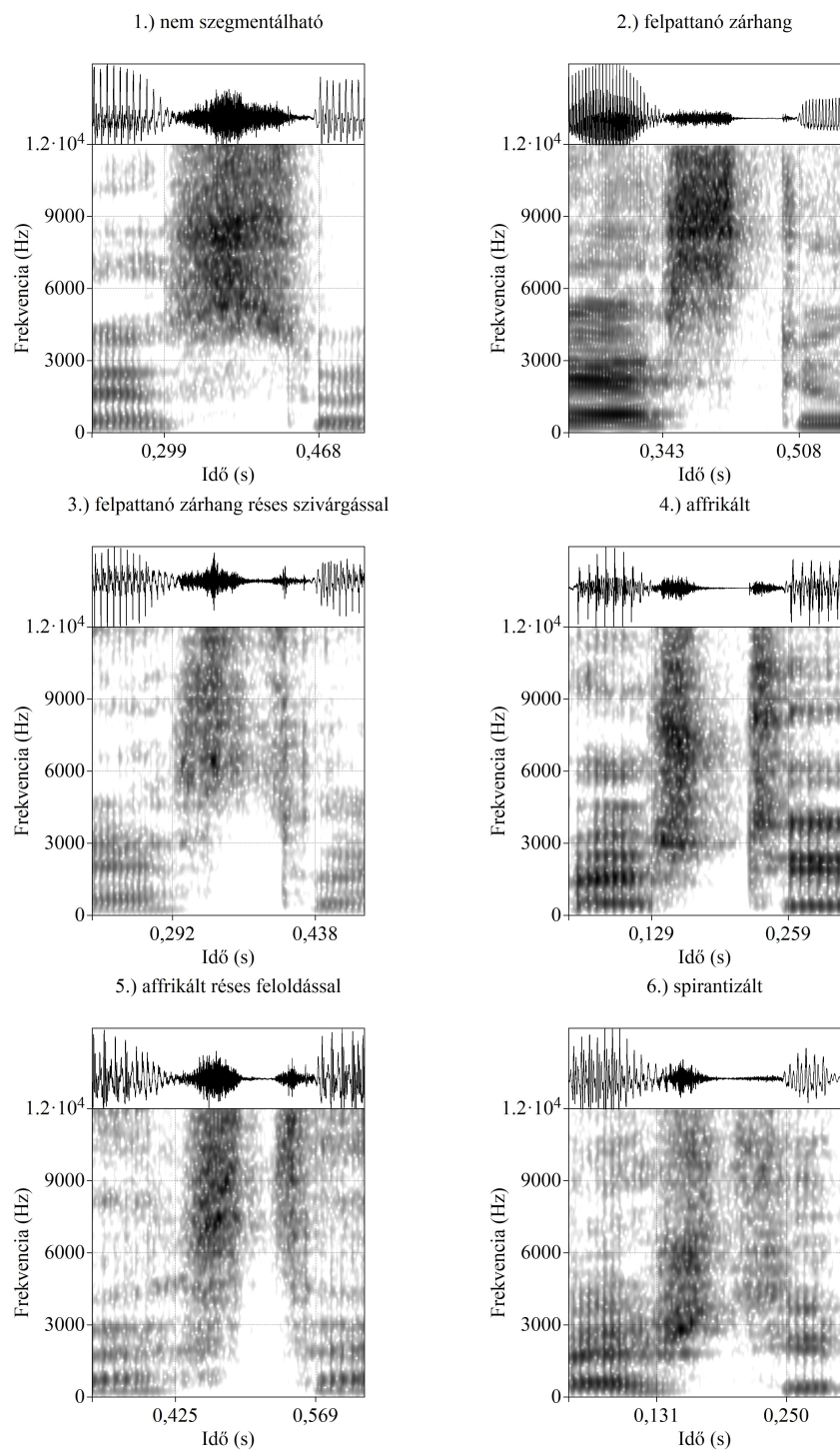
A harmadik főkategória a réshangszerű /t/-realizáció volt (6. ábra 6. regisztrátum). Ebben az esetben a réses elem és a /t/ előtti spiráns között inkább hangátmenet volt felfedezhető, nem zárra utaló lenyomat. Kevés esetben elő-

fordult a hangátmenet intenzitásának oly mértékű csökkenése, hogy kérdéses volt, hogy zárképzés jött létre erős szivárgással vagy réshang. A fenti hasonló átmeneti kategóriától a zárfelnyílás elkülöníthetősége alapján külön csoportosítottuk. Azaz ide azokat a megvalósulásokat soroltuk, amelyek esetében nem volt zárfelnyílás tapasztalható. Ugyanakkor megjegyzendő, hogy az intenzitáscsökkenés nyilvánvalóan a zárképzésre törekvés következtében jelenik meg, és hogy az összes bemutatott realizáció leginkább az explozíva–réshang skálán meghúzott határok mentén értelmezhető, mintsem egzakt kategóriákként.

A leggyakoribb realizációk explozívaként voltak besorolhatóak (több mint az esetek felében mindkét spiráns utáni pozícióban), valamivel nagyobb arányban jelentek meg az /ʃ/ után, mint az /s/ után. Az esetek közel negyedében jelent meg affrikált /t/-ejtés. A réshangszerű ejtés gyakoribb volt az /s/ után, az esetek ötödét tette ki, míg az /ʃ/ után csak a tizedét. A mássalhangzók között nem találtunk szignifikáns eltérést a /t/-realizációk gyakoriságában ($\chi^2(3) = 4,512$, $p = 0,211$) (2. ábra).

A nők ejtésében gyakoribb volt, hogy explozívaként realizálódott a /t/ (60,76%), mint a férfiakéban (43,09%). A férfiak mindkét alternatív képzésmódot jellemzően gyakrabban ejtették (affrikáta: 30,32%, réshang: 22,87%), mint a nők (affrikáta: 23,63%, réshang: 13,50%). A nem elkülöníthető /t/-realizációk hasonló arányban jelentek meg mindkét nem ejtésében (nők: 2,11%, férfiak: 3,72%). A nemek között ez az eltérés szignifikánsnak bizonyult. ($\chi^2(3) = 14,134$, $p = 0,003$) (2. ábra).

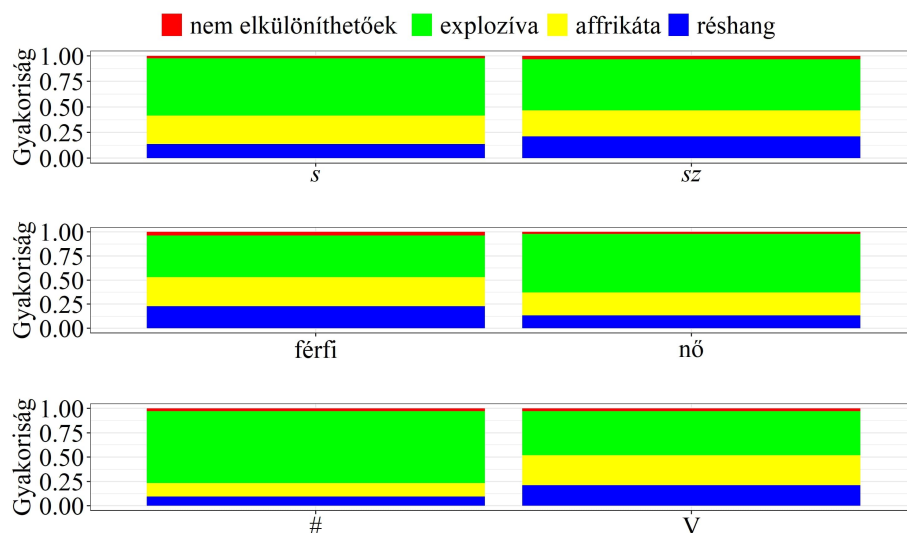
Mivel két szó esetében a spiráns + /t/ szóvégi pozícióban helyezkedik el, négy esetben intervokális pozícióban, elemeztük, hogy ennek mentén eltér-e az explozíva realizációja. Az eredmények szerint igen ($\chi^2(6) = 38,736$, $p < 0,001$). Míg szóvégi pozícióban a /t/ leggyakrabban a prototipikus explozívarealizációt mutatta (73,72%), affrikataként (13,97%) vagy réshangként (9,49%) összesen az esetek negyedében jelent meg, addig magánhangzó előtt explozívaként 45,37%-ban, affrikataként 30,56%-ban, réshangként 21,30%-ban jelent meg. A két hang elkülöníthetősége nem tért el, 3% alatt volt mindkét pozícióban.



1. ábra. A /t/-realizációk főbb típusai

1. táblázat. A /t/-realizációk típusainak gyakorisága

főkategória	alkategória	N	/ʃ/ arány%	N	/s/ arány%
nem elkülöníthetők		5	2.39	7	3.24
felpattanó zárhang	felpattanó zárhang	114	54.55	101	46.76
	zár + 0 felpattanás	1	0.48	1	0.46
	felpattanó zárhang zár nélkül	1	0.48	6	2.78
	felpattanó zárhang felpattanásokkal a zár alatt	1	0.48	0	0
	összesen	117	55,98	108	50,00
	zár + réses felnyílás	10	4.78	20	9.26
affrikáta	zár + hosszabb (affrikátához közelibb) felnyílás	47	22.49	33	15.28
	zárszakasz, felpattanás a zár alatt + réses felnyílás	1	0.48	1	0.46
	zárszakasz felpattanásokkal + hosszabb (affrikátához közelibb) felnyílás	0	0	1	0.46
	összesen	58	27,75	55	25,46
	réshang	26	12.44	41	18.98
réshang	réshang vagy réses felnyílás zár nélkül vagy erős szivárgással	2	0.96	5	2.31
	réshang + felpattanás	1	0.48	0	0
	összesen	29	13,88	46	21,30
	összesen	209	100	216	100



2. ábra. A /t/-realizációk fő képzési módjának gyakorisága a két réshang után (fent), a nemek között (középen) és a szóbeli pozíció függvényében (lent)

Összefoglalva tehát három fő kategóriát állítottunk fel a szibiláns + /t/ kapcsolatokban a /t/ realizációira: explozíva, affrikáta, spiráns. Ezek megjelenésének eloszlása a mássalhangzók között nem, de a két nem és a fonémakapcsolat szóbelseji vs. szóvégi pozíciója mentén szignifikáns eltérést találtunk.

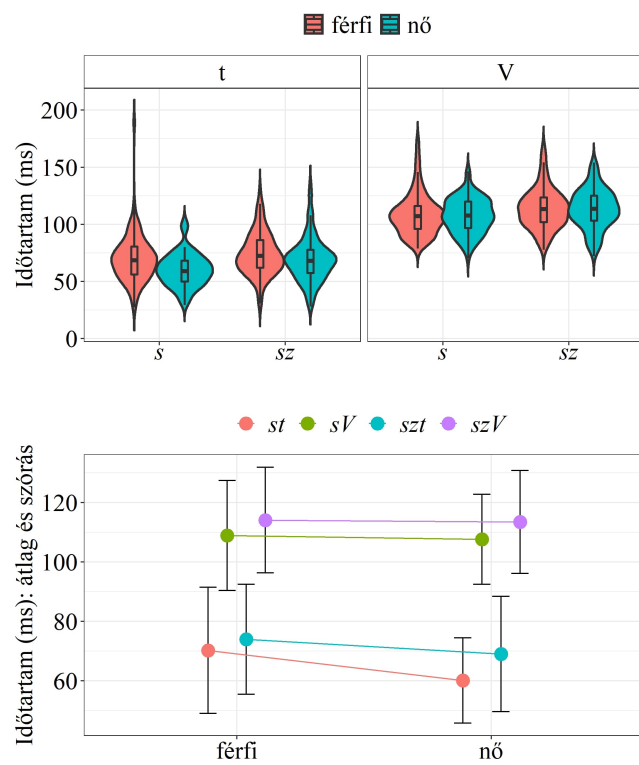
3.2. Időtartam

Az /s/ és az /ʃ/ időtartamának eltérését a [3](#) ábra szemlélteti. A felső panelen hegedűábrák és dobozdiagramok láthatóak. A hegedűábrák elfordított sűrűségfüggvényként szolgálnak, azaz minél szélesebbek egy tartományban, annál több adat szerepel ott. A dobozdiagram pedig a mediánt és a kvartiliseket szemlélteti. Az alsó panelen átlag és szórás látható. Az átlagokat és a szórásokat a [2](#) táblázat tartalmazza minden mért változóra.

Az időtartamra illesztett modellek közül az bizonyult az adatokat legjobban jellemzőnek, melyben a három vizsgált független változó (mássalhangzó,

hangkörnyezet, nem) is szerepelt faktorként, és csak random konstansokat tartalmazott. A hangkörnyezet mint főhatás szignifikánsan hatott az időtartamra ($F(1, 830) = 193,868, p < 0,001$), a /t/ előtti környezetben rövidebb volt mindkét réshang, mint intervokális pozícióban. A hangkörnyezet és a nem interakciója bírt még szignifikáns hatással a mássalhangzók időtartamára ($F(1, 830) = 12,511, p < 0,001$). Míg intervokális helyzetben nem volt eltérés a mássalhangzók időtartamában a nők és a férfiak között, addig a /t/ előtti helyzetben a nők ejtésében rövidebb volt az /ʃ/, mint a férfiakéban. A nők és a férfiak esetében a két mássalhangzó közötti eltérés hasonlóan alakult intervokális helyzetben, és a V_/t/ környezetben a nők ejtésében, míg a férfiak esetében ebben a CC-pozícióban nem találtunk eltérést a két réshang időtartama között. A modell marginális hatásnagysága 0,607, a kondicionális 0,764, azaz a fix hatások is nagymértékben magyarázzák az adatokat, a random hatás (random konstans a beszélőre) pedig tovább javítja.

A kiugró értékek az /ʃ/ ejtésében hiperartikulált ejtések voltak, de kizárásuk nem változtat lényegesen a modelleken sem az időtartam, sem a további elemzések tekintetében.



3. ábra. A frikatívák időtartama (ms) a követő beszédhang és a nem függvényében. Fent: teljes tartomány, lent: átlag és szórás

2. táblázat

2.a táblázat. Az /ʃ/ akusztikai jellemzőinek átlaga és szórása a beszélő nemének és a hangkörnyezetnek a függvényében (s. = spektrális, lognorm. = lognormált)

C	/ʃ/			
nem	férfi		nő	
követő hang	/t/	V	/t/	V
összes adat	90	88	114	113
időtartam (ms)	70.25	108.93	60.11	107.68
	±	±	±	±
	21.27	18.53	14.34	15.17
CoG (Hz)	3974.07	3915.74	4520.75	4612.86
	±	±	±	±
	735.85	700.52	726.33	785.74
S. szórás (Hz)	1766.53	1647.47	1701.23	1662.07
	±	±	±	±
	382.89	355.50	327.79	426.87
S. ferdeség	1.65	1.84	1.69	1.68
	±	±	±	±
	0.91	1.11	0.89	1.11
S. csúcsosság	6.58	7.96	7.63	8.42
	±	±	±	±
	6.89	10.38	6.88	12.58
S. csúcsosság lognorm.	1.42	1.61	1.65	1.48
	±	±	±	±
	1.14	1.15	1.12	1.47

2.b táblázat. Az /s/ akusztikai jellemzőinek átlaga és szórása a beszélő nemének és a hangkörnyezetnek a függvényében (s. = spektrális, lognorm. = lognormált)

C	/s/			
nem	férfi		nő	
követő hang	/t/	V	/t/	V
összes adat	91	96	118	120
időtartam (ms)	73.95	114.13	69.01	113.52
	±	±	±	±
	18.52	17.81	19.39	17.32
CoG (Hz)	6677.47	6755.37	8386.82	8915.37
	±	±	±	±
	926.22	991.21	1183.72	1323.30
S. szórás (Hz)	2298.03	2133.33	2367.78	2197.15
	±	±	±	±
	509.11	398.03	604.31	519.11
S. ferdeség	0.64	0.87	−0.35	−0.31
	±	±	±	±
	0.71	0.77	0.91	1.02
S. csúcsosság	2.91	3.23	3.70	4.02
	±	±	±	±
	2.64	3.21	5.42	5.37
S. csúcsosság lognorm.	0.65	0.78	0.79	0.93
	±	±	±	±
	1.11	1.00	1.15	1.01

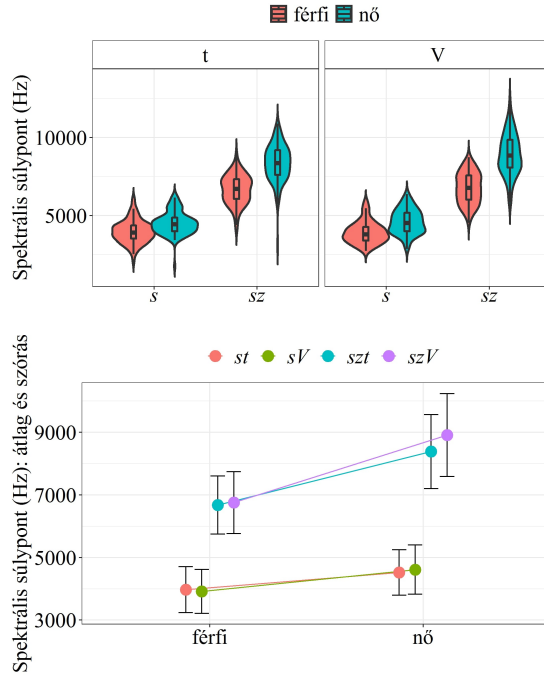
3.3. Spektrális súlypont (CoG)

A CoG jellemzőit a 4. ábra és a 2. táblázat mutatja. A kevert modellek közül az bizonyult az adatokat legjobban leírónak, melyben mindhárom független változó szerepelt, és a mássalhangzóra és a hangkörnyezetre random meredekséget tartalmazott. Ennek eredményeit a 3. táblázat mutatja be. Az eredményekre a főhatások közül a mássalhangzó és a beszélő neme, az interakciók közül a mássalhangzó és a hangkörnyezet is a nemmel való interakcióban bizonyult szignifikánsan meghatározónak. A hármas interakció tendenciaszintre utaló p-értéket ért el. Az /s/ minden esetben magasabb CoG-vel valósult meg, mint az /f/. Az /s/ intervokális környezetben magasabb CoG-vel realizálódott a nők ejtésében, mint /t/-s környezetben, míg a férfiak ejtésében nem találtunk ilyen összefüggést egyik zöngétlen spiráns esetében sem. A nők ejtésében mindkét réshang mindkét környezetben magasabb spektrális súlyponttal valósult meg, mint a férfiakéban, de ez az eltérés nagyobb volt az /s/ ejtésében.

A marginális R^2 0,798, a kondicionális 0,919 volt, ami azt jelenti, hogy a fix hatások jól magyarázzák az adatokat, és a random hatások (beszélőnkénti random meredekség a mássalhangzóra és a hangkörnyezetre, random konstans az itemekre) ezt tovább erősítik.

3. táblázat. A spektrális súlypontra kapott statisztikai eredmények. (Szürke cellák jelölik a szignifikáns hatásokat.)

faktor	$F(1, 830)$	p
mássalhangzó	587,434	<0,001
beszélő neme	78,1236	<0,001
hangkörnyezet	2,4375	0,149
mássalhangzó * beszélő neme	37,024	<0,001
mássalhangzó * hangkörnyezet	1,862	0,210
hangkörnyezet * beszélő neme	6,198	0,015
mássalhangzó * hangkörnyezet * beszélő neme	3,645	0,0567

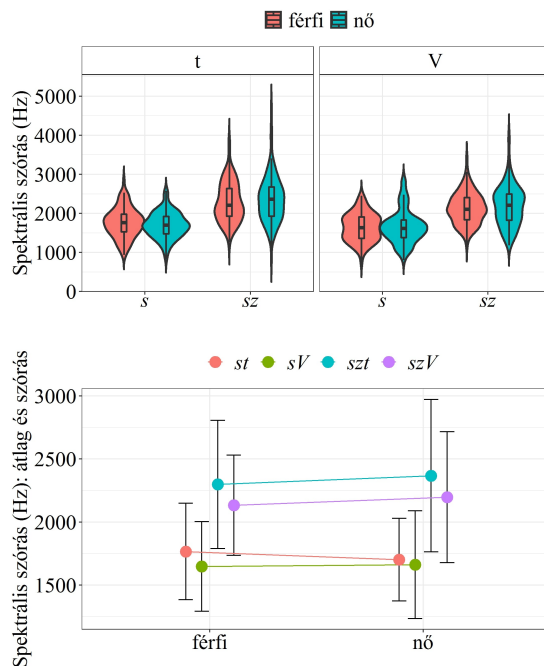


4. ábra. A spektrális súlypont (Hz) a vizsgált spiránsokban a követő beszédhang és a nem függvényében. Fent: teljes tartomány, lent: átlag és szórás

3.4. Spektrális szórás

A spektrális szórás esetében az a modell bizonyult az adatokat legjobban leírónak, melyben a nem nem szerepelt, csak a mássalhangzó és a hangkörnyezet, továbbá ezekre random meredekséget is tartalmazott. A spektrális szórás (5. ábra) esetében azt találtuk, hogy a mássalhangzó ($F(1, 830) = 71,398$, $p < 0,001$) és a követő hang ($F(1, 830) = 7,0107$, $p < 0,021$) főhatások gyakoroltak szignifikáns hatást az eredményekre. Az /s/ a hangkörnyezettől és a beszélő nemétől függetlenül magasabb tartományban realizálódott, mint az /ʃ/. A férfiak ejtésében /t/-s környezet mindkét mássalhangzóban magasabb értéket eredményezett, mint az intervokális pozíció, a nők ejtésében csak az /s/-re kaptunk ilyen eltérést. A marginális R^2 0,284, a kondicionális 0,605 volt. Ez azt jelenti, hogy a fix hatások mellett további meghatározó tényezők is hatnak a spiránsok spektrális szórására, de ezek is hozzájárulnak annak értékéhez, emellett

pedig a mássalhangzókra illesztett beszélőnkénti random meredekség nagyban hozzájárul a modell hatékonyságához.



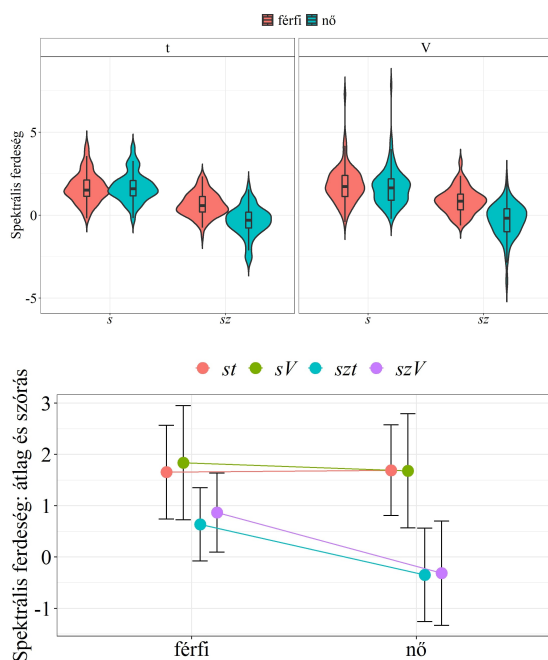
5. ábra. A spektrális szórás (Hz) a vizsgált spiránsokban a követő beszédhang és a nem függvényében. Fent: teljes tartomány, lent: átlag és szórás

3.5. Spektrális ferdeség

A spektrális ferdeségre [6] ábra) illesztett modellek közül az adatokat az írta le a legjobban, amelyben mindhárom független változó (mássalhangzó, nem, kontextus) szerepelt, és a mássalhangzóra illesztett random meredekséget tartalmazott. A mássalhangzó ($F(1,830) = 136,835$, $p < 0,001$) és nem ($F(1,830) = 20,714$, $p < 0,001$) főhatások és ezek interakciója ($F(1,830) = 21,552$, $p < 0,001$) esetében kaptuk azt az eredményt, hogy szignifikánsan határozzák meg az eredményeket. Míg a vizsgált hangkörnyezetek között nem tértek el a két mássalhangzóra kapott eredmények, az /s/ spektrális ferdesége

mindkét pozícióban és mindkét nem esetében alacsonyabb volt, mint az /j/-é. A nők ejtésében ez az eltérés nagyobb mértékű volt, mint a férfiak ejtésében.

A marginális R^2 0,465, a kondicionális 0,710 volt, azaz a vizsgált fix hatások nagy magyarázó erővel bírtak, a random hatás (beszélőnkénti random meredekség a mássalhangzóra és itemenkénti random konstans) pedig tovább erősítette a modellt.



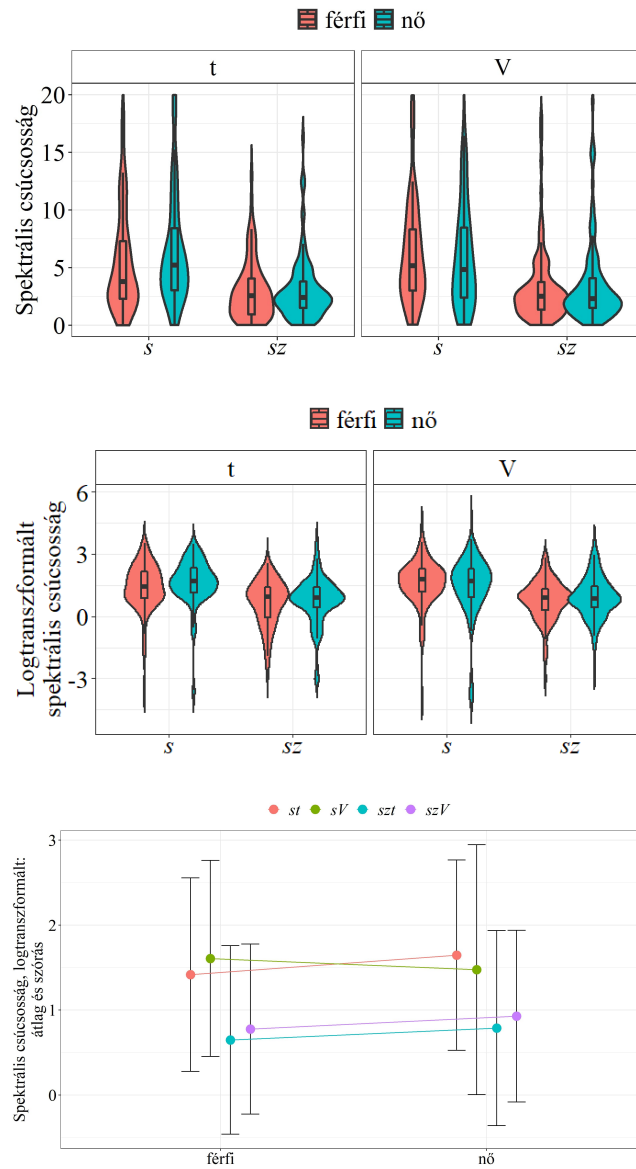
6. ábra. A spektrális ferdeség a vizsgált spiránsokban a követő beszédhang és a nem függvényében. Fent: teljes tartomány, lent: átlag és szórás

3.6. Spektrális csúcsosság

A spektrális csúcsosság esetében az a modell írta le az eredményeket a legjobban, amelyikben a nem nem, csak a mássalhangzó és a követő hangzó szerepelt fix hatásként, és csak a mássalhangzóra tartalmazott random meredekséget. A 7. ábrán két hegedű-dobozábra látható. Az első a kinyert adatokat, a második a logtranszformált értékeiket mutatja. Az átlag-szórás ábrát csak a

logtranszformált adatokra közöljük. A logaritmus alapú transzformációra azért volt szükség, mert az adatok sűrűsége nagymértékben emelkedett a pozitív értékektől 0 felé közeledve. A transzformáció előtt ellenőriztük a magas értékeket, a 20 és 40 feletti adatokat. Ezek esetében egy-egy frekvenciatartomány a szibilánsokra jellemző, erős intenzitású sávnál is intenzívebbnek bizonyult, de a legnagyobb érték esetében nem volt ilyen jellemző felfedezhető. Az adatok értelmezéséhez fontos, hogy készítettünk modellt a nem transzformált adatokra úgy, hogy minden értéket megtartottunk, úgy, hogy a 40 felettieket, majd a 20 felettieket kizártuk. Emellett elkészült logtranszformált adatokra is a modell. Egyik esetben sem találtunk eltérést a fix faktorok és azok interakcióinak hatásában, a p-értékek minimálisan változtak, de az érdemi változást nem okozott. Emellett a hatásnagyságok elemzésében sem értünk el jelentősen jobb értékeket egyik módon sem. A [2](#) táblázatban mindkét adatsor átlagát és szórását közöljük, mert fontos, hogy bár a transzformáció eredményez negatív értékeket, de a ténylegesen mért csúcsosság adatai között nincsenek negatív értékek, ami pedig a csúcsosságról fontos információ, hogy a Gauss-normál görbéhez képest csúcsosabb irányban jelennek csak meg adatok.

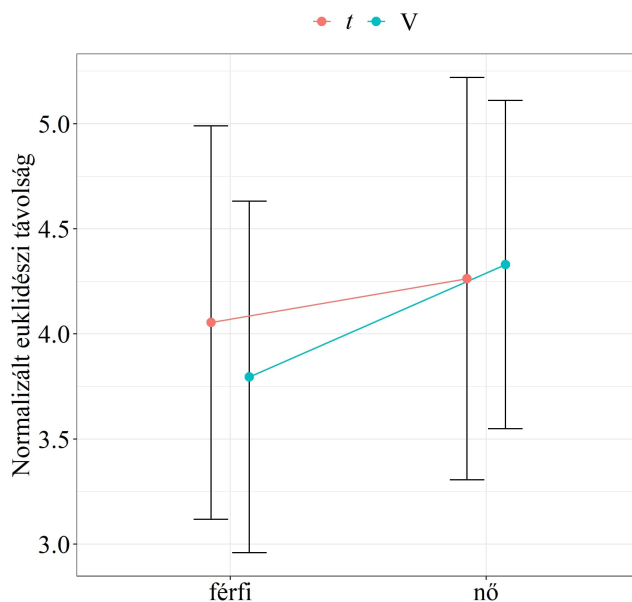
A logtranszormált adatokra készült modell eredményeit ismertetve tehát a mássalhangzó mint főhatás bizonyult szignifikánsan meghatározónak ($F(1, 830) = 19,624$, $p < 0,001$), az R^2 -értékek pedig 0,075 és 0,291 voltak. Vagyis más, nem vizsgált jelentős tényezők is hatnak a csúcsosság értékeire. Bár minimális eltérések látszanak a követő beszédhang és a beszélő neme mentén, de az adatok szóródása nagy, így ezek nem jelentenek eltérést. Az /ʃ/ spektrális csúcsossága mindkét hangkörnyezetben és mindkét nem esetében magasabb volt az /s/-énél.



7. ábra. A spektrális csúcosság a vizsgált spiránsokban a követő beszédhang és a nem függvényében. Fent: teljes tartomány, középen: teljes tartomány logtranszformált eloszlása, lent: átlag és szórás a logtranszformált értékekben

3.7. Euklidészi távolság

A két mássalhangzó normalizált euklidészi távolságát is összevetettük (8. ábra). Ebben az időtartamot nem, csak a négy spektrális jellemzőt használtuk fel. Mind a négy jellemző szignifikánsan eltért a két konszonáns között, a csúcosság kivételével pedig a két kontextus között, mint főhatás között vagy legalább a nemmel való interakcióban mutattak eltérést. A két kontextus között számított távolság valamivel alacsonyabb volt a férfiak ejtésében, és míg a nők beszédében nem találtunk eltérést a két kontextus között, a férfiak esetében az intervokális kontextushoz képest nagyobb eltérést kaptunk a /t/-s környezetben. Az ismételt mintás ANOVA eredménye szerint a két nem között ($F(1, 62) = 0,558$, $p = 0,005$) és a nem és a kontextus interakciójában találtunk szignifikáns eltérést ($F(1, 62) = 0,832$, $p = 0,008$).



8. ábra. Az /s/ és /f/ normalizált euklidészi távolsága a két kontextus és a két nem között

4. Következtetések

A zöngétlen alveoláris és posztalveoláris spiránst vetettük össze intervokális és V_/t/ pozícióban. Egyrészt a réshangok /t/ realizációjára gyakorolt hatását, másrészt a spiránsok időtartamát és spektrális szerkezetét elemeztük. A réshangok koartikulációs agresszióját, azaz a környezetében lévő hangra gyakorolt hatását elemeztük a /t/ különböző realizációinak feltárásával, illetve a koartikulációs rezisztenciáját, ellenállását is megvizsgáltuk a réshangok spektrális szerkezetének mérésével.

A két mássalhangzó utáni /t/-realizációk eloszlása nem tért el szignifikánsan annak függvényében, hogy melyik mássalhangzó előzte meg, de a két nem és a /t/ pozíciója (szóvégi vagy szóbelseji és magánhangzó követi) eltért. A férfiak esetében, illetve a szóbeli helyzetben gyakrabban fordul elő, hogy spirantizált vagy affrikátaközeli realizációval jelent meg.

A két frikatíva között szignifikáns különbséget találtunk a négy spektrális mérőszámban, de az időtartamban nem (hasonlóan a spontán beszédre kapott eredményekhez: Beke & Gyarmathy, 2010; Grácsi & Krepsz, 2018). Az /s/ spektrális súlypontja, szórása és ferdesége magasabb, csúcsossága pedig alacsonyabb volt. A csúcsosság esetében azonban fontos kiemelni, hogy valójában a két mássalhangzóra jellemző tartomány tért el, az /f/ magasabb értékeket is gyakrabban vett fel.

A két elemzett spiráns realizációi esetében korábban azt találták magyar beszédben (Beke & Gyarmathy, 2010; Grácsi & Krepsz, 2018), hogy a spektrális súlypont és a spektrális szórás értéke az /s/ esetében magasabb posztalveoláris párjához képest. A spektrális ferdeség és csúcsosság esetében pedig eltérő eredményeket kaptak. A jelen vizsgálatban a spektrális ferdeséget és csúcsosságot az /f/-ben mértük magasabbnak, a hangkörnyezettől és a beszélő nemétől függetlenül, a spektrális súlypont és a spektrális szórás pedig az /s/-ben volt magasabb ugyancsak a másik két tényezőtől függetlenül. Figyelembe kell venni azt is, hogy a nemzetközi szakirodalomban is találtak ellentmondásos eredményeket a CoG

kivételével. Mindezen eltérő eredmények feltárásához célzott vizsgálatokra van szükség.

A kontextus, hogy intervokális vagy V_/t/ pozícióban állt a vizsgált réshang, önmagában az időtartamra és a szórásra gyakorolt szignifikáns hatást. Az időtartam az intervokális, a szórás a /t/-s környezetben volt magasabb.

A két nem mint főhatás között a spektrális súlypont és ferdeség mutatott szignifikáns eltérést. A spektrális súlypont esetében a korábbi eredményekhez hasonlóan a nők értékei voltak magasabbak, a ferdeség esetében a férfiaké. Az elemzéseinkhez hasonlóan a korábbi vizsgálatok is kimutattak nemek közti különbségeket a spektrális súlypont és a ferdeség értékeiben. Ezekben a vizsgálatokban a CoG és a spektrális szórás értéke magasabb nőknél, mint férfiaknál (Jongman et al., 2000; Koenig et al., 2013; Perkell et al., 2004), a spektrális ferdeség alacsonyabb, a csúcsosság pedig magasabb volt a nők ejtésében, mint a férfiakéban (Jongman et al., 2000; Fuchs & Toda 2010). Habár a mi vizsgálatunkban a csúcsosság nem mutatott eltérést a férfiak és nők ejtésében.

Mindezeket az eredményeket több esetben a nem és a mássalhangzó közötti vagy a nem és a kontextus közötti interakció szignifikáns volta árnyalta tovább. A spektrális súlypont esetében mindkét interakció szignifikáns eredményt adott. Ennek oka, hogy az alveoláris konsonáns esetében nagyobb arányú eltérést kaptunk az értékekben a két nem között, illetve, hogy a /t/-s környezet kis mértékben csökkentette ezt az eltérést. Emellett az /s/ esetében nagyobb eltérést kaptunk a nők és a férfiak ejtése között, mint az /j/ esetében. A *Bevezetőben* leírt irodalmi eredmények szerint nem kizárólag fiziológiai oka van az eltérésnek, hanem az artikulációban is eltérő mintázat feltételezhető. Rácz & Schepács (2013) laminálisabb ejtést feltételez a nők ejtésében, ezért percepcióss tesztükben férfi ejtésben a nem laminális és laminális ejtéssel rögzítettek /s/ és /j/ réshangokat tartalmazó felolvasást. Akusztikai eltérést az /j/ esetében találtak e két ejtés között (egy férfi beszélő ejtésében). A percepcióss tesztben a laminális szibilánsokat tartalmazó felolvasást jelentős mértékben femininebbnek tekintették az adatközlők –, ugyanakkor Icht & Boaz (2017) héberben nem talált szignifikáns eltérést a laminális és az apikális ejtésű alveoláris réshangok között.

A valós eltérés feltárásához tehát artikulációs vizsgálatok járulhatnak hozzá. Az artikuláció esetleges eltérése a koartikulációs mintázatok közötti eltérésre is magyarázatot adhat.

Az időtartam esetében a férfiak ejtésében az /ʃ/ /t/-s környezetben nem rövidült szignifikáns mértékben, míg a nőknél mindkét mássalhangzó igen, illetve a férfiak ejtésében az /s/ is.

A spektrális ferdeség esetében a két nem között ismét az /s/ ejtésében mutatkozott nagyobb eltérés, mint az /ʃ/ ejtésében.

Az első hipotézisünk szerint (H1) feltételeztük, hogy a két mássalhangzó a V_/t/ pozícióban az intervokális pozíciótól eltérő akusztikai mintázatot mutat. Ez az időtartamban, a spektrális súlypont és a spektrális ferdeség értékében mutatkozott meg. Az időtartam és a spektrális súlypont esetében a nemek között eltért. Az amerikai, ausztrál, skót angol és német nyelvben /st/ és egyéb /s/ + felpattanó zárhang hangkapcsolatban a réshang CoG értéke ugyancsak alacsonyabb volt, mint amikor magánhangzó követte (Baker et al., 2011; Stevens et al., 2015; Stuart-Smith et al., 2019) hasonlóan a vizsgálatunkban talált magyar beszédbeli eredményekhez. A CC-kapcsolatban ugyanakkor beszélőspecifikusnak találták, hogy megjelenik-e az eltérő ejtémintázat az intervokális helyzethez képest, vagy sem (Stevens & Harrington, 2016).

A második hipotézisünk szerint (H2) az /s/ és /ʃ/ közötti akusztikai távolság csökken a /t/-s pozícióban. Ez a hipotézisünk nem nyert alátámasztást. A négy spektrális mérőszám alapján kapott normalizált euklidészi távolság a nők ejtésében nem mutatott eltérést a kontextusok között, a férfiak ejtésében pedig a /t/-s környezetben kaptunk magasabb értéket.

(H3) Feltételeztük, hogy a nők és a férfiak ejtése között nem csak a két spiráns CoG-értékében találhatunk eltérést, hanem a /t/-vel való koartikuláció eltérő mintázattal jelenik meg a két nem esetében. A két nem között a spektrális súlypont és a ferdeség esetében találtunk eltérést. A koartikulációs mintázatokban pedig egyrészt az időtartam, a spektrális súlypont és ferdeség, valamint a normalizált euklidészi távolság esetében is eltérést kaptunk.

Negyedik hipotézisünkben feltételeztük, hogy (H4) a /t/ ejtésében a tipikus zár + felpattanás realizációktól eltérő képzési mód is megjelenik. Az eredmények azt mutatták, hogy az esetek közel felében jelentkezik affrikátaszerű vagy spirantizált ejtés, esetleg nem szegmentálható a réshang és a /t/. Érdekes figyelembe venni, hogy ez utóbbi esetben a /t/-törlését is jelentheti a realizáció. A további elemzésekben ezeket a nem szegmentálható eseteket nem vettük figyelembe, mert nem határoltuk el, hogy /t/-törlés vagy ennyire összeolvadt ejtés jelent-e meg, másrészt ha törlődik is a /t/, nem evidens, hogy milyen koartikulációs csoportba kell sorolni a réshang realizációját. Fontos továbbá megjegyezni, hogy a tipikus explozív-, affrikata- és spiránsrealizációk helyett gyakran kevésbé kategoriálisan, inkább skálaszerűen képzelhetők el ezek a megvalósulások, tehát a prototipikus lenyomatok mellett több határeset is előfordult. A /t/ nem prototipikus explozív realizációinak és ezek skálaszerű megjelenésének az oka feltehetően a nem teljes zárképzés, mivel a zárszakasszal megjelenő beszédhangok ideje alatt is gyakorta gyengébb-intenzívebb réses szivárgás volt tapasztalható. Hasonló okok feltételezhetők, mint a Bevezetőben ismertetett angol /tɪ/, /dɪ/ esetében tapasztalható, a két képzési mód közötti nem teljes váltás (Magloughlin, 2018). Az alveoláris explozív + /ɹ/ esetében azt találták, nyelv elülső és hátulsó területe eltérő hangzóra jellemző pozíciót vett fel (az elülső affrikátákra jellemző, a hátulsó az /ɹ/-re jellemző formát). A réshang + /t/ esetében feltehetően arról van szó, hogy a réshang képzése után a felpattanóhoz képzendő zár nem minden ejtés során jön teljesen létre, kisebb-nagyobb szivárgás jelenik meg. Ennek vizsgálatához azonban célzott artikulációs kísérletre van szükség.

4.1. Limitációk

A módszertan leírásában említettük, hogy a BEA adatbázis olvasott anyagából összeválogatható kváziminimálpárok nem minden esetben teljesítenek olyan feltételeket, amelyeket kontrollálni kell további vizsgálatokban (egy szópár esetében homoszillabikus, szóvégi mássalhangzó-kapcsolat, míg a többi esetben heteroszillabikus, szóbelseji klaszterek; egy esetben a szópár között eltér, hogy első szótagi, tehát szóhangsúlyt viselő /f/ áll szemben többedik szótagbeli /s/

ejtésével; továbbá a szóvégi pozícióban az /f/ és a /t/ között morfémahatár húzódik, az /s/ és a /t/ között nem). Azért választottuk mégis első lépésnek ezt az anyagot használni, mert a kutatási kérdések megalapozásához nagy beszélőszámon céloztuk feltárni az esetleges nemek közötti eltéréseket a spektrális és koartikulációs mintázatokban, hogy a további, célzott anyagok felvételéhez megfelelő támpontot nyújtson. A marginális R^2 értéke a spektrális ferdeség és csúcosság esetében nagyon alacsony volt, így további, a jelen tanulmányban nem elemzett faktorok hatását is feltételezhetjük. Fontos megemlíteni, hogy a nemzetközi irodalomban is ellentmondásosak ezek az értékek, így feltehetőleg nem csak ebben az anyagban rejlik ennek az oka.

5. Összegzés

Összefoglalva elemzéseink megmutatták, hogy a réshang CoG értéke alacsonyabb volt /t/ mássalhangzó előtt, mint amikor magánhangzó követte. Az /s/ és /f/ közötti akusztikai távolság ugyanakkor nem csökkent a /t/-s környezet hatására. A nők és férfiak ejtésében nemcsak a két spiráns CoG-értékei tértek el egymástól, hanem a /t/-vel való koartikuláció eltérő mintázattal jelent meg a két nem esetében.

Köszönetnyilvánítás

A kutatás az NKFIH FK-128814-es projekt keretében készült. Köszönjük Jakus Julianna segítségét a hanghatárok javításában. Továbbá hálásan köszönjük a névtelen bírálók gondolatait, javaslatait, amelyek a kutatáshoz nemcsak ebben a tanulmányban, hanem a további lépések megtervezésében is hatalmas segítséget nyújtottak.

Hivatkozások

Algeo, J. (1978). What Consonant Clusters Are Possible? *Word*, 29, 206–224.
doi:[10.1080/00437956.1978.11435661](https://doi.org/10.1080/00437956.1978.11435661)

- Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language variation and change*, 23, 347–374.
- Bartoń, K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. URL: <https://CRAN.R-project.org/package=MuMIn>.
- Bates, D., Maechler, M., & Bolker, B. (2015). Walker, S. lme4: Linear mixed-effects models using S4 classes. *Journal of Statistical Software*, 67, 1–48.
- Beke, A., & Gyarmathy, D. (2010). Zöngétlen résmássalhangzók akusztikai szerkezete. *Beszédkutatás*, (p. 57–75).
- Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer. URL: <http://www.praat.org/>.
- Bóna, J., & Beke, A. (2013). A zöngétlen réshangok akusztikai szerkezete fiatal és idős korban. In T. Gecső, & Cs. Sárdi (Eds.), *Az interkulturális kommunikáció elmélete és gyakorlata* (p. 38–43). Budapest: Tinta Könyvkiadó.
- Czaplicki, B., Žigis, M., Pape, D., & Jesus, L. M. T. (2016). Acoustic Evidence of New Sibilants in the Pronunciation of Young Polish Women. *Poznań Studies in Contemporary Linguistics*, 52, 1–42.
- Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *The Journal of the Acoustical Society of America*, 84, 115–123.
- Fuchs, S., & Toda, M. (2010). Do differences in male versus female /s/ reflect biological or sociophonetic factors? In S. Fuchs, M. Toda, & M. Žygis (Eds.), *An interdisciplinary guide to turbulent sounds* (p. 281–302). Berlin: deGruyter Mouton.
- Grácz, T. E., & Kohári, A. (2014). Multiple bursts in Hungarian voiceless plosives and VOT measurements. In S. Fuchs, M. Grice, A. Hermes, L. Lancia,

- & D. Mücke (Eds.), *Proceedings of the 10th International Seminar on Speech Production (ISSP)* (p. 158–161). Cologne, Germany.
- Grácz, T. E., & Krepsz, V. (2018). Intervokális réshangok akusztikai szerkezete tinédzser és felnőttkorban. *Alkalmazott Nyelvtudomány*, 18, 1–19.
- Van der Harst, S., Van de Velde, H., & Schouten, B. (2007). Acoustic characteristics of standard Dutch /x/. In J. Trouvain, & W. Barry (Eds.), *Proceedings of the 16th ICPHS* (p. 1469–1472). Saarbrücken: Pirrot GmbH.
- Icht, M., & Boaz, M. B. D. (2017). Sibilant production in Hebrew-speaking adults: Apical versus laminal. *Clinical Linguistics & Phonetics*, 32, 193–212.
- Janda, R. D., & Joseph, B. D. (2003). Reconsidering the Canons of Sound-Change: Towards a “Big Bang” Theory. In B. Blake, & K. Burridge (Eds.), *Historical Linguistics 2001. Selected Papers from the 15th International Conference on Historical Linguistics, Melbourne, 13–17 August 2001* (p. 205–219). Amsterdam: John Benjamins Publishing Co.
- Jannedy, S., & Weirich, M. (2016). The acoustics of fricative contrasts in two German dialects. In Chr. Draxler, & F. Kleber (Eds.), *Proceedings of Phonetics and Phonology in German Speaking Areas* (p. 70–73). Munich: LMU.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of Acoustic Society of America*, 108, 1252–1263.
- Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research*, 56, 1175–1189.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82, 1–26.
- Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package. version 4.4-0. URL: <https://CRAN.R-project.org/package=ez>

- Linville, S. E. (1998). Acoustic correlates of perceived versus actual sexualorientation in men's speech. *Folia Phoniatrica et Logopaedica*, 50, 35–48.
- Magloughlin, L. (2018). */tɪ/ and /dɪ/ in North American English: Phonologization of a Coarticulatory Effect*. Ph.D. thesis University of Ottawa.
- Munson, B. (2004). Variability in /s/ production in children and adults. *Journal of Speech, Language, and Hearing Research*, 47, 58–69.
- Neuberger, T., & Grácsi, T. E. (2013). Az alveoláris zöngétlen explozívá variabilitása. *Beszédkutató*, (pp. 160–172).
- Neuberger, T., Gyarmathy, D., Grácsi, T. E., Horváth, V., Gósy, M., & Beke, A. (2014). Development of a large spontaneous speech database of agglutinative Hungarian language. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *International Conference on Text, Speech, and Dialogue* (p. 424–431). Cham: Springer.
- Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, 118, 2570–2578.
- Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language, and Hearing Research*, 32, 120–132.
- Ohala, J. J., & Solé, M. J. (2010). Turbulence and phonology. In S. Fuchs, M. Toda, & M. Žygis (Eds.), *Turbulent sounds: an interdisciplinary guide* (p. 37–102). Berlin: De Gruyter Mouton.
- Olaszy, G. (2007). *Mássalhangzó-kapcsolódások a magyar beszédben*. Budapest: Tinta Kiadó.
- Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., & Stockmann, E. (2004). The distinctness of speakers' /s/–/ʃ/ contrast is

- related to their auditory discrimination and use of an articulatory saturation effect. *Journal of speech, language, and hearing research*, 47, 1259–1269.
- Pharao, N., Maegaard, M., Møller, J. S., & Kristiansen, T. (2014). Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society*, 43, 1–31. doi:[10.1017/s0047404513000857](https://doi.org/10.1017/s0047404513000857).
- Phillips, J. B. (2020). *Sibilant Categorization, Convergence, and Change: The case of /s/-retraction in American English*. Doctoral dissertation The University of Chicago.
- Rácz, P., & Schepács, A. (2013). The perception of high frequency sibilants in Hungarian male speech. *Acta Linguistica Hungarica*, 60, 457–468. URL: <https://doi.org/10.1556/aling.60.2013.4.3>. doi:[10.1556/aling.60.2013.4.3](https://doi.org/10.1556/aling.60.2013.4.3).
- R Core Team (2020). R: A language and environment for statistical computing. URL: <https://www.R-project.org/>.
- Recasens, D. (2014). *Coarticulation and sound change in Romance*. Amsterdam–Philadelphia: John Benjamins Publishing Company.
- Recasens, D., & Pallarès, M. D. (2001). Coarticulation, assimilation and blending in Catalan consonant clusters. *Journal of Phonetics*, 29, 273–30. URL: <https://doi.org/10.006/jpho.2001.0139>.
- Recasens, D., & Rodríguez, C. (2016). A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics*, 59, 58–75.
- Shadle, C. H. (1990). Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (p. 187–209). Dordrecht: Springer.

- Shadle, C. H., & Mair, S. J. (1996). Quantifying spectral characteristics of fricatives. In *Proceeding of Fourth International Conference on Spoken Language Processing, ICSLP'96* (p. 1521–1524). Philadelphia: IEEE.
- Shapiro, M. (1995). A case of distant assimilation: /str/ → /ftr/. *American Speech*, 1, 101–107.
- Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. (2021). afex: Analysis of Factorial Experiments. R package version 0.28-1. URL: <https://CRAN.R-project.org/package=afex>.
- Smith, B. J., Mielke, J., Magloughlin, L., & Wilbanks, E. (2019). Sound change and coarticulatory variability involving english /ɹ/. *Glossa: a journal of general linguistics*, 4, 1–51.
- Smorenburg, L., & Heeren, W. (2020). The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues. *The Journal of the Acoustical Society of America*, 147, 949–960.
- Smyth, R., Jacobs, G., & Rogers, G. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society*, 32, 329–350.
- Stevens, K. N. (2000). *Acoustic phonetics*. Cambridge, Massachusetts: MIT Press.
- Stevens, M., Bukmaier, V., & Harrington, J. (2015). Pre-consonantal /s/-retraction. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th international congress of phonetic sciences* (p. 1–5). Glasgow, UK: the University of Glasgow.
- Stevens, M., & Harrington, J. (2016). The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics*, 58, 118–134.

- Stuart-Smith, J., Sonderegger, M., Macdonald, R., Mielke, J., McAuliffe, M., & Thomas, E. (2019). Large-scale acoustic analysis of dialectal and social factors in English /s/-retraction. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (p. 1273–1277). Canberra, Australia: Australasian Speech Science and Technology Association Inc.
- Varga, Á., Tarján, B., Tobler, Z., Szaszák, G., Fegyó, T., Bordás, C., & Mihajlik, P. (2015). Automatic close captioning for live hungarian television broadcast speech: A fast and resource-efficient approach. In A. Ronzhin, R. Potapova, & N. Fakotakis (Eds.), *Speech and Computer* (pp. 105–112). Cham: Springer International Publishing.
- Winter, B. (2019). *Statistics for linguists: An introduction using R*. New York: Routledge.

A rövid és hosszú zöngétlen explozívák észlelése felnőtteknél és kisiskolás gyermekeknél

Neuberger Tilda¹

¹*Nyelvtudományi Kutatóközpont, Általános és Magyar Nyelvészeti Intézet*

Abstract

There has been much research indicating that a specific speech sound is characterized by wide variability in timing of articulation. Despite the wide variability and the considerable overlap in the duration of singleton and geminate realisations, it is also confirmed by previous production studies that the primary acoustic attribute that distinguishes geminates from singletons is duration. In speech perception research, it is an essential question how listeners discriminate phonological categories of quantity along a continuous durational scale of the phonetic realisations. This perceptual task seems to be more difficult for children than for adults. The aim of this paper is to examine the role of duration as a perceptual distinction between single and geminate stops using a binary discrimination test of stops with systematically manipulated closure duration. The main question was what kind of similarities and differences can be found between 7-8-year-old children' and adults' perceptual shifts from singleton to geminate. Results confirmed that closure duration is a sufficient and adequate perceptual cue in the distinction of single and geminate stops both in children and adults; however, perceptual shifts differed between children and adults, and depended on the place of articulation. Moreover, displacement of the response curves based on the original quantity of stimuli (originally geminate or originally singleton) indicated that cues other than closure duration may contribute to the length contrast in Hungarian stops.

Keywords: perception, stop consonant, geminate, closure duration

1. Bevezetés

1.1. Elméleti háttér

Általános kiindulópont a laboratóriumi fonológiai kutatásokban, hogy a fonológiai releváns megkülönböztető jegyek (pl. zöngésség, hosszúság) a beszédhangok szintjén is realizálódnak az artikulációban, az akusztikumban és a percepcióban. A kvantitásra vonatkozó [$\pm$ hosszú] disztinktív jegyet az időtartamhoz, illetőleg az ejtés, hangoztatás tartamához szokás kapcsolni, eszerint

Email address: `neuberger.tilda@nytud.hu` (Neuberger Tilda)

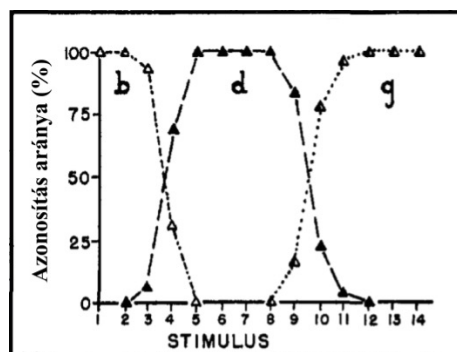
megkülönböztetünk rövid (szingleton) és hosszú (gemináta) mássalhangzókat. A nemzetközi szakirodalomban használt terminológia szerint a tanulmányban szinonimaként használjuk a *rövid mássalhangzó* és a *szingleton*, valamint a *hosszú mássalhangzó* és a *gemináta* terminusokat.

Számos akusztikai fonetikai kutatás alátámasztotta, hogy a rövid és a hosszú mássalhangzók megkülönböztetésében az elsődleges akusztikai attribútum az időtartam (lásd [Ridouane, 2010](#) összefoglalója 24 nyelvről). Univerzális jelenséggént elmondható, hogy a geminátákra a világ nyelveiben átlagosan 1,5–3-szor hosszabb időtartam jellemző a rövid konszonánsokhoz képest ([Ladefoged & Maddieson, 1996](#)). Az időtartam adatok azonban nagyfokú variabilitást mutatnak beszélőn belül és beszélők között is, amely a rövid és hosszú hangzók abszolút időtartamainak eloszlása között jelentős átfedést okoz (pl. [Hirata & Whiton, 2005](#); [Olaszy, 2006](#)). Ennek különféle okai lehetnek a beszédhang kontextusfüggő időviszonyaitól kezdve a beszélő artikulációs sajátosságaiig. Az artikulációs tempó különbségeinek és változékonyságának eredményeképpen például egy gyors artikulációs tempóban ejtett gemináta akár rövidebb időtartamú is lehet, mint egy rövid konszonáns.

Felmerül a kérdés, hogy a beszédhangok sokfélesége ellenére miként lehet az észlelésük a mindennapi kommunikációban mégis sikeres? A beszédpercepció kísérletek nagy része fókuszál annak vizsgálatára, hogy a beszédhangok különféle megvalósulásait figyelembe véve mik azok a releváns jegyek, amelyek lehetővé teszik az absztrakt fonémákhoz való hozzárendelésüket. Másképp fogalmazva: a jelentős mértékű variabilitás ellenére léteznek-e állandó, invariáns jegyek, amelyek a percepció alapjául szolgálnak? Ez a kérdés áll az invariancia-alapú beszédpercepció elméletek középpontjában, de a legtöbb beszédészlelési modell kitér rá. A következőkben megemlítenek néhányat, amelynek téziseit a jelen kutatás témájához közelebb állónak érezzük.

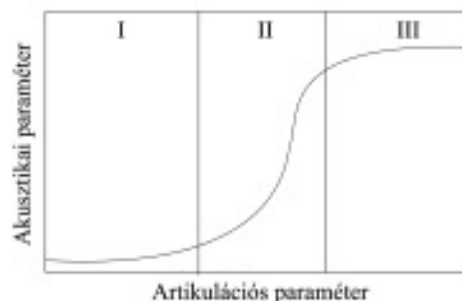
A **motoros elmélet** (*motor theory*: [Lieberman & Mattingly, 1985](#)), valamint a **közvetlen realista elmélet** (*direct realist theory*: [Fowler, 1986](#)) közös pontja, hogy összeköti az artikulációt a percepcióval. Feltételezik, hogy a hallgatók nem akusztikai fonetikai információkat, hanem gesztusokat nyernek ki az észlelés

során. Előbbi képviselői az észlelés invariáns alapegységét a beszélő szándékolt artikulációs gesztusaiban (*intended phonetic gestures*) látják, amelyeket a hallgató az akusztikai jelen keresztül a saját, produkcióról tárolt ismeretei alapján dekódol. Utóbbi középpontjában pedig az aktuális artikulációs gesztusok állnak, és nem az azokat megelőző szándékolt gesztusok, vagy az absztrakt fonémák. Vélhetően az akusztikai jelben található információ elegendő a gesztus felismeréséhez. Liberman et al. (1957) azon kísérletük alapján, amelyben a hallgatók egy kontinuum mentén folyamatosan változó szintetizált hangokat (/ba/ - /da/ - /ga/) diszkrét kategóriákként azonosítottak, felvetették a kategorikus észlelés fogalmát (1. ábra).



1. ábra. Kategóriák azonosítása (Liberman et al., 1957: 361)

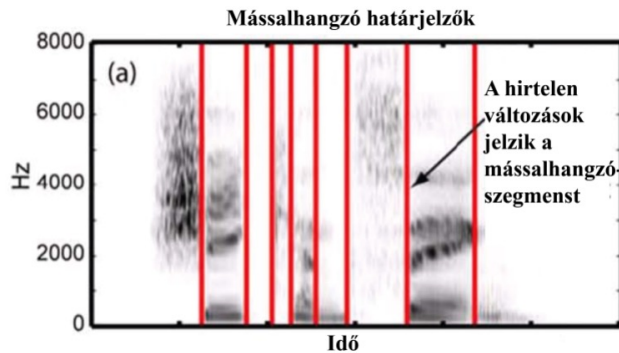
A kvantális elmélet (*quantal theory*: Stevens, 1989) az akusztikai jelben keresi az invariáns jegyeket. Kimondja, hogy az artikuláció és az akusztikum, valamint az akusztikum és a percepció közötti kapcsolat nem lineáris, hanem ugrásszerű (2. ábra). Ezt bizonyítja, hogy az akusztikai szerkezetben tapasztalható relatíve nagy eltérés esetén a hallgató nem feltétlenül észlel különbséget, ugyanakkor a percepció különbségtételben olykor apró akusztikai eltérések is szerepet játszanak. Az észlelést kategoriálisnak tartja. Az akusztikai paraméterek közül az időtartam kihívást jelent a kvantális interpretáció számára: nem tisztázott, hogy a szegmensek fizikai időtartamának széles változatosságát tekintve a hallgatók miként használják fel az időalapú információt a rövid és hosszú jegyek megkülönböztetéséhez.



2. ábra. Az artikuláció és az akusztikum kapcsolata (Stevens, 1989: 4)

Stevens (2002) későbbi modelljében hangsúlyozza a **megkülönböztető jegyek** (mint a fonológiai kontraszt alapegységeinek) szerepét, amelyekhez akusztikai kulcsok kapcsolhatók, és amelyek mentén a hallgató bináris döntéseket hoz a percepció során (vagyis az észlelés alapegységeinek tekinti). A megkülönböztető jegyek (pl. [mássalhangzó], [szonoráns]) a legkönnyebben az akusztikai jel azon részén azonosíthatók, ahol ún. **akusztikai határjelző** (*landmark*) található. Ezek olyan régiói a beszédnek, ahol hirtelen változás jelenik meg a beszéd spektrumában, mint például egy felpattanó zárhang esetében a felpattanás pillanata (3. ábra). A hallgató a határjelzők közvetlen környezetében elemzi az akusztikai paramétereket, melynek során eljut a megfelelő megkülönböztető jegyekhez, illetve az ezek kötegéből felépülő szegmensekhez (fonémák), majd megtalálja az ezek sorozatából álló szavakat a lexikonban.

Az **akusztikai invariancia elmélete** (*acoustic invariance theory*: Blumstein & Stevens, 1981) szerint az állandó, invariáns jegyek közvetlenül az akusztikai jelben találhatók. Az észleléshez szükséges információt a hallgatók jellemződetektáló mechanizmusok által nyerik ki a beszédjelből. Vizsgálták többek közt az explozívák képzési helyét, s kimutatták, hogy a zárfeloldás rövid idejű spektruma tartalmazza a képzési helyüket meghatározó invariáns jegyeket. Ezt olyan percepciók kísérletek eredményeivel támasztották alá, melyekben (szintetizált) CV hangsorokban a feloldás és a magánhangzó-átmenet akusztikai para-



3. ábra. Határjelzők az akusztikai jelben (Stevens, 2002 alapján Slifka, 2006: 95)

métereit szisztematikusan manipulálták, majd a zárhang azonosítását kérték a hallgatóktól (pl. Stevens & Blumstein, 1978). Az eredmények szerint a felpattanás utáni 10-20 ms-os szakasz jellemzőiben mutatkozik a zárhangok képzéshely szerinti különbsége. Mások a zöngés-zöngétlen megkülönböztetés invariáns jegeit kívánták felderíteni a zárhangok produkciós és percepció elemzésével (pl. Stevens & Klatt, 1974). Továbbá a rövid-hosszú megkülönböztetés akusztikai és felismerési kulcsait a tanulmányok jelentős hányada szintén a felpattanó zárhangok osztályán keresztül vizsgálta (pl. Lisker, 1958; Hankamer et al., 1989; Pickett et al., 1999; Tsukada et al., 2018; Zhou et al., 2021). Összetett akusztikai szerkezetük különösképpen indokoltá teszi a kérdés vizsgálatát az expozívak (illetve az affrikáták) körében: felvetődik ugyanis a kérdés, hogy a kvantitás a mássalhangzó komplex időszerkezetének mely elemében manifesztálódik. A munkák egy része a zárszakasz abszolút időtartamát vette górcső alá, mások a relatív időtartamokat, vagyis a különböző paraméterek egymáshoz viszonyított szerepét hangsúlyozzák az invariáns jegyek feltérképezésében.

A zárszakasz abszolút időtartama elegendő felismerési kulcsnak bizonyult azokban a percepció tesztekben, amelyek a záridőtartam mesterséges manipulációjával létrehozott stimulusokat alkalmaztak. Ilyen teszteket a világ számos nyelvén elvégeztek, többek között a maráthi nyelvben (Lisker, 1958), az arabban (Obrecht, 1965), a törökben és a bengáliban (Hankamer et al., 1989), a japánban

és a finnben (Yoshida et al., 2015; Zhou et al., 2021) vagy az olaszban (Esposito & Di Benedetto, 1999). Ezen kutatások közös célja az volt, hogy megtalálják, hol található az a pont, amikor a hallgatók észlelése a rövid hang megítéléséből átvált a hosszú hang ítéletébe, vagyis hogy hol található a percepció határ a singleton-gemináta kategóriák között. Az eredményeik szerint a singleton és a gemináta közötti percepció határvonal a zöngétlen zárhangok esetén valahol a 120–180 ms közötti tartományba esik (pl. Esposito & Di Benedetto, 1999; Yoshida et al., 2015; Zhou et al., 2021).

A **relációs akusztikai invariancia** (*relational acoustic invariance*) elmélete szerint az invariáns jegyek nem önmagukban találhatóak meg az akusztikai szerkezetben, hanem a jellemzők egymással való kapcsolatában (pl. Sussman et al., 1991; Amano & Hirata, 2010, 2015). A mássalhangzó-hosszúság tekintetében ez azt jelenti, hogy a relatív időtartamok, időviszonyok döntő szerepet játszanak a kategóriák észlelésében. Más szóval: nem csak (és nem elsődlegesen) a zárszakasz abszolút időtartamában érdemes a hosszúság akusztikai kulcsát keresni, hanem az időtartamok (pl. a zárszakasz és a megelőző magánhangzó időtartamának) egymáshoz viszonyított arányában. Továbbá nem hagyható figyelmen kívül az artikulációs tempó variabilitása sem. Amano & Hirata (2010) alátámasztotta azt, hogy míg a rövid-hosszú kategóriák közötti percepció határhoz kapcsolódó záridőtartamok a tempótól függően változatos képet mutatnak (34–213 ms, átlag: 110 ms), addig a záridőtartamnak a szóidőtartamhoz viszonyított aránya egy állandó (invariáns) paraméternek tűnik a két fonológiai kategória elkülönítésében.

Az **anyanyelvi mágnes elmélete** (*native language magnet theory*; Kuhl et al., 2008) kimondja, hogy a gyermekek a születés utáni korai szakaszban (10 hónapos korukig) még képesek a világ nyelveiben előforduló legtöbb fonéma megkülönböztetésére, ami lehetővé teszi, hogy bármely nyelv fonémakészletét elsajátítsák, ám ezután átrendeződik a fonémakategória-rendszerük, amelyben a saját környezetük nyelve lesz az irányadó. Ettől fogva észlelésük csak az anyanyelvük fonémái közötti különbségekre lesz érzékeny. Mindez a prototípusok elvén alapul; a mindaddig hallott számos beszédhang-példányból kialakítják az

adott fonémához tartozó prototípust. A prototípusok perceptuális mágnesként funkcionálnak, amelyek a bemeneti beszédjelen a hasonló akusztikai jellemzőkkel bíró szegmenseket „magukhoz vonzzák”. Így a kategórián belül csökkennek a különbségek, viszont a prototípushoz nem hasonlító hangok esetében nő a különbség, vagyis ezek eltérő kategóriába fognak kerülni. Ez, az univerzálisból a nyelvspecifikus észlelésbe történő váltás a későbbiekben nehézségeket okozhat az idegen nyelv tanulásában, különösképpen a L2 fonológiai kategóriák elsajátításban. A mássalhangzó-hosszúságot illetően ez a jelenség mutatkozik meg azokban a percepció tesztekben, amelyekben nem anyanyelvi beszélők hosszúságészlelését vizsgálták (pl. [Wilson et al., 2005](#), [Sonu et al., 2013](#), [Yoshida et al., 2015](#); [Tsukada et al., 2018](#)). Egyrészt azt találták, hogy a nyelvtanulók javarészt az abszolút időtartamokra támaszkodnak a döntéseikben, és nem használják a kontextus-, illetőleg tempófüggetlen felismerési kulcsokat, mint ahogyan az anyanyelvi beszélők; emiatt számukra nehézséget jelent a hosszúsági kontraszt azonosítása, különösképp az artikulációs tempó változatossága esetén. Másrészt felhívták a figyelmet arra, hogy a hallgatók a kvantitásnak a saját nyelvükre jellemző nyelvspecifikus vonásait veszik alapul a megkülönböztetésben akkor is, amikor idegen nyelvű rövid és hosszú mássalhangzókról kell döntést hozniuk.

1.2. A mássalhangzó-hosszúság vizsgálata a beszédfejlődésben

Az idegen nyelv tanulása mellett az anyanyelv elsajátításában ugyancsak elengedhetetlen a releváns jegyek felismerését lehetővé tevő akusztikai kulcsok szerepe, és azoknak a redundáns jegyeknek a számbavétele, amelyek irrelevánsak, tehát ignorálhatók. Csecsemő- és gyermekkorban a bemeneti beszédjel (input) elemzésén keresztül fejlődik a percepció és a produkció mechanizmus, vagyis az akusztikai jelenségek észlelése, illetőleg az artikulációs parancsok végrehajtása. A gyermekek fonológiai fejlődésével kapcsolatban a kutatók egy része úgy tartja, hogy a fent említett folyamatokhoz minden bizonnyal szükség van invariáns jegyekre, amelyek alapján a gyermek felfedezi anyanyelvének fonémikus szerkezetét ([Klatt, 1979](#)).

Az anyanyelvi fonológia elsajátításának vizsgálatában fontos kérdés, hogy a gyermek hogyan képes a hozzá intézett folyamatos beszédet szegmentálni, különálló egységekre bontani, illetve mik lehetnek a beszédelemzés, illetőleg az észlelés alapegységei. A fonológiai fejlődés leírását célzó munkákban legtöbbször két hagyományos megközelítéssel találkozhatunk, amelyek máig jelentősek: a generatív nyelvtanban általánosan alkalmazott szabály-, vagy megszorítás-alapú elemzéssel, amelyben a fonológiai szerveződés alapegysége a szegmens (pl. Gnanadesikan, 2004), valamint a kognitív megközelítéssel, melyben általában az alapegység a szó mint teljes egész (pl. Vihman & Velleman, 2000).

A gyermekek beszédészlelésére és szófelismerésére vonatkozó kutatások felfedték, hogy a gyermekek különös figyelmet fordítanak a spektrum globális mintázatára, míg a felnőttek észlelése az akusztikai jel részleteire fogékonyabb (Jusczyk, 1993; Nittrouer & Lowenstein, 2009). Alátámasztották azt is, hogy – a nem anyanyelvi beszélőkhöz hasonlóan – a gyermekek másképp súlyozzák a felismerési kulcsokat az észlelésükben, mint a felnőttek (Nittrouer & Lowenstein, 2009). Ezt azt jelenti, hogy a gyermekek és a felnőttek között eltérés mutatkozik abban, hogy az akusztikai jel mely tulajdonságai lesznek számukra fonetikailag informatívabbak, vagyis kiemelt jelentőségűek (szálieensek), melyek segítik a hallgatót a beszédhangokat fonémaosztályokba sorolni (fonémadóntés) vagy nagyobb, jelentéssel bíró egységekbe (szófelismerés) rendezni.

A mássalhangzók kvantitásának kérdése a gyermekek beszédfejlődésében kevésbé kutatott terület a nemzetközi szakirodalomban, mint a felnőtteknél. Vihman & Velleman (2000) megállapította, hogy az első szavaikban még azok a gyermekek is ejtenek hosszú (időtartamú) mássalhangzókat, akik olyan nyelvet sajátítanak el, ahol a rövid és hosszú mássalhangzók nem állnak fonológiai oppozícióban (pl. angol, francia). Ez azonban az anyanyelv-elsajátítás egyszavas szakaszán túl már csak azokban a nyelvekben figyelhető meg, amelyekben a kvantitás fonológiai releváns jegy (pl. finn, japán).

A nyelvek közötti vizsgálatok arra is felhívták a figyelmet, hogy a különböző anyanyelvű gyermekek nem azonos életkorban kezdik el stabilan megkülönböztetni a produkcióban a mássalhangzókat a kvantitás alapján: a finn gyermekek

például gyorsabban sajátítják el a kontrasztot, mint a japánok (Aoyama, 2000; Kunnari et al., 2001). Ennek okát a bemeneti beszéd (input) különbségeiben látják a kutatók: egyrészt a kvantitáskontraszt gyakoribb, másrészt az időtartamkülönbség jelentősebb a finnben, mint a japánban. A gyakran hallott kategóriák közötti kontrasztok ugyanis előbb épülnek be a nyelvspecifikus elsajátítás során, mint a ritkábban hallottak (Anderson et al., 2003).

(Khattab & Al-Tamimi, 2013) a libanoni arab gyermekek produkciójában U-alakú fejlődési görbét talált a mássalhangzó-hosszúság oppozíciójában, mely szerint a kontraszt pontos realizációja egy korai fejlődési stádiumban már jelen van, majd egy visszaesés tapasztalható, és csak később alakul ki a felnőtt ejtésbeli fonológiai struktúra. Hasonlót mutattak ki a norvég szonoránsok és zöngétlen explozívák vizsgálatában a 2 és fél, a 4 és a 6 évesek produkciójában (Payne et al., 2017).

A percepció oldaláról csekély számú nemzetközi tanulmány foglalkozott a geminátákkal gyermekeknél. Egy japán csecsemőkkel vizuális habituáció-diszhabituáció (ingerre irányuló figyelem, nézéspreferencia) technikát alkalmazó kutatás (Sato et al., 2012) arról számol be, hogy a gyermekek 4 hónapos korukban még nem, de 9 és fél hónapos korukban már képesek szingleton-gemináta diszkriminációra a zöngétlen explozívák körében, és akkor is csak a természetes ejtésű (vagyis nem a mesterségesen manipulált záridőtartamú) stimulus esetén. Ez utóbbi eredmény a természetes beszédben megjelenő kiegészítő kulcsok szerepére hívja fel a figyelmet, amelyre úgy tűnik, a kisgyermekek is szenzitívek. A tanulmány eredményei arra engednek következtetni, hogy ilyen fiatal korban már érzékenyvé válnak a gyermekek a kontraszt észlelésére, azonban ekkor még nem beszélhetünk fonológiai tudatosságról. A szavak belső szerkezetéhez való tudatos hozzáférés, illetve a szavak eltérő méretű egységeire történő bontás képessége későbbi életkorban (3-4 éves kortól) alakul ki, és a fonológiai tudatosság egyes területeinek (pl. rímképzés, szótagszintézis) fejlődése nyelvspecifikus sajátosságokat mutathat (Jordanidis, 2015).

1.3. A mássalhangzó-hosszúság vizsgálata magyar nyelven

Az utóbbi évtizedben egyre fokozódó érdeklődés övezi a magyar gemináták produkciós sajátosságait (pl. Siptár & Grácz [2014], Neuberger [2015], Neuberger & Beke [2017], Deme et al. [2019]). A tanulmányok elsősorban a komplex akusztikai szerkezetű felpattanó zárhangokat elemezték felnőttek ejtésében. A mássalhangzó-hosszúság észlelésére vonatkozó elemzések kisebb számban készültek, és főként tágabb beszédfejlődési vizsgálatok részeként jelentek meg, amelyek célja a gyermekek fonéमतudatosságának feltérképezése volt (pl. Gósy [2006], Gyarmathy & Horváth [2010], Jordanidisz [2015]). Közös eredményeik alapján elmondható, hogy a gyermekkor egy bizonyos szakaszáig nehézséget jelent a kvantitáskategóriák tudatos megkülönböztetése. A fonológiai tudatosság fejlődése során a szenzitivitás és a manipulációs készség fokozatosan fejlődik a nagyobb egységektől (szótagok) a kisebbek (fonémák) felismeréséig. Továbbá a gyermekek a nagyobb különbségektől a kisebbekig haladnak az észlelés fejlődésében: a beszédhangok differenciálása során a legkönnyebb a képzés módjának eltérését felismerniük, ezt követi a képzés helye szerinti különbség, majd a zöngés-zöngétlen eltérés észlelése. Legnehezebb a hangok időtartam-különbségének felismerése.

A fenti sajátosságokat veszik figyelembe a beszédészlelést vizsgáló eljárások is. Például a 4 és 8 éves kor közötti gyermekek számára készült DIFER beszédhanghalló (fonémaészlelési) készséget vizsgáló diagnosztikus teszt (Nagy et al. [2004]) arra épül, hogy a hangok differenciálása a megkülönböztető jegyek számának csökkenésével párhuzamosan nehezedik. A beszédhanghallás készsége akkor mondható kialakultnak, ha a gyermekek a csupán egy jegyben történő eltérést is érzékelni tudják.

A sztenderdizált GMP-diagnosztika a 3–13 évesek beszédfeldolgozási folyamatait méri fel (Gósy [1995/2006]). A beszédhang-differenciálást mérő alteszben a gyermekek feladata annak megítélése, hogy két elhangzott hangsor egyforma (pl. *begi–begi*) vagy különböző (pl. *voka–vokka*, *ípi–íppi*, *tegő–teggő*). Nagy mintán (hatszáz 4 és 9 év közötti gyermek) vizsgálva Gósy és Horváth ([2006]) azt találták, hogy a beszédhang-differenciálásban jelentős változás hét- és nyolc-

éves kor között tapasztalható. Mind a produkciós, mind a percepciós fejlődésben általában az utolsó lépcsőfok a fonológiai hosszúsági oppozíció megkülönböztetése; ennek stabilizálódását 7 éves korra szokás tenni (Gósy, 2006). Horváth és Gyarmathy (2010) a 6–10 év közötti, hallászavarral küzdő óvodások és kisiskolások beszédhang-megkülönböztetési képességét vette górcső alá. A képzési jegyek közül a mássalhangzó hosszúsága szerinti differenciálás jelentette a legnagyobb nehézséget a gyermekek számára, sőt még a tízéveseknek is csupán a fele tudta elkülöníteni a hangsorokat a konszonánsok időtartama szerint.

Jordanidisz (2015) 4 és 10 év közötti magyar anyanyelvű gyermekek ($N = 390$) fonológiai tudatosságát elemezte. A fonématudatosságot mérő tesztek alapján kimutatta, hogy 6 éves kor után a gyermekek teljesítménye ugrásszerűen nő (feltehetően az olvasástanulás hatására). Az egyik altesztben a gyermekek feladata a hosszú beszédhangok azonosítása volt egy-egy adott szóban. Az eredmények szerint 6 éves korban 45%-os, 7 éves korban 73%-os, 8 éves korban pedig már 93%-os átlagteljesítményt nyújtottak a gyermekek. A hétévesek eredményei szignifikánsan eltértek a nyolcévesekétől ebben a feladatban. A szerző a beszédhangok temporális jegyeinek pontos észlelését 8 éves korra (2. osztályos korra) teszi, bár hozzáteszi, hogy eredményei alapján sokan még 3. osztályban sem észlelik a beszédhanghosszúságot megfelelően.

1.4. A jelen kutatás célja

A jelen kutatás arra keresi a választ, hogy miként különböztetik meg a rövid és a hosszú mássalhangzókat az észlelésben az időparaméter mentén a magyar anyanyelvű felnőttek és 7-8 éves kisiskolás gyermekek. A zöngétlen explozívákon végzett kutatásunk hipotéziseit az alábbiakban fogalmaztuk meg: 1. A zárképzés időtartama elégséges kulcs a kvantitásra vonatkozó döntés meghozatalában mind a felnőttek, mind a gyermekek számára. 2. A gyermekek kvantitásra vonatkozó döntései kevésbé kategorikus észlelési mintázatot (laposabb választásgörbét) mutatnak, mint a felnőtteké, valamint 3. a rövid és hosszú hangok közötti határ a felnőttekétől eltérő időtartamértéknél húzódik az észlelésükben. 4. Továbbá feltételezzük, hogy a záridőtartamukban manipulált zárhangok ere-

deti kvantitása (szingletonból vs. geminátából módosított hang) befolyásolja a hallgatók ítéleteit (az akusztikai jelben található esetleges másodlagos kulcsok hatására, amelyeket nem módosítottunk a jelen kísérletben) – függetlenül az életkortól vagy a mássalhangzók képzési helyétől.

2. Anyag, módszer, kísérleti személyek

A jelen kutatásban bináris döntési teszt segítségével vizsgáljuk a különböző életkorú (gyermek és felnőtt) hallgatók mássalhangzó-hosszúságra vonatkozó percepciók ítéleteit. A kísérleti anyagot a Nyelvtudományi Intézet Fonetikai osztályának csendesített szobájában rögzítettük egy magyar anyanyelvű, 27 éves női beszélőtől, aki gyakorlott fonetikus lévén a hangsorok felolvasása alatt végig azonos szinten tartotta az artikulációs tempót, a dallammenetet és az átlagos intenzitást. A felvétel AT 4040 kardioid kondenzátor mikrofonnal történt a Gold-Wave programban,¹ 44,1 kHz mintavételezési frekvencián, 16 bit-es tárolással. A teszthez felhasználandó anyag meghatározásához egy előkísérletet végeztünk, melynek eredményeit Neuberger (2016) ismerteti. Ennek eredményei alapján választottuk ki a jelen kísérlet anyagát: olyan VCV hangszekvenciákat (logatomokat), amelyekben az intervokális helyzetben lévő mássalhangzó szingleton vagy gemináta [p,t,k], a környező magánhangzó pedig minden esetben az [i]. Az előkísérlet résztvevői az eredeti *ipi*, *ippi*, stb. hangszekvenciákban szereplő mássalhangzókat 100%-os egyetértéssel a megfelelő kategóriaként azonosították.

A hat kiválasztott tokenből (3 képzési hely $\times$ 2 hosszúsági kategória) a zárszakasz szisztematikus időtartam-manipulációjával 61 stimulust hoztunk létre a következőképpen: a rövid explozívák néma szakaszát 10 ms-onként növeltük, míg elérte a hosszú pár időtartamát (II táblázat). A kezdő rövid [p] hang például 123 ms-os volt, ezt nyolc 10 ms-os lépésben növeltük, amíg elérte a hosszú [p:] hangnak megfelelő 203 ms-ot. A hosszú explozívák zárszakaszát pedig 10 ms-os lépésekben csökkentettük, míg elérte a rövid pár időtartamát. Így a két szélső érték, valamint minden köztes lépésben kétszer szerepelt egy inger: egy

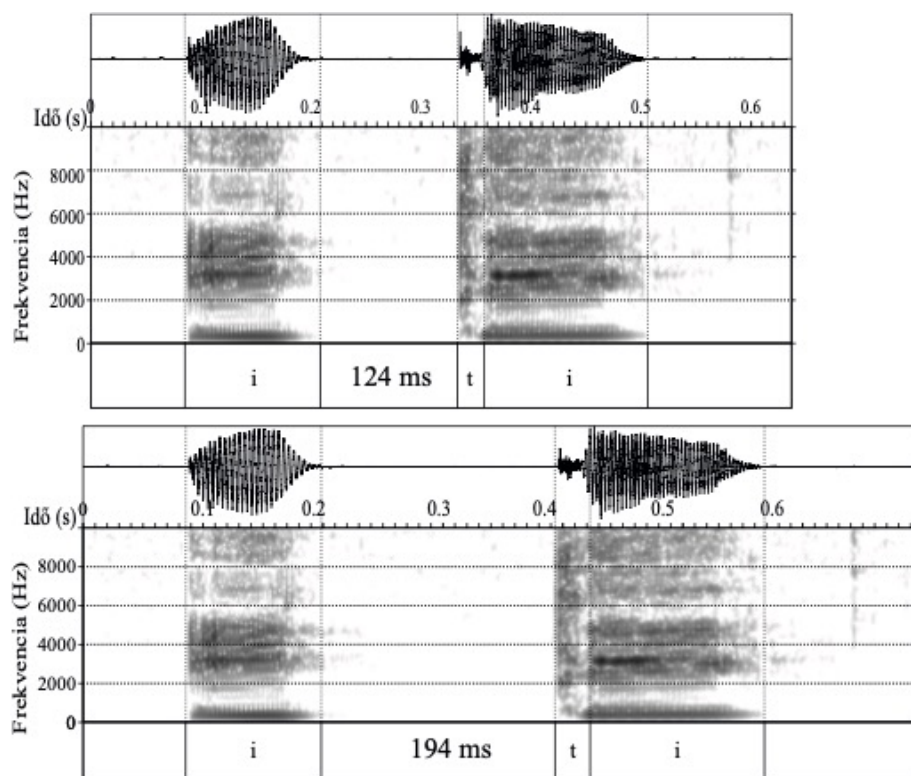
¹5.12 verzió, www.goldwave.com

szingletonból és egy geminátából előállított, azonos záridőtartamú hang. Ezt követően további lépésekkel is kibővítettük az időmanipulálást: a szingletonok időtartamán túl is csökkentettük a gemináták zárídejét a 100 ms-ot megközelítő értékekig. Arra voltunk kíváncsiak, hogy a természetes ejtésű (olvasott) szingletonnál rövidebb zárszakaszok esetén vajon egyöntetű lesz-e a hallgatók ítélete a „rövid” kategória felé. Az így kialakított kísérleti anyagban jól megfigyelhető lett a 100–200 ms közötti záridő-intervallum. Bár a képzési hely szerint különbséget mutattak az eredeti zöngétlen explozívák időértékei a szakirodalmi adatoknak megfelelően (a képzési helyben hátrafelé haladva egyre rövidebbek a záridőtartamok, pl. Grácz, 2013; Neuberger, 2015), a csökkentést egységesen a 100 ms körüli időértékekig folytattuk, a labiális hang esetén tíz lépésben, az alveolárisnál tizenkettő lépésben, a velárisnál pedig kilenc lépésben (innen adódik a stimulusok nem egyenlő száma az egyes képzéshelyeken). A maximum értékek pedig a felolvasott gemináták időtartamához illesztve a jelen anyagban [p] 203 ms, [t] 224 ms és [k] 194 ms-osak voltak.

1. táblázat. A kísérletben használt stimulusok adatai

Mássalhangzó	A záridőtartam minimum értéke	A záridőtartam maximum értéke	Stimulus száma
	(ms)	(ms)	(db)
[p] – [p:]	103	203	20
[t] – [t:]	104	224	23
[k] – [k:]	104	194	18

Az időtartamok manipulálása a Praat programban (Boersma & Weenink, 2020) történt, melynek során csupán a zöngétlen zárhangok néma fázisát növeltük; a felpattanást, a zöngékezdési időt, valamint a környező magánhangzókat változtatlanul hagytuk. A 4. ábrán látható egy példa, melyben a 124 ms-os zárszakaszt 194 ms-osra növeltük, az *i*-vel jelölt magánhangzókat és a *t*-vel jelölt felpattanást és VOT-t az eredeti ejtés szerint meghagytuk.



4. ábra. Példa a záridőtartam manipulálására a Praat programban

A percepció tesztet minden hallgatóval egyénileg végeztük el a Praat szoftverben az ExperimentMFC nevű modul segítségével. A gyermekeket egy csendes iskolai teremben, a felnőtteket egy egyetemi szobában teszteltük. Ugyanazt a rögzített anyagot hallgatták meg Sennheiser HD 419 fejhallgatón keresztül. A feladatuk az volt, hogy egy rövid stimulus (pl. [ipi]) meghallgatása után döntést hozzanak arról, hogy a hangmintában szereplő mássalhangzó rövid vagy hosszú (pl. [ipi] vagy [ip:i]). Ha a hallgató észlelése alapján rövidnek ítélte a hangot, akkor a „rövid”, ha hosszúnak, akkor a „hosszú” gombra kellett kattintania (vagy a kisiskolások esetében: rámutatnia a „rövid” vagy „hosszú” kategóriát jelképező képre). A 7-8 éves gyermekekkel – életkorukból fakadóan – játékos formában, egy kerettörténet alkalmazásával végeztük el a kísérletet. Azon gyermekek számára, akik a teszt elvégzése idejében még nem tudtak (jól) olvasni, két kép is

segítette a döntést. Egy rövid és egy hosszú testű tacsót ábrázoló kép közül kellett kiválasztaniuk, hogy gazdájuk aktuálisan melyiket hívja: a „rövid” nevűt (pl. [ipi]) vagy a „hosszú” nevűt (pl. [ip:i]), és erre a kutyák testalkata utalt. A kísérlet anyagát jelentő logatomokat random sorrendben játszottuk le. Előzetesen minden adatközlőnek bemutattuk a feladatot, az első néhány próbát nem számítottuk bele a kísérleti anyagba. Az előtesztelés alapján elmondható, hogy a vizsgálat-hoz választott résztvevők (beleértve a gyermekeket is) a feladatot megértették.

Minden adatközlő magyar anyanyelvű, egynyelvű beszélő volt. A vizsgálat-hoz eredetileg felkért 40 gyermek között többen voltak, akik egy idő után szisztematikusan csak rövid vagy csak hosszú válaszokat adtak a tesztelés során, ezért az ő válaszaikat kizártuk az elemzésből, mert kétséges, hogy a válaszok valóban a gyermekek percepciós működését tükrözték, vagy a figyelem fenntartása a feladat végéig túlzottan nehéznek bizonyult számukra. A kísérleti személyek két csoportját így 32 gyermek és 44 fiatal felnőtt alkotta. Náluk nem volt tapasztalható figyelemfenntartási nehézségből fakadó véletlenszerű jelölés.

Az összes gyermek ép halló és ép értelmű, tipikus fejlődésűnek mondható, beszédhibájuk nem volt, közöttük 16 lány és 16 fiú szerepelt. A gyermekek 7 és 8 évesek (átlagos életkor: 7,5 év); egy budapesti általános iskola első és második osztályos tanulói voltak. A vizsgálatot mindkét osztályban a tanév második félévében végeztük el. A korábbi szakirodalmi adatok alapján (lásd Bevezetés) választottuk ezt az életkort vizsgálatunk alapjául. Alátámasztották ugyanis azt, hogy a mássalhangzók hosszúságának észlelésében kiemelt jelentőségű e korosztály: az anyanyelv-elsajátítás ebben a szakaszában jelentős változás történik a kvantitásoppozíció észlelésében a korábbi életévekhez képest, illetve 7-8 éves korra elvárható a beszédhang-hosszúság szerinti diszkrimináció pontos működése.

A kísérletben részt vett felnőttek 18–27 év közöttiek (átlagos életkor: 21 év). Egy budapesti egyetem magyar nyelv és irodalom szakos alapképzésének hallgatói voltak, kevés fonetikai ismerettel rendelkeztek a teszt elvégzésének idején. A 44 felnőtt között 33 nő és 11 férfi szerepelt. Minden adatközlőtől ($N = 76$) a 61 stimulusra kaptunk egy „rövid” vagy „hosszú” (bináris: 0 vagy 1) választ, így összesen 4636 választ elemeztünk.

Az elemzés során összesítettük az egyes stimulusokra adott válaszok arányát. Meghatároztuk a „hosszú” válaszok százalékos arányát a zárszakasz-időtartam függvényében. Eszerint például ha egyetlen válaszadótól sem kaptunk „hosszú” választ, akkor az arány 0%; amennyiben viszont minden válaszadó hosszúnak ítélte meg az adott hangot, az arány 100%-os. Ahol 50%-os az arány, ott a válaszadók fele-fele arányban ítélték a hangot singletonnak, illetve geminátának (vagyis ennél a pontnál egyenlő a valószínűsége, hogy „rövid” vagy „hosszú” az elhangzott hang a percepció szerint). A kapott válaszokra logisztikus függvényt illesztettünk, és ábráztuk a válaszgörbékét. A statisztikai elemzéshez bináris logisztikus regressziót alkalmaztunk az SPSS 20-as verziószámú szoftverben². Általánosított lineáris kevert modellt (GLMM) építettünk az adatokra, amelyben a függő változó a 0:1-es 'válaszok' voltak, a 'záridőtartamot' fix hatásként, az 'adatközlőket' random hatásként tüntettük fel. Továbbá elemeztük az 'életkor' (gyermek, felnőtt) és a 'képzéshely' (labiális, alveoláris, veláris) eredményekre gyakorolt hatását is (a 'záridőtartam' és az 'életkor', valamint a 'záridőtartam' és a 'képzéshely' interakciója is szerepelt a modellünkben mint fix hatás). Összehasonlítottuk továbbá a gyermekek csoportján belül a 7 éves és a 8 évesek eredményeit, valamint az eredetileg singletonból és az eredetileg geminátából létrehozott stimulusokra adott válaszokat; majd nem parametrikus próbákkal (Wilcoxon, Friedman-teszt) kiértékeljük a különbségeket.

3. Eredmények

3.1. A mássalhangzó-hosszúság észlelése gyermekeknél és felnőtteknél

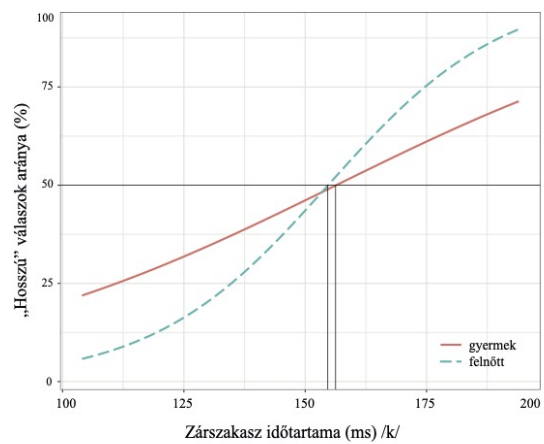
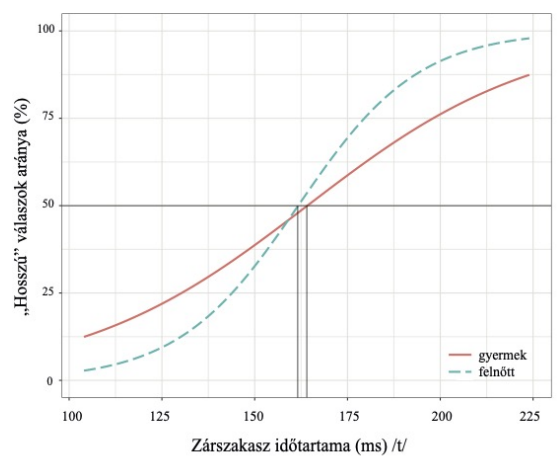
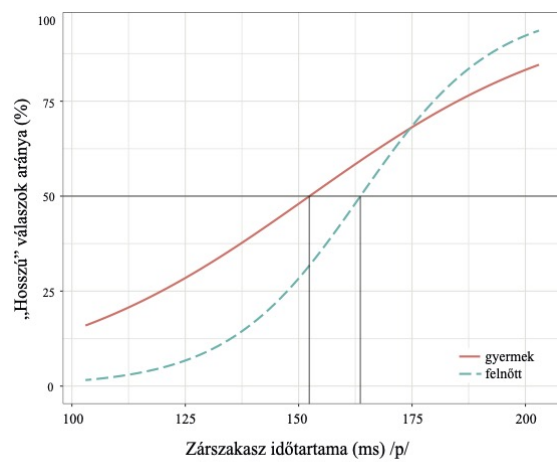
Az elemzés első lépéseként megvizsgáltuk a hallgatók ítéleteit az elhangzott mintában található explozívák zárszakaszának időtartamértékei alapján. Mivel a korábbi produkciós eredmények (Ridouane, 2010, magyarra: Neuberger, 2015) azt mutatják, hogy a záridőtartam a legfőbb akusztikai attribútum a fonológiai hosszúság megkülönböztetésében, azt vártuk, hogy a hallgatók is felhasználják

²www.ibm.com/analytics/spss-statistics-software

ezt a paraméter a percepcióban. A hallgatói ítéletek alapján kialakult válaszgörbék (5 ábra) azt mutatják, hogy mind a felnőttek, mind a gyermekek döntéseit meghatározta a záridőtartam mindegyik expozíva esetén: relatíve rövid zárszakaszhoz „rövid” ítéletek, hosszú időtartamhoz pedig „hosszú” ítéletek adódtak nagy arányban. A statisztikai elemzés kimutatta, hogy a záridőtartam szignifikánsan befolyásolja a hallgatók döntését: $F(1, 4629) = 869,142$, $p < 0,001$.

A felnőttek észlelésében a kísérleti anyagban szereplő minimum (100 ms körüli) záridőtartam esetén elvértve jelentek meg „hosszú” válaszok, míg a maximum (>200 ms) záridőtartamoknál a hallgatók döntése szinte egyöntetű volt a „hosszú” válaszokban. A felnőtt hallgatók egyöntetűen rövidnek ítélték a geminátákból 100 ms körüli záridőtartamúra csökkentett [p] és [t] hangokat, ám a [k] esetében 6,8%-uk „hosszú” választ adott a 100 ms körüli (legrövidebb záridőtartamú) expozívára. A 200 ms körüli expozívákra 82–97%-os arányban „gemináta” ítéletek születtek (itt is a [k] kategóriába sorolása mutatkozott a legbizonytalanabbnak). A felnőttek esetében a szigmoid (S-alakú) függvényt megközelítő eredmények a mássalhangzó-hosszúság észlelésének kategorikus voltára utalnak. A gyermekek válaszait tekintve a felnőttektől eltérő eredményt kaptunk. Egyfelől nem találtunk náluk 0% vagy 100%-os egyöntetű eredményt egyik hangnál sem, még a minimum és a maximum záridőtartam esetében sem. Másfelől az adatokra illesztett függvény alakja nem S-alakúnak mutatkozott, hanem megközelítőleg lineárisnak.

A statisztikai elemzés szerint annak valószínűségét, hogy a hallgató rövidnek vagy éppen hosszúnak ítél meg egy adott időtartamú zárhangot a zárszakasz hosszán túl az adott hang képzéshelye is befolyásolja: $F(2, 4629) = 4,160$, $p = 0,016$. Azt is kimutatta az elemzés, hogy a záridőtartam a képzési hely szerint nem minden életkori csoportban van azonos hatással a döntésekre: $F(5, 4629) = 7,964$, $p < 0,001$. Az adatokra illesztett görbékben jól látható, hogy labiális hang esetén található a legnagyobb különbség a gyermekek és a felnőttek válaszadása között (tekintve az 50%-nál meghúzható „percepciók határhoz” tartozó záridőtartamok különbségét a gyermekek és a felnőttek között).



5. ábra. Válaszgörbék és percepció határok gyermekeknél és felnőtteknél

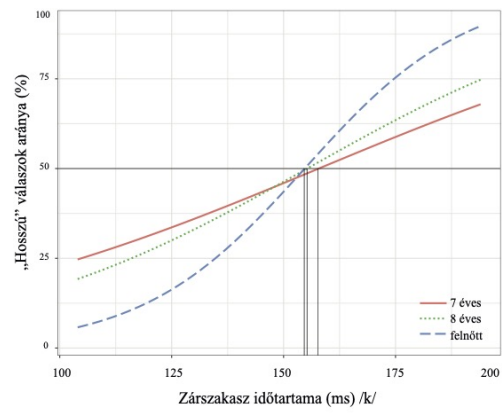
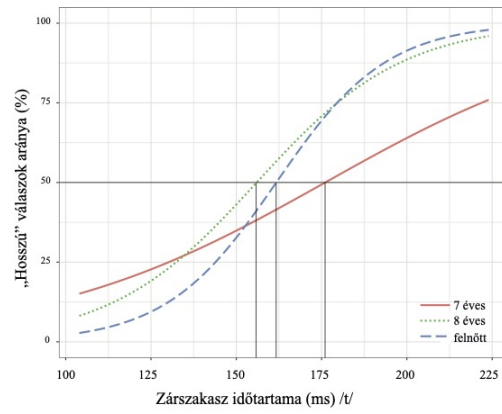
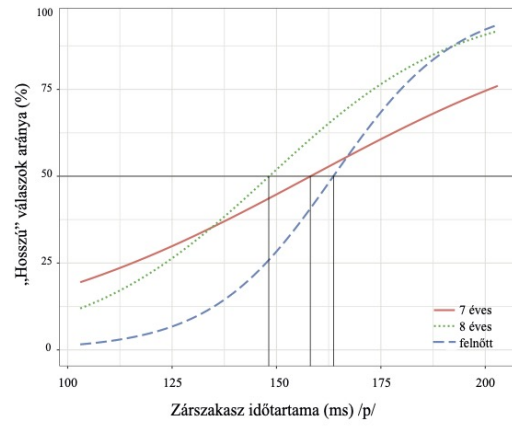
A szingletonok és a gemináták közötti percepció határt a nemzetközi szakirodalomban bevett gyakorlat szerint annál a záridőtartamnál húztuk meg, ami az adatokra illesztett logisztikus görbén az 50%-os válaszadáshoz kapcsolható (ahol a válaszadók fele rövidnek, fele hosszúnak ítélte meg az adott hangot). Ezt úgy határozhatjuk meg, hogy megvizsgáljuk azt a pontot, ahol az adatokra illesztett logisztikus görbe az 50%-os válaszadás egyenesét metszi. Az explozívák képzéshely szerinti különválasztása a percepció határ kérdésében azért is indokolt, mivel a produkciós adatok is alátámasztják, hogy a zárszakasz időtartamát befolyásolja a zárhang képzési helye. Magyar nyelven [Gósy & Ringen \(2009\)](#) azt találta, hogy az átlagos záridőtartam a képzéshely hátrébb tolódásával csökkenő tendenciát mutat az intervokális zöngés explozíváknál izolált szavakban. [Grácsi \(2013\)](#) álszavakban mérte a zöngétlen zárhangok átlagos záridőtartamát, és szintén az előbb említett trendet támasztotta alá: 90(± 13) ms a labiális, 72(± 15) ms az alveoláris, valamint 70(± 13) ms a veláris explozíváknál. A rövid zöngétlen explozívák esetében spontán beszédanyagban is megjelent a tendencia: 79(± 11) ms a labiális, 71(± 18) ms az alveoláris, 63(± 18) ms a veláris mássalhangzók esetében [\(Neuberger, 2015\)](#). Ugyanebben a kutatásban a gemináták átlagos záridőtartama pedig: 115(± 20) ms-os, 122(± 31) ms-os és 106(± 27) ms-os átlagértékeket mutatott a képzéshellyel hátrafelé haladva. A percepció határt illetően is hasonló különbségeket vártunk a hangok között a képzéshely függvényében. A percepció eredmények a produkcióban megfigyelhető tendenciát mutatták a felnőtteknél: a képzési hely hátrébb tolódásával párhuzamosan a rövid-hosszú határ csökkenést mutatott [\(2\)](#) táblázat). A gyermekek percepció határa azonban a labiális explozívánál rövidebb záridőtartamnál jelent meg, mint a másik két képzéshely esetén.

2. táblázat. Záridőtartamok a rövid-hosszú percepció határnál a gyermekek és a felnőttek észlelésében

	[p]	[t]	[k]
Gyermekek	152 ms	164 ms	156 ms
Felnőttek	163 ms	161 ms	154 ms

3.2. Beszédészlelési különbségek a gyermekek életkora tekintetében

Az elemzés következő szakaszában különválasztottuk a 7 évesek és a 8 évesek eredményeit a korábbi szakirodalmi tapasztalatok okán, melyek szerint ez az az életkor, amely során jelentős változás tapasztalható az anyanyelvi fejlődés bizonyos területein, többek között a beszédhang-differenciálásban (lásd Bevezetés). Az így kialakított három csoport eredményeit a [6](#) ábrán szemléltetjük. Az adatokra épített általánosított lineáris kevert modell megerősítette, hogy az életkor befolyásolja a hallgatók válaszadását a záridőtartamok szerint: $F(2, 4624) = 16,774$, $p < 0,001$. A vizsgált életkorok közül a 7 évesek eredményei tértek el szignifikánsan a felnőttek eredményeitől ($p = 0,018$), de a gyermekek két csoportja, valamint a 8 évesek és a felnőttek eredményei között nem találtunk szignifikáns eltérést. Az adott hang képzéshelye függvényében szintén eltérő eredményeket kaptunk a különböző életkorú hallgatók döntéseit illetően: $F(6, 4624) = 4,283$, $p < 0,001$. A veláris hang esetén találtuk a legeggyöntebb válaszokat: a logisztikus függvények lefutása mind a hét- és a nyolcévesek, mind a felnőttek esetében hasonló képet mutat. A statisztikai elemzés a labiális és az alveoláris explozív esetén szignifikáns különbséget mutatott az életkori csoportok között a „hosszú” válaszok százalékos arányát tekintve (Friedman $\chi^2 = 8,553$, $df = 2$, $p = 0,014$; illetve $\chi^2 = 8,435$, $df = 2$, $p = 0,015$), a veláris explozív esetén azonban nem.



6. ábra. Válaszgörbék és percepciós határok életkoronként

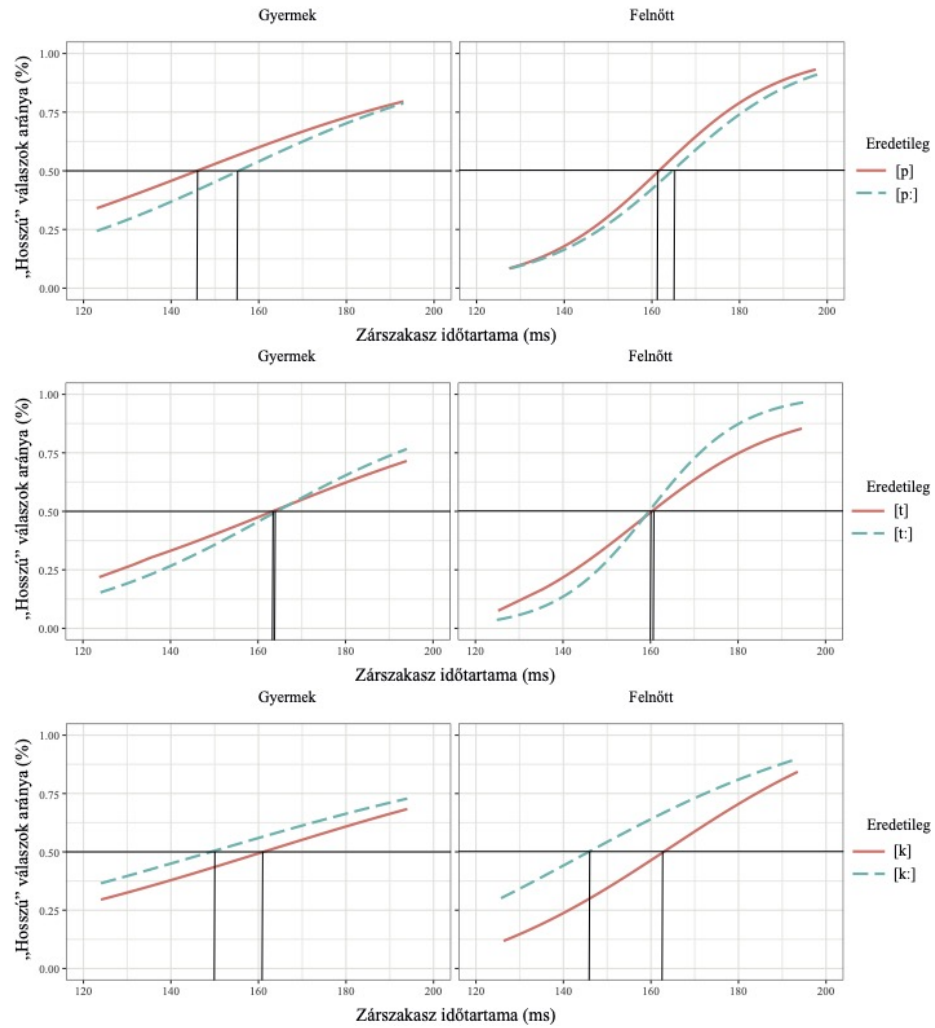
A rövid és a hosszú hangok közötti határ változó záridőtartam-értékeknél jelent meg a különböző életkorú hallgatóknál (3 táblázat). Korábban láthattuk már, hogy a felnőtteknél a képzéshely szerint hátrafelé haladva egyre rövidebb időtartamokat kaptunk. A gyermekek egyik csoportjában sem jelent meg a tendencia. Két szélsőséges értéket találtunk az eredményekben: a hétévesek percepció határa a [t] hang esetében hosszabb, a nyolcévesek percepció határa pedig a [p] hang esetében rövidebb volt, mint a másik két hangnál vagy másik két csoportnál mért értékek.

3. táblázat. Záridőtartamok a rövid-hosszú percepció határnál életkoronként

	[p]	[t]	[k]
Hétévesek	157 ms	175 ms	157 ms
Nyelcévesek	148 ms	154 ms	154 ms
Felnőttek	163 ms	161 ms	154 ms

3.3. Másodlagos felismerési kulcsok az észlelésben

Az elemzés utolsó lépésében megvizsgáltuk, hogy az a tényező, hogy az időtartam-manipuláció előtt az adott zárhang eredetileg szingleton vagy eredetileg gemináta volt, miként befolyásolja a hallgatók ítéleteit. Vajon a zárképzés időtartamának változtatása után marad-e valamilyen akusztikai jellemző, amely az eredeti hang hosszúsági jegyét magában hordozza, és amit a hallgatók felismernek az észlelés során? Ezt a kérdést azért érdemes körüljárni, mert a korábbi szakirodalmi eredmények (pl. Muller, 2003; Sato et al., 2012) felvetették a záridőtartam mellett további, másodlagos felismerési kulcsok lehetőségét az észlelésben. Ennek érdekében összesítettük a gyermek és felnőtt hallgatók válaszait különválasztva az eredetileg szingleton hangot tartalmazó hangszekvenciákat (iCi) az eredetileg geminátát tartalmazóktól (iCi). A 0:1-es adatokra illesztett válaszgörbéket és a percepció határt ismét a képzési hely szerint vizsgáltuk a gyermekeknél és a felnőtteknél (7 ábra).



7. ábra. Válaszgörbék és percepció határok életkoronként

Az eredetileg szingletonból manipulált beszédhangokra adott válaszok csupán kis mértékben különböztek az eredetileg geminátákból alkotott hangokétól a válaszgörbék alakját és elhelyezkedését megfigyelve. A statisztikai elemzés szerint a felnőtteknél a veláris hang esetében tértek el a válaszok szignifikánsak az eredeti kvantitás szerint (Wilcoxon $Z = -2,524$; $p = 0,012$). A gyermekek válaszaiban egyik hangnál sem volt kimutatható szignifikáns különbség.

A rövid-hosszú percepció határt hasonló záridőtartamoknál húzhatjuk meg az eredetileg szingeton és az eredetileg gemináta stimulus esetén is (4 táblázat). A gyermekeknél a [p] és [k] hangnál 10 ms-os eltérést, míg a [t] hangnál semmilyen eltérést nem találtunk a válaszgörbék 50%-os értékei között az eredeti kvantitás függvényében. A felnőtteknél a [p] és a [t] hangnál 3 ms, illetve 1 ms-os különbséget találtunk. A legnagyobb eltérés a felnőttek [k] hangra adott válaszaiban adatolható (15 ms-os eltérés): a geminátából képzett rövid záridőtartamú stimulusokra a hallgatók nagyobb arányban adtak „hosszú” választ, mint az ugyanilyen zárszakaszú szingletonból képzett stimulusokra. Ez azt sugallja, hogy a hallgatók még rövidebb záridőtartam esetén is hajlamosabbak voltak hosszúnak ítélni a hangot, amely a manipuláció előtt gemináta volt. A percepció határ a szingletonból alkotott stimulusoknál 162 ms, míg a geminátából létrehozott stimulusoknál 147 ms. Ez azt jelenti, hogy előbbi esetben 162 ms-tól ítélték inkább hosszúnak a hangot, míg az eredetileg gemináta hangszekvenciákat már 147 ms-tól nagyobb arányban tartották hosszúnak, mint rövidnek.

4. táblázat. Záridőtartamok a rövid-hosszú percepció határnál az eredeti hang kvantitása szerint (er. = eredetileg, szin. = szingleton, gem. = gemináta)

	[p]		[t]		[k]	
	Er. szin.	Er. gem.	Er. szin.	Er. gem.	Er. szin.	Er. gem.
Gyermekek	146	156	164	164	161	151
Felnőttek	162	165	161	160	162	147

Összesítve az eredményeket (képzéshelytől és eredeti kvantitástól függetlenül), a rövid és a hosszú zöngétlen zárhangok közötti percepció határ a gyermekek észlelésében a zárszakasznak a 146–164 ms-os tartományában, a felnőtteknél pedig a 147–165 ms-os záridőtartamoknál található a jelen kísérlet anyagán elemezve.

4. Következtetések

A jelen tanulmányban a mássalhangzó-hosszúság észlelését vizsgáltuk a zöngétlen explozívák záridőtartamának függvényében. A beszédészlelés kutatásában régóta központi kérdés, hogy a hallgatók a beszédjel mely jellemzőit használják fel ahhoz, hogy fonetikai ítéleteket hozzanak. A korábbi nemzetközi és hazai kutatási eredmények alapján feltételeztük, hogy a hallgatók kvantitásra vonatkozó ítéleteihez elégséges kulcs a zárképzés időtartama. Eredményeink megerősítették ezt a hipotézist a vizsgálatban részt vett gyermekeknél és felnőtteknél is. Mindkét életkori csoportban szignifikánsan befolyásolta a záridőtartam az explozívákra adott „rövid” vagy „hosszú” válaszokat a hallott hangszekvenciákban. A 100 ms és 200 ms közötti tartományban megfigyelhető volt, ahogyan a hallgatók észlelése a „rövid” kategóriából átvált a „hosszú” kategóriába mindhárom explozíva esetén. Jóllehet a hallgatók döntése nem minden esetben mutatkozott egyöntetűnek még a minimum és maximum záridőtartamok esetén sem.

A kutatás második és a harmadik hipotézise a gyermekek és a felnőttek észlelése közötti eltérésekre vonatkozott. Az adatokra illesztett válaszgörbék alakjában megmutatkozó különbségek, valamint a statisztikai eredmények is megerősítették ezeket a feltevéseket. A felnőtteknél megjelent S-alakú görbék azt jelezték, hogy a záridőtartam kontinuum mentén van egy időintervallum, amikor a hallgatók ítéletei átváltanak szingletonból geminátába, míg a relatíve rövid vagy hosszú időtartamok (főként a skála két végén található minimum és maximum értékek) egyöntetűbb döntést eredményeznek. Az akusztikum és a percepció közötti kapcsolat nem lineáris volta (Stevens, 1989) a jelen kutatás felnőttekre vonatkozó eredményeiben is jól megmutatkozott. A gyermekek válaszaira illesztett függvényekben azonban nem rajzolódott ki a tipikus S-alakú görbe, az ő eredményeik sokkal inkább lineáris elrendezést mutattak. Még a vizsgálatunkban alkalmazott időtartamskála minimum és maximum értékeinél sem született egyöntetű döntés a hosszúság megítélésében. A kvantitás kategóriális észlelése ezek szerint fejlődés eredménye, és a kategóriák megítélésében

a 7-8 éves kornál későbbi életkorban válik a felnőttekéhez hasonló szintűvé a magyar anyanyelvű gyermekeknél.

A szingletonok és a gemináták közötti határhoz kapcsolódó időtartamértékeket már több nemzetközi kutatás vizsgálta mind a produkciós, mind a percepció oldalról (pl. [Esposito & Di Benedetto, 1999](#), [Amano & Hirata, 2010, 2015](#)). Tudomásunk szerint nem született még olyan tanulmány, amely a gyermekek észlelésében vizsgálta a kategóriák közötti választóvonalat. Vizsgálatunk fényt derített arra, hogy – bár a rövid és hosszú zárhangok észlelésében különbség mutatható ki gyermekek és felnőttek között –, a kategóriák észlelési határa nem feltétlenül tér el a felnőtteknél megjelenő időtartamértékektől. A különbségek akkor váltak szembetűnővé a kísérletünkben, amikor a 7 és a 8 évesek eredményeit különválasztva elemeztük, de ekkor is maximum 15 ms-os különbség volt adatolható a zárszakaszok időtartamában a felnőttekhez képest. Mindent összevetve a rövid és hosszú kategóriák közötti percepció határ anyagunkban a 146 és 175 ms közötti tartományban nyilvánult meg: a labiális explozívánál 146–165 ms között, az alveoláris explozívánál 154–175 ms között, a veláris explozívánál pedig 147–162 ms között. Ezek az értékek összhangban vannak a nemzetközi kutatásokban talált percepció határvonal értékeivel (a zöngétlen explozívák záridőtartamára: 120–180 ms, pl. [Esposito & Di Benedetto, 1999](#), [Yoshida et al., 2015](#), [Zhou et al., 2021](#)).

A felnőtteknél a veláris explozíva esetében különbséget találtunk aszerint, hogy eredetileg szingleton vagy eredetileg gemináta szerepelt a hangsorban. Ebből arra következtethetünk, hogy a zárszakasz időtartamán túl más akusztikai jellemzők is szerepet játszhatnak a kvantitás jegyének felismerésében, amelyek a három vizsgált hang közül a [k] hang esetében bizonyult száliensnek. A hosszúsági oppozíció másodlagos akusztikai attribútumának tekinthető például a felpattanás intenzitása, amelyben a produkciós vizsgálatok alapján különbség figyelhető meg a rövid és hosszú felpattanó zárhangok között ([Neuberger & Beke, 2017](#)). Továbbá a szingletonokat és a geminátákat megelőző magánhangzók időtartamának produkcióban megmutatkozó különbsége alapján feltételezhetjük, hogy ez a paraméter is hozzájárul a kontraszt pontos észleléséhez ([Deme et al.,](#)

2019). Az eredetileg szingleton és gemináta hangok válaszgörbéinek ilyenfajta „elcsúszását”, azaz a percepciók határok különbségét találták más nyelvekben is. A [p] explozíva esetén mind a japán, mind a finn hallgatók válaszaiban megjelent az, hogy a szingleton/gemináta határ alacsonyabb záridőtartamhoz kapcsolódik abban az esetben, ha a manipuláció előtt a hang gemináta volt (Yoshida et al. 2015). A [t] explozíva esetén az eredeti geminátákból képzett stimulusokban szintén kisebb záridőtartamokhoz (167–168 ms körül) kapcsolódott az 50%-os válaszadás, mint az eredetileg szingletonoknál (170–175 ms körül) (Zhou et al. 2021).

A gyermekeknél egyik hangnál sem volt kimutatható a válaszokban különbség aszerint, hogy a stimulus szingleton avagy gemináta volt az időtartam-manipuláció előtt. Ennek egy magyarázata lehet az, hogy az ő percepciók működésük ebben az életkorban még kevésbé érzékeny ezekre a másodlagos felismerési kulcsokra, illetőleg hogy más felismerési kulcsokra támaszkodnak az egyes fonológiai megkülönböztető jegyek felismerésekor, mint a felnőttek. Szakirodalmi adatok alátámasztják, hogy a kisgyermekek a hallásalapú információkat nagyobb (szótag méretű) egységekben dolgozzák fel, az idősebb gyermekek és a felnőttek pedig kisebb egységekre (fonémákra) támaszkodnak. Bekövetkezik tehát egy fejlődésbeli váltás, amelynek során a gyermekek már másképp súlyozzák be az észlelés alapegységeit, ahogyan a növekvő nyelvi tapasztalattal párhuzamosan a percepciók stratégiáik is változnak (Nitttrouer & Lowenstein, 2009).

A mindennapi életben az egyes beszédhangok fonémaosztályokba sorolása ritkán korlátozódik csupán egyetlen beszédhangra, általában nagyobb egységekben történik, olykor egy teljes szó azonosítását követően jön létre. Ebben az esetben a hosszúsági jegy megítélését az akusztikai jellemzőkön túl a szó szemantikai tartalma is támogatja. A fonémadöntésre számos kontextuális tényező hatással van: ilyen a beszédhang helye a szótagban, a szótagszerkezet vagy a szomszédos beszédhangok jellemzői, amelyek viszonylatában az adott hang akusztikai sajátosságai (többek között az időtartamértékei is) változatos képet mutathatnak.

Továbbá a fonológiai fejlődésben meghatározó az adott nyelvben a hangelfordulások gyakorisága; a gyakran hallott kategóriák közötti kontrasztok előbb épülnek be a nyelvspecifikus elsajátítás során, mint a ritkábban hallottak (Anderson et al., 2003). Mivel a magyarban a kvantitásoppozíciónak a mássalhangzók körében alacsony a funkcionális terheltsége, disztinktív szerepe főként levezetett (és nem lexikális, morfémán belüli) alakokban, valamint hangutánzó, indulatszó, jövevéyszó vagy tulajdonnév tagú minimálpárokban előforduló szembenállásokra korlátozódik, ezért a kihasználtság, illetve a gyakoriság szempontjából periférikus jelenségről van szó. Ezzel állhat összefüggésben az, hogy a magyar gyermekek relatíve későn válnak képessé a mássalhangzókat hosszúság szerint fonématudatosan elkülöníteni a percepcióban. Holott az anyanyelvi oktatás-nevelés, az írás- és olvasástanítás során szükséges lenne, hogy a gyermekek beszédészlelése, fonológiai tudatossága elérjen egy olyan szintet, amely a sikeres tanulás alapjául szolgálhat.

Lényeges hangsúlyozni, hogy a jelen kísérletben kapott percepciós határok csak bizonyos megszorításokkal hasonlíthatók össze más vizsgálati eredményekkel. Eredményeink korlátozottsága részben a kutatáshoz felhasznált kísérleti anyagból fakad. Ez egyetlen női beszélő felolvasott hangszekvenciáin alapul, amelyben a magánhangzó spektrális-temporális jegyeit változatlanul hagytuk. Ebből kifolyólag egyrészt az artikulációs tempó szerinti különbségekre, illetve ezeknek a percepcióra gyakorolt hatására itt nem derülhetett fény. Másrészt, ismeretes, hogy a felpattanó zárhangok észlelése kontextusfüggő, a mássalhangzó és a környező szegmensek relációja további befolyásoló tényező lehet az kategóriák azonosításában. Jelenlegi anyagunkhoz csupán az [i] magánhangzó-környezetben ejtett zöngétlen explozívákat választottuk. A továbbiakban mindenképpen fontosnak tartjuk megvizsgálni nem csupán az abszolút időtartamok hatását a kategóriák közötti különbségek felismerésében, hanem a szingleton vs. gemináta mássalhangzók és a kontextus időviszonyainak szerepét az észlelésben – ahogyan azt a relációs akusztikai invariancián alapuló kutatások kiemelik. Mindez pontosabb képet adhat a statikus nyelvi kategóriák és a beszédprodukció, illetve beszédpercepció dinamikus működése közötti kapcsolatról.

Köszönetnyilvánítás

A kutatást a Bolyai János Kutatási Ösztöndíj és az MTA Posztdoktori Kutatói Pályázat támogatta.

Hivatkozások

- Amano, S., & Hirata, Y. (2010). Perception and production boundaries between single and geminate stops in Japanese. *The Journal of the Acoustical Society of America*, 128, 2049–2058.
- Amano, S., & Hirata, Y. (2015). Perception and production of singleton and geminate stops in Japanese: Implications for the theory of acoustic invariance. *Phonetica*, 72, 43–60.
- Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, 46, 155–182.
- Aoyama, K. (2000). *A psycholinguistic perspective on Finnish and Japanese prosody: Perception, production and child acquisition of consonantal quantity distinctions*. Boston/Dordrecht/London: Kluwer Academic Publishers.
- Blumstein, S., & Stevens, K. (1981). Phonetic features and acoustic invariance in speech. *Cognition*, 10, 25–32.
- Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1., retrieved 10 January 2020 from <http://www.praat.org>.
- Deme, A., Bartók, M., Grácsi, T. E., Csapó, T. G., & Markó, A. (2019). Gemináltak artikulációs szerveződése a magyarban. *Beszédkutatás*, 27, 54–74.
- Esposito, A., & Di Benedetto, M. G. (1999). Acoustical and perceptual study of gemination in Italian stops. *The Journal of the Acoustical Society of America*, 106, 2051–2062.

- Fowler, C. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14, 3–28.
- Gnanadesikan, A. (2004). Markedness and faithfulness constraints in child phonology. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (p. 73–108). Cambridge University Press.
- Gósy, M. (2006). A beszédhangok megkülönböztetésének fejlődése. *Beszédkutatás*, 14, 147–159.
- Gósy, M., & Horváth, V. (2006). Beszédfeldolgozási folyamatok összefüggései gyermekkorban. *Magyar Nyelvőr*, 130, 470–481.
- Gósy, M., & Ringen, C. O. (2009). Everything you always wanted to know about vot in hungarian. In *IXth International Conference on the Structure of Hungarian*. Debrecen, Hungary. URL: http://icsh9.unideb.hu/pph/handout/Ringen_Gosy_handout.pdf.
- Grácz, T. E. (2013). Explozívák és affrikáták zöngességének időviszonyai. *Beszédkutatás*, 21, 94–120.
- Gyarmathy, D., & Horváth, V. (2010). A beszédhallás szerepe a beszédhang-differenciálásában. *Gyógypedagógiai Szemle*, 38, 126–135.
- Gósy, M. (1995/2006). *GMP-diagnosztika*. Budapest: Nikol.
- Hankamer, J., Lahiri, A., & Koreman, J. (1989). Perception of consonant length: Voiceless stops in Turkish and Bengali. *Journal of Phonetics*, 17, 283–298.
- Hirata, Y., & Whiton, J. (2005). Effects of speaking rate on the single/geminate stop distinction in japanese. *The Journal of the Acoustical Society of America*, 118, 1647–1660.
- Jordanidisz, Á. (2015). *Magyar anyanyelvű gyermekek fonológiai tudatosságának fejlődése 4 és 10 éves kor között*. Budapest: Doktori disszertáció. ELTE.

- Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3–28.
- Khattab, G., & Al-Tamimi, J. (2013). Influence of geminate structure on early Arabic templatic patterns. In M. M. Vihman, & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches and cross-linguistic evidence* (p. 374–414). Cambridge University Press.
- Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279–312.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophic Transactions of the Royal Society B*, 369, 979–1000.
- Kunnari, S., Nakai, S., & Vihman, M. M. (2001). Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication*, 5, 13–24.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages* (Vol. 1012). Oxford: Blackwell.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology*, 54, 358–368.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Lisker, L. (1958). The Tamil occlusives: short vs. long or voiced vs. voiceless. *Indian Linguistics, Turner Jubilee*, 1, 294–301.
- Muller, J. S. (2003). The production and perception of word-initial geminates in Cypriot Greek. In *ICPhS-15*.

- Nagy, J., Józsa, K., Vidákovich, T., & Tibor-Fazekasné Fenyvesi, M. (2004). *DIFER Programcsomag: Diagnosztikus fejlődésvizsgáló és kritériumorientált fejlesztő rendszer 4–8 évesek számára*. Szeged: Mozaik Kiadó.
- Neuberger, T. (2015). Durational correlates of singleton-geminate contrast in Hungarian voiceless stops. In *ICPhS 2015*. URL: <http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0422.pdf>.
- Neuberger, T. (2016). Perception of consonant length opposition in Hungarian stop consonants. *The Phonetician*, 113, 6–23.
- Neuberger, T., & Beke, A. (2017). A zárfelpattanás spektrális jegyei a hosszúsági oppozíció függvényében. *Beszéd kutatás*, 25, 7–23.
- Nittrouer, S., & Lowenstein, J. H. (2009). Does harmonicity explain children’s cue weighting of fricative-vowel syllables? *The Journal of the Acoustical Society of America*, 125, 1679–1692.
- Obrecht, D. H. (1965). Three experiments in the perception of geminate consonants in Arabic. *Language and Speech*, 8, 31–41.
- Olaszy, G. (2006). Hangidőtartamok és időszerkezeti elemek a magyar beszédben. *Nyelvtudományi Értekezések*, 155.
- Payne, E., Post, B., Garmann, N. G., & Simonsen, H. G. (2017). The acquisition of long consonants in Norwegian. In H. Kubozono (Ed.), *The phonetics and phonology of geminate consonants* (p. 130–162). Oxford University Press volume 2.
- Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica*, 56, 135–157.
- Ridouane, R. (2010). Geminate at the junction of phonetics and phonology. *Papers in laboratory phonology*, 10, 61–90.

- Sato, Y., Kato, M., Mazuka, R., & R. (2012). Development of single/geminate obstruent discrimination by Japanese infants: Early integration of durational and nondurational cues. *Developmental Psychology*, 48, 18–34.
- Siptár, P., & Grácz, T. E. (2014). Degemination in Hungarian: Phonology or phonetics? *Acta Linguistica Hungarica*, 61, 443–471.
- Slifka, J. (2006). Acoustic cues, landmarks, and distinctive features: a model of human speech processing. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, 2, 91–96.
- Sonu, M., Kato, H., Tajima, K., Akahane-Yamada, R., & Sagisaka, Y. (2013). Non-native perception and learning of the phonemic length contrast in spoken Japanese: training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics*, 22, 373–398.
- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of phonetics*, 17, 3–45.
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, 111, 1872–1891.
- Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64, 1358–1368.
- Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, 55, 653–659.
- Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*, 90, 1309–1325.

- Tsukada, K., Cox, F., Hajek, J., & Hirata, Y. (2018). Non-native Japanese learners' perception of consonant length in Japanese and Italian. *Second Language Research*, 34, 179–200.
- Vihman, M. M., & Velleman, S. L. (2000). The construction of a first phonology. *Phonetica*, 57, 255–66.
- Wilson, A., Kato, H., & Tajima, K. (2005). Native and non-native perception of phonemic length contrasts in Japanese: Effects of speaking rate and presentation context. *The Journal of the Acoustical Society of America*, 117, 2425–2425.
- Yoshida, K., Jong, K. J., Kruschke, J. K., & Pääviö, P. M. (2015). Cross-language similarity and difference in quantity categorization of Finnish and Japanese. *Journal of Phonetics*, 50, 81–98.
- Zhou, Y., Nakamura, Y., Mugitani, R., & Watanabe, J. (2021). Influence of prior auditory and visual information on speech perception: Evidence from Japanese singleton and geminate words. *Acoustical Science and Technology*, 42, 36–45.

Az alapfrekvencia jellemzőinek longitudinális összevetése felnőtt beszélők felolvasásában

Markó Alexandra¹, Huszár Anna^{2,3}, Krepsz Valéria^{3,4}, Grácsi Tekla Etelka³

¹*ELTE BTK Alkalmazott Nyelvészeti és Fonetikai Tanszék*

²*ELTE Nyelvtudományi Doktori Iskola*

³*Nyelvtudományi Kutatóközpont*

⁴*Humboldt University of Berlin*

Abstract

Speech changes with age, in several aspects. With respect to fundamental frequency (f0), there is a wealth of cross-sectional analyses of changes. However, longitudinal analyses are rare, due to the methodological difficulties, even though these studies are able to control many of the variables which are not controlled in most of the cross-sectional analyses. In the present paper, the change of the fundamental frequency (f0) was studied in young adults' speech within a ten-year-long interval. Most studies that describe f0 across lifespan embrace larger time intervals, however, the observations on short-term changes can be applied with various respects. The present study analysed the f0 characteristics in 20 young adults' sentence reading with ten year difference between the recordings of the same healthy speakers. At the time of the first recording, the speakers (10 males and 10 females) were at the age of 20-45 years, while they were between 30 and 55 years at the second recording's time. Besides the mean f0, the overall f0 range of the recordings and the utterance level f0 range were measured.

The results showed that the f0 mean decreased in female speakers after 10 years, but no clear tendency was found in male speakers within the analysed 10 years. The overall f0 range of the recordings did not change in the analysed time interval, however, at the utterance level, the f0 range was higher in the second recording. The data showed considerable interspeaker variability. The results (and the database) are relevant for forensic phonetic applications and can serve as baselines for speech pathology research.

Keywords: fundamental frequency, longitudinal changes, f0 range

1. Bevezetés

Jól ismert, hogy az alapfrekvencia (f0) az életkor előrehaladtával folyamatosan változik. Általánosságban elmondható, hogy kisgyermekkorban a beszélők alapfrekvenciája magasabb, mint fiatal felnőttkorban, és az első néhány évben

Email addresses: marko.alexandra@btk.elte.hu (Markó Alexandra),
huszar.anna@nytud.hu (Huszár Anna), krepsz.valeria@nytud.hu (Krepsz Valéria),
graczi.tekla.etelka@nytud.hu (Grácsi Tekla Etelka)

nem mutatható ki különbség a két nem között. A fiúk és lányok alapfrekvenciájának elkülönülése körülbelül pubertáskorra tehető, amikor a fiúk alapfrekvenciája jelentősen, a lányoké pedig kisebb mértékben csökken (Pedersen et al. 2015, Berger et al. 2019), így a pubertáskor után és a fiatal/középkorú felnőttkorra az általánosan tapasztalható tendenciát hozva létre: a férfiak alapfrekvenciája alacsonyabb lesz a nőkéénél (összefoglalóan lásd pl. Traunmüller & Eriksson 1995). Időskorban szintén jelentős változások feltételezhetők, általánosan a következő tendenciák megjelenése valószínűsíthető: a nők alapfrekvenciája csökken, a férfiaké pedig valamelyest növekszik, azaz a két nem beszélőinek f_0 -értékei közelítenek egymáshoz (Brown et al. 1991, de Pinto & Hollien 1982). Kérdés azonban, hogy a fiatalok felnőtt beszédének alapfrekvenciája, amelyhez általában viszonyítják a más életkorokban kapott adatokat, valóban állandónak és változatlanul tekinthető-e, vagy a hang esetleg ebben a rövidebb időintervallumban is változik-e. Ugyanis ha a fiatalok felnőtt beszédében is található változás az alapfrekvencia jellemzőit tekintve, akkor a nagyobb életkori távlatban tett összevetéseknél ezt is figyelembe kell venni.

A különböző szakirodalmi eredmények sokszor ellentmondásosak abban a tekintetben, hogy a fiatalok és a középkorú felnőttek beszédében mérhető-e változás az egyénen belül vagy eltérés az (egymáshoz közelebb álló) életkori csoportok között (vö. pl. Nishio & Niimi 2008 vs. Cox & Selent 2015).

Az elemzések során azt is figyelembe kell venni, hogy az alapfrekvenciát a beszélő életkora mellett más tényezők is meghatározzák, így például a különböző fiziológiai változások, amelyeket az elsődleges és a másodlagos öregedés vált ki (Busse 2002). Az elsődleges öregedés az életkor előrehaladása miatt bekövetkező szükségszerű fiziológiai változás, míg a másodlagos öregedés az egyén betegségei, szokásai/életvitele (pl. dohányzás, sport) és a külső tényezők (pl. szmog) hatására lezajló változások. Ez utóbbit további kategóriákra is bontják a betegségek és az egyéb jellemzők mentén. A két öregedési hatás önmagában is változatosan jelenik meg az egyének között (Belsky et al. 2015). Emellett a különböző módszertani (pl. a felvétel körülményei, az adott feladat) és a beszélők

között kimutatható jelentős egyéni különbségek is magyarázatként szolgálnak az eltérésekre.

1.1. Keresztmetszeti vizsgálatok

Hollien & Ship (1972) egy korai vizsgálatában 175 férfi beszélő felolvasását elemezte keresztmetszeti vizsgálati módszertan segítségével, azaz különböző életkorú beszélői csoportok összevetésével (20 és 90 év között). Eredményeik szerint kora felnőttkorban, 20 és 40 év között, magasabb átlagos alapfrekvencia-értéket kaptak, mint az idősebbek beszédében (112 Hz vs. 107 Hz), majd később, 60 és 80 éves kor között egyre magasabbakat, 70–79 éves kor között átlagosan 132 Hz, 80–89 éves kor között 146 Hz-es átlagos alapfrekvencia-érték volt megfigyelhető. Hasonló eredményeket: eleinte egyre alacsonyabb értékeket, majd stagnáló, később egyre magasabb értékeket adatoltak más férfi beszélőknél is. Stathopoulos és munkatársai (2011) 200 beszélőt vizsgáltak egy szinte teljes életen átívelő – 4 és 93 év közötti beszélőkkel elvégzett – keresztmetszeti vizsgálatban. 4 és 20 éves kor között az életkorral előrehaladva egyre alacsonyabb értékeket mértek. 20 és 50 év közötti férfiaknál ugyancsak csökkentek az értékek az életkor előrehaladtával, de ezek az eltérések kisebb mértékűek voltak, mint a 4 és 20 év közöttiek esetében. Az 50 év feletti beszélőik esetében az életkor mentén növekvő értékeket kaptak. A szerzők nem közöltek konkrét adatokat az egyes korosztályok esetében, a bemutatott ábra (Stathopoulos et al. 2011: 1015) alapján a legmagasabb értékek a fiúgyermekek esetében érték a 350 Hz-et, majd felnőttkorra több adatközlő átlagos f_0 -értéke is 100 Hz alá csökkent. Ehhez képest némely idős beszélő átlagértékei 150 Hz fölötti értéket értek el. Schötz (2006) is változó tendenciát igazolt 268 férfi eredményeinek összevetése alapján, ahol 20-tól 50 éves korig egyre alacsonyabb értékeket mért (átlagosan 120 Hz vs. 112 Hz), 50 éves kor felett pedig egyre magasabbakat (a legidősebb, átlagosan 90 éves beszélőknél 143 Hz-es átlagot mértek). Nishio és Niimi (2008) 374 egészséges beszélő (azonos számú nő és férfi) alapfrekvenciáját vizsgálta tinédzserkor és idős kor között. A férfiaknál ők is egyre alacsonyabb átlagos alapfrekvencia-értékeket találtak időskorban (a 19 és 69 éves kor között 119–123

Hz-esnek mért átlagos alapfrekvencia 80 és 89 éves kor között elérte az átlagosan 136 Hz-es értéket), ám ez csupán 70 éves kor fölött mutatkozott meg.

Ezzel szemben Eichhorn és munkatársai (2018), akik Hollienékhez hasonlóan 96 (43 férfi és 53 női), 20 és 92 év közötti angol beszélő alapfrekvenciájának értékét vizsgálták 3 életkori csoportban felolvasásban és egy szótagos szavakban, a fenti csökkenést mutató eredményekkel szemben már a két fiatalabb korcsoport esetében is az idősebb esetében magasabb értékeket találtak a férfiaknál az életkor előrehaladtával (20–30 éves kor között: 102–112 Hz, 40–60 éves kor között: 108–115 Hz, 70–92 éveseknél: 119–126 Hz). Szintén magasabb átlagos alapfrekvencia-értéket igazolt az idősebb vizsgálati csoport esetében Debruyne és Decoster (1999). Ebben a vizsgálatban a fiatalabb korosztály, azaz 40, életkorát tekintve 21;3–29;5 év közötti férfi átlagos alapfrekvencia-értéke 112 Hz volt; az idősebb beszélőké, azaz 60;7–94;3 év közötti életkorú 40 férfié 125 Hz volt. Sebastian és munkatársai (2012) 60 és 80 éves kor között, illetve Torre és Barlow (2009) átlagosan 25 és 75 éves beszélők értékeinek összevetésekor ugyancsak magasabb átlagos alapfrekvencia-értékeket talált az idősebb korosztályokban: a fiatalabb beszélők (20–35 év) átlagértéke 114 Hz, az idősebbeké (60–89 év) 142 Hz volt. A korábbi eredményekkel szemben Cox és Selent (2015) fiatalabb korosztályok között egyre magasabb, idősebb korosztályok között egyre alacsonyabb értékeket kapott 35 férfi beszélő esetében, akiket öt életkori csoportba osztottak (20–29 éves kor között átlagosan 122 Hz-et mértek, 30–39 év között 128 Hz-et, a 40–49 éves korcsoportban 117 Hz-et, 50 és 59 éves kor között 114 Hz-et, 60–69 éveseknél pedig 113 Hz-et, azonban a beszélők létszáma nem volt azonos).

A női beszélőknél 60 éves korig fokozatosan egyre alacsonyabb alapfrekvencia-értékeket mértek az életkor előrehaladtával a következő kutatások: Nishio és Niimi (2008) 19 éves korban 226 Hz vs. 89 éves korban 168 Hz), Debruyne és Decoster (1999) a 22–28 évesek átlagértéke 197 Hz, a 60–90 éves beszélőké 180 Hz volt), Kaur és Narang (2015) 5–80 év között, az átlagértékek 388 Hz és 170 Hz között változtak), Sebastian és munkatársai (2012) 60–80 év között 136–145 Hz közötti átlagos f0 ötévenkénti korcsoportokban), Torre és Barlow (2009) a

fiatalabb beszélők átlaga 204 Hz, az idősebb beszélőké 180 Hz volt), Eichhorn és munkatársai (2018) 20–30 éves korban: 200–212 Hz, 40–60 évesen 165–178 Hz), valamint Stathopoulos és munkatársai (2011) is. Ez utóbbi tanulmány pontos adatokat nem közöl, de a 2. ábrájuk (Stathopoulos et al. 2011: 1015) alapján a legfiatalabb korosztálynál (a 4 éves beszélőknél) mért f_0 értéke 250–260 Hz között volt, a 90 éves kort követően pedig átlagosan 150–170 Hz körül valósult meg, a trendvonalak szerint pedig folyamatos csökkenés mutatható ki az életkor előrehaladtával. Eichhorn és munkatársainak vizsgálatában a 60 évnél idősebbek esetében az életkor növekedésével egyre magasabb f_0 -értéket mértek (70 éves kor felett 166–188 Hz), míg Stathopouloséknál nem volt kimutatható egységes tendencia az idősebb beszélőknél.

Nishio és Niimi (2008) nem csupán az f_0 -átlagértékeket, de az egyénenkénti legalacsonyabb és legmagasabb f_0 -érték különbségét is elemezte. Eszerint a férfiaknál fiatal felnőttkorban 66 Hz, idősebb korban 62 Hz, a legidősebb vizsgált korcsoportnál pedig 73 Hz hangtartományt adatoltak, míg a nőknél ugyanezen sorrend szerint 82, 64 és 77 Hz-es tartományt.

Látható tehát, hogy az átlagos alapfrekvencia értéke már fiatal és középkorú felnőttek esetében is eltérést mutat az életkor mentén, egyes kutatások szerint a 20 és 40 éves kor között igazolódó tendencia eltér a későbbi, idősebb korban tapasztaltaktól.

1.2. Longitudinális vizsgálatok

A jelentős egyéni különbségek, valamint a módszertani nehézségek kiküszöbölése érdekében a keresztmetszeti kutatások mellett az alapfrekvencia változását longitudinális módszertan segítségével is vizsgálták. Az igazolt tendenciák – a keresztmetszeti kutatásokkal szemben – közel egységesnek tekinthetők: a hosszútávú idővizsgálatok csökkenő tendenciát igazoltak a beszélők f_0 -értékeiben a korábbi és a későbbi felvételek összevetése alapján. Ennek feltehetően részben az az oka, hogy a keresztmetszeti kutatások általában nem vagy kevéssé kontrollálnak olyan tényezőket, amelyek bizonyítottan hatással vannak az alapfrekvencia alakulására. A longitudinális mérések esetében az egyénen

belüli változásokat az idősödés mint alapvető hatás befolyásolja. A longitudinális kutatások elvégzése jóval nehezebb, hiszen hosszú távú elköteleződést kíván meg minden résztvevőtől, a kutatókat is beleértve, és az életútban bekövetkező változások nem mindig teszik lehetővé az egyes egyének újbóli bevonását az egymást követő vizsgálati időszakokban. Ennélfogva longitudinális kutatásból kevesebbet ismerünk, és ezek jellemzően kisebb beszélőszámot vonultatnak fel, változó időintervallumokban.

50 év távlatából (1945-ben 28 fővel, 1981-ben közülük 5 fővel, majd 1993-ban 15 fővel készített felvételek alapján) hasonlították össze 15 ausztrál nő hangmin-táját Russel és munkatársai (1995). Az első felvétel alkalmával a beszélők 18–25 évesek, a másodiknál 53 és 56 év közöttiek, a harmadiknál pedig 65–68 évesek voltak. Az életkori szóródásbeli eltérés azzal magyarázható, hogy az eredeti adatbázis 28 fővel készült, később azonban csak közel egyhatodukat, még később pedig az eredeti adatközlők közel felét tudták elérni. Az eredmények szignifikáns csökkenést mutattak az életkor előrehaladtával az első és az utolsó felvételek között, amelyet minden beszélőnél alátámasztottak az adatok. Ezzel szemben a második és a harmadik felvétel között jelentős eltérések mutatkoztak a beszélők között a változás irányában: 3 beszélőnél kismértékű növekedés/stagnálás, 2 beszélőnél kismértékű csökkenés mutatkozott. Egy másik kísérletben Harrington és munkatársai (2007) II. Erzsébet királynő „Christmas broadcasts” korpuszát (az 1950-es évektől a 2000-es évekig elmondott karácsonyi köszöntők felvételeit) elemezték. Ennek fontos előnye, hogy a jó minőségű rögzítés mellett ugyanaz a beszéd stílusa, és tartalmaz minden alkalommal megismétlődő mondatot is, így az alaphérfkvencia elemzése viszonylag állandó körülmények között valósulhatott meg. Harrington és kollégái megállapították, hogy az életkor előrehaladtával csökkent az alaphérfkvencia értéke 253 Hz-ről 189 Hz-re a közel 50 év alatt. Ezt a beszédképző szervek folyamatos öregedésével és a fiziológias változásokkal magyarázták.

Szintén II. Erzsébet királynő karácsonyi beszédeit alapul véve igazolták Mwangi és munkatársai (2009), valamint Reuboldt és munkatársai (2010) az alaphérfkvencia fokozatos csökkenését a beszélő 25 és 75 éves kora között. A legjelentő-

sebb változás körülbelül 25 és 45 éves kor között, valamint 70 éves kor fölött volt megfigyelhető a két kutatás eredményei alapján. Egy másik kutatásban (Harrington et al. 2007) II. Erzsébet királynő hanganyagát másik három beszélő értékeivel vetették össze 24–50 év távlatából. Az egyikük Margaret Lockwood, a BBC rádióbemondója volt, továbbá egy ausztrál férfi beszélő és egy új-zélandi férfi beszélő (az utóbbiakat nem nevezte meg a tanulmány). Sajnos a kutatásból levonható következtetések hatását csökkenti, hogy a beszélőkkel eltérő körülmények között rögzítették az egyes felvételeket. Emellett a beszélők más életkorúak voltak (Erzsébet királynő hanganyagát 26, illetve 76, Margaret Lockwood-ét 40, illetve 64 éves, az ausztrál férfi hanganyagát 40 és 79 éves, az újzélandi férfi közléseit pedig 36 és 73 éves kora között rögzítették különböző időpontokban), illetve más típusú professzionális beszélőknek tekinthetők (uralkodó, rádiós, illetve televíziós bemondók), más időközönként rögzítették a hanganyagaikat (50, 24, 39 és 37 év elteltével), és más adatbázisból származtak a felvételek. Az eredmények azonban azt mutatták, hogy mindettől függetlenül mind a négy beszélő esetében az alaphfrekvencia csökkenése volt megfigyelhető.

Ezzel szemben Decoster és Debruyne (2000) 20 holland nemhivatásos férfi beszélő hanganyagának összevetése során jelentős eltéréseket talált az életkor mentén az egyes beszélők alaphfrekvencia-értékei között. Noha az átlagértékek 14 Hz-cel csökkentek, valójában 5 beszélőnél növekedés, míg 15 főnél csökkenés volt kimutatható.

1.3. A jelen kutatás célkitűzése és hipotézisei

Ahogy mind a keresztmetszeti, mind a longitudinális módszertannal készült vizsgálati eredmények alapján látható, a változás egész életünk során megfigyelhető, az időskor felé mutató, nemenként, sőt sokszor beszélőnként is eltérő tendencia már fiatal felnőttkorban megállapítható. Ennek mértéke, iránya, módja vizsgálatonként eltérő lehet, így ezen eredmények alapján az sem állapítható meg teljes bizonyossággal, hogy vajon mely életkortól kezdődően figyelhető meg változás a fiatal beszélők esetében.

Összefoglalóan elmondható, hogy a különböző kutatások gyakran egymásnak is ellentmondó eredményekre jutottak az alapfrekvencia értékeinek alakulását tekintve fiatal felnőtt beszélők esetében. Míg a longitudinális kutatások többsége viszonylag ritkán vizsgálta a beszélők jellemzőit fiatal korban, illetve rövid intervallumon belül, továbbá a módszertani sokszínűség is nehezíti az összevetést, a vizsgálatok többsége csökkenő tendenciát igazolt fiatal felnőttkorban (elsősorban II. Erzsébet királynő eredményei alapján). Ezzel szemben a keresztmetszeti kutatások esetében egymásnak ellentmondó eredmények születtek: míg az egyes kutatások esetében 20 és 40 éves kor között mutatkozott a legmeredekebb csökkenés (vö. pl. [Hollien & Ship, 1972](#)), mások növekedést (vö. pl. [Stathopoulos et al., 2011](#); [Eichhorn et al., 2018](#)) adatoltak.

A jelen kutatás célja felmérni fiatal felnőtt (20 és 40 év közötti), nem professzionális beszélők alapfrekvenciájának változását 10 éves távlatból, amely eredmények többek között a beszélőazonosítás szempontjából is meghatározóak lehetnek. Decoster és Debruyne ([2000](#)), Harrington és munkatársai ([2007](#)), valamint Reubold és munkatársainak ([2010](#)) korábbi eredményei alapján a vizsgálat megkezdése előtt azt feltételeztük, hogy (i) az átlagértékekben kismértékű eltérések, ám az egyes beszélők között jelentős individuális különbségek mutatkoznak meg az alapfrekvencia értékének változásában.

A nők esetében gyakoribb volt, hogy a korábbi kutatások csökkenő tendenciát találtak (de náluk is változó volt az eredmény a tanulmányok között), így azt feltételeztük, hogy (ii) a nők felolvasásában az alapfrekvencia csoportszinten csökken, de nagy egyéni különbségeket vártunk. A férfiak esetében sokkal több ellentmondó eredményt adtak ebben az életkorban a fenti elemzések, ezért feltételeztük, hogy (iii) nem mutatható ki az átlagos alapfrekvencia egységes csökkenése vagy növekedése. Az alapfrekvencia átlagos értékein túl elemeztük a hangterjedelmet, azaz a beszédmintában produkált legmagasabb és legalacsonyabb frekvenciaérték arányát, illetve a hangközt, az egy-egy mondaton belül mérhető legmagasabb és legalacsonyabb alapfrekvencia arányát (félhangban). (iv) Ezekben az értékekben nem vártunk csoportszinten igazolható eltérést a tízéves intervallumban.

2. Kísérleti személyek, anyag és módszer

2.1. Kísérleti személyek

A vizsgálathoz a Longitudinális korpuszból (Grácsi et al., 2020) választottunk adatközlőket. A Longitudinális korpuszt a BEA adatbázis (Neuberger et al., 2014) beszélőinek 10-11 évvel később újra rögzített beszédmintái alkotják, azaz rendelkezésre áll két olyan felvétel minden beszélőtől, amelyek között egy évtized telt el. A korpusz jelenleg 25 fő felvételét tartalmazza. Közülük választottunk 10 nőt és 10 férfit, akik fiatal vagy középkorú beszélők voltak az első felvétel idején. A BEA adatbázis felvételének készültekor az életkoruk 20 és 45 év közötti, átlagosan $26,4 \pm 6,2$ év volt. (A tanulmányban az átlag $\pm$ szórás alakot használjuk.) A férfiak, 20 és 34, a nők 21 és 45 év közöttiek voltak az első felvétel idején, és 30–44 illetve 31–55 évesek a második felvétel készültekor (a 45 éves nő eredményei nem tértek el a csoportra jellemző tendenciáktól, ezért nem zártuk ki az elemzésből). A beszélőknek sem hallásproblémájuk, sem beszédhibájuk nem volt a felvételek készültekor.

2.2. Anyag

Mind a BEA adatbázis, mind a Longitudinális korpusz protokolljában szerepel mondatfelolvasási feladat. A BEA-felvételeken 25 mondatot olvastak fel az adatközlők, melyek közül az újabb felvételek során 15 szerepelt ismét. Ezt a 15, eltérő hosszúságú és összetettségű mondatot elemeztük a vizsgálatban az alábbi szempontok szerint. A Longitudinális korpusz esetében egy olyan kérdőívet is kitöltöttek az adatközlők, amelyben a beszédüket esetlegesen befolyásoló tényezőkre kérdeztünk rá. Ezek az alábbiak: beszédproblémák, hangterhelés, fül-orr-gégészeti, mozgásszervi, hormonális, neurológiai, krónikus légzőszervi megbetegedések, reflux, gyógyszerszedés, dohányzás, rendszeres alkoholfogyasztás, átlagos folyadékfogyasztás.

2.3. Módszer

A 15 mondat felolvasását a Praat szoftverben (Boersma & Weenink, 2020) annotáltuk szakaszszinten. A félreolvasást, irreguláris zöngét vagy zajt tartal-

mazó szakaszokat kizártuk az elemzésből (egy olyan mondat volt egy beszélő egyik hangfelvételén, amelyet teljesen ki kellett zárunk az elemzésből zajosság miatt). Az alaphfrekvencia-mérés határértékeit egy előzetes mérés alapján állítottuk be 75 és 600 Hz közé, az ablak hosszát pedig a Praat alapbeállításával (0,04 s) tartottuk meg. Ezen első mérést követően átnéztük a mondatokra kapott alaphfrekvencia-meneteket. Az ellenőrzés alapján a 400 Hz feletti értékek oktávugrással jöttek létre, ezért ezeket megfeleztük. Ezek után az automatikus mérés eredményeit minden egyes mondat esetében ellenőriztük, és szükség szerint javítottuk a beszélőre nem jellemző tartományba eső értékeket. A nők esetében ez a 150 Hz alatti, egy nő (N01) esetében a 100 Hz alatti értékek eltávolítását jelentette, a férfiaknál pedig a 250 Hz feletti (F01, F02, F03, F06, F08, F09), 200 Hz feletti (F04, F05, F010) vagy 160 Hz feletti (F07) értékek eltávolítását. További egy férfi esetében a 100 Hz alatti értékeket is eltávolítottuk (F04).

Az alaphfrekvencia teljes tartományának eloszlását és jellemzőit vizsgáltuk. Az átlagos alaphfrekvencia értékét a teljes felvételen és mondatonként is, a hangterjedelmet (a jelen esetben az adott felvételen mért legmagasabb és legalacsonyabb f_0 -érték eltérése félhangban) és a hangközt (az adott felvételkor az adott mondatban mért legmagasabb és legalacsonyabb érték eltérése félhangban) számítottuk ki és vetettük össze a beszélők két felvételkori beszédmintáiban. A minimum- és maximumértékek összevetésére a félhangban való számítást választottuk (a korábbi kutatásokban alkalmazott frekvenciakülönbségek számítása helyett), mert így az értékek az arányokat mutatják, és így a hangok közötti különbségek pszichoakusztikai képzetét reprezentálják.

A legmagasabb értéket a mondat első negyedében mértük, mivel egy mondat kivételével (F07 beszélő második felvételének m5-ös mondata – itt a mondat második feléből mértük a csúcs értékét) ebben a tartományban jelent meg a dallamcsúcs. Az alaphfrekvencia-minimumot pedig az ez utáni tartományban mértük. Az időbeli elkülönítésre azért volt szükség, mert egy-egy kiugró, de nem frekvenciacsúcsához tartozó érték megjelenhet zöngétlen mássalhangzók környezetében is. Ezek olyan téves adatok, amelyek értéke a mondat dallamcsúcs-értékéhez

közel eshet, de akár meg is haladhatja. A félhangszámításokat a HQmisc csomaggal végeztük (Quené, 2014). A hangterjedelem és a hangköz esetében a legmagasabb értéket viszonyítottuk a legalacsonyabb értékhez, az alaphfrekvencia átlagának a két felvételi időpont közötti eltérésének vizsgálatában pedig a második felvételi (azaz későbbi) érték félhangbeli értékét számítottuk ki az első felvételi (azaz korábbi) felvétel értékéhez képest. Így ez utóbbi számításban a negatív érték alacsonyabb, a pozitív érték magasabb alaphfrekvencia-átlagot jelent a második felvételen mért adatokban.

Az így kapott adatokat az R szoftverben (R Core Team, 2020) elemeztük. Mivel a mért értékeket kézzel ellenőriztük, további adattisztítást nem végeztünk. A mondatonkénti átlagos alaphfrekvencia és a hangköz esetében beszélőnként 15 adat (egy nő egy felvételére 14) állt rendelkezésre felvételi időpontonként, ezekre lineáris kevert modelleket készítettünk az lme4 csomagban (Bates et al., 2015), melyek összevetését és a p -érték számítását az lmerTest csomaggal (Kuznetsova et al., 2017) végeztük Satterthwaite-approximációval. Az R^2 értékét a MuMIn (Barton, 2020) és az afex (Singmann et al., 2021) csomagok használatával nyertük ki. A legbővebb modelltől a kisebb felé haladtunk az elemzésekben, azaz a nemet és felvételi időt tettük fix hatásnak a legbővebb modellben, a szűkítés során pedig ezek közül hagyttunk el egyet-egyet, illetve mindkettőt. A fix hatásokra random meredekséget illesztettünk. A legszűkebb modell random konstans tartalmazott. Az lmerTest csomag *anova()* módszerével vetettük össze a modelleket, és ha a bővebb és a szűkebb szignifikánsan eltért, a bővebbet tartottuk meg. A mondatokra random konstans, a felvételre beszélőnkénti random meredekséget állítottunk be, és összevetettük az egyszerűbb, a mondatra és a beszélőre is csak random konstans tartalmazó modellel az lmerTest *anova()* próbájával. Amennyiben nem volt szignifikáns eltérés, az alacsonyabb AIC-számú modellt tartottuk meg, vagyis az ezen mérőszám alapján kisebb előrejelzési hibával rendelkező, jobb minőségű modellt (Akaike, 1974). A kiinduló modellben függő változóként a mondatonként mért átlagos alaphfrekvencia, illetve a hangköz szerepelt, fix hatásként a nem és a felvételi időpont, azok interakcióját is megengedve.

A teljes felvételre kapott átlagos alapfrekvencia-értékek és a hangterjedelem esetében beszélőnként egy-egy érték állt felvételenként rendelkezésre, ezért ezeket ismételt méréses próbával vetettük össze. Mivel egyik változó sem volt normál eloszlású, nem parametrikus próbát választottunk. Az ismételt mintás ANOVA nonparametrikus „megfelelője” a Scheirer–Ray–Hare-próba (rcompanion csomag: [Mangiafico, 2021](#)). Ennek azonban feltétele, hogy a mérési eredmények között ne álljon fenn korreláció. Ezt Spearman-féle rangkorrelációval ellenőriztük (*cor.test()* függvénnyel). Az alapfrekvencia-átlagok esetében fennállt korreláció, a hangterjedelem esetében azonban nem. Ezért a hangterjedelmet a Scheirer–Ray–Hare-próbával vetettük össze, ahol a hangterjedelem szerepelt függő változóként, fix faktorként pedig a nem és a felvétel interakciója. Az alapfrekvencia-átlagokat pedig Wilcoxon-próbával (a páros *t*-próba nemparametrikus megfelelője) vetettük össze. Ebben csak egy fix faktor adható meg, ezért nemenként készítettük el az elemzést.

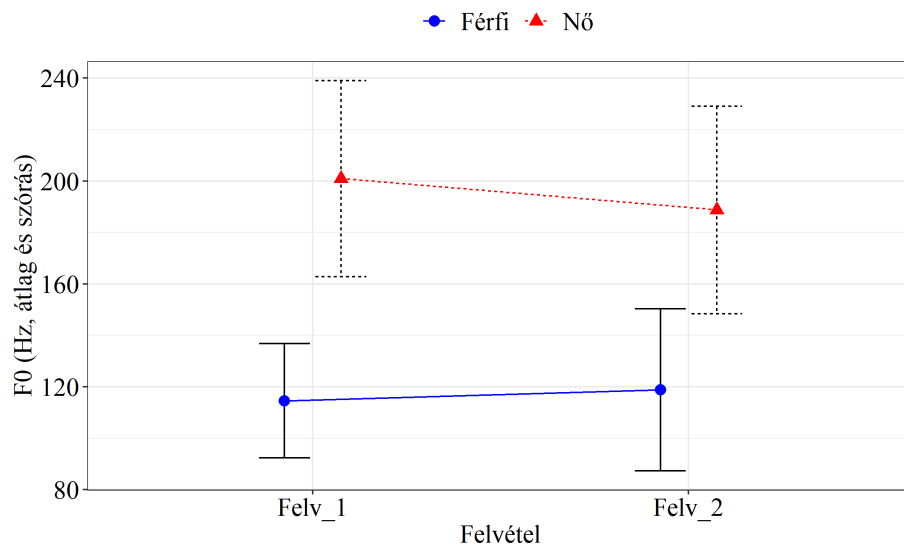
A beszélők eredményeit egyesével is elemeztük, összevetettük, kiknél kaphattunk az adott nemű beszélőktől eltérő tendenciát, mintázatot. A kérdőívben adott válaszaikkal összevetettük, hogy volt-e olyan tényező, ami jellemzően befolyásolta a kapott eredményeket.

3. Eredmények

3.1. Az alapfrekvencia változása

A két felvételi időpontban a két nemi csoportban mért *f*₀-értékek átlagát és szórását az [1](#) ábra mutatja. A férfiak átlagos alapfrekvenciája az első felvételen 115 ± 22 Hz, a második felvételen pedig 119 ± 31 Hz volt. A nők átlagos alapfrekvenciája az első felvételen 201 ± 38 Hz, a második felvételen 189 ± 40 Hz volt. Ez az átlagok között $-1,08$ félhang eltérés a nők, $+0,63$ félhang a férfiak beszédében, azaz a nők teljes felvételre mért átlagos *f*₀-jára csökkenés, míg a férfiak *f*₀-jára növekedés volt jellemző 10 év elteltével, a csoportszintű eltérés azonban alacsony. A Wilcoxon-próba alapján az átlagértékek a nők esetében

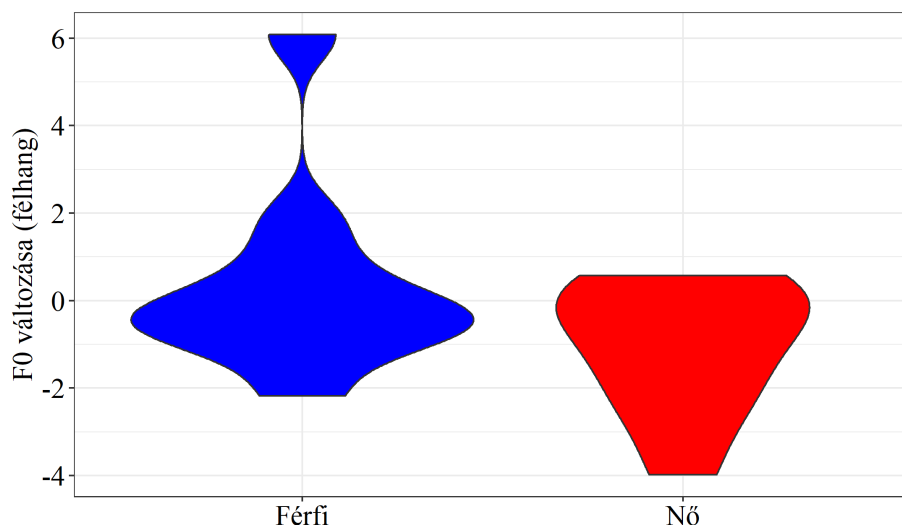
szignifikánsan eltértek a két felvétel között ($V = 3, p = 0,009$), míg a férfiak esetében nem.



1. ábra. Az alapfrekvencia-értékek (Hz) tartománya nemenként és felvételi időpontonként

A [2](#) ábrán két violinplot, azaz hegedűábra látható, melyek gyakorlatilag 90° -ban elforgatott sűrűségfüggvények, tehát azt mutatják meg, hogy az y -tengely értékei mennyire gyakoriak (amelyik y -értéknél szélesebb a forma, az az y -érték gyakoribb). Az ábra az f_0 beszélőnként számolt eltérését mutatja, tehát azt, hogy mennyivel változott a férfiak és nők f_0 -ja 10 év elteltével. Beszélőnként mindkét felvételre kiszámoltuk az átlagértéket, és azoknak a félhangban mért eltérése szerepel az ábrákon (a második felvétel f_0 -átlagértéke az első felvétel f_0 -átlagértékéhez viszonyítva). Az ábrán a negatív tartomány jelzi a csökkenő, a pozitív tartomány pedig a növekvő tendenciát 10 év elteltével. Az f_0 -különbségek mediánja mindkét nem esetében a nullához közeli, negatív értéket vesz fel, ez a nők esetében $-0,68$ félhang (az átlag $-1,06 \pm 1,50$ félhang), a férfiak esetében pedig $-0,28$ félhang (az átlag $+0,43 \pm 2,29$ félhang). Egy férfi esetében jelentősen nagyobb eltérést kaptunk a két felvételi időpont között, mint a többiek beszédében: F09 átlagos alapfrekvenciája 108 ± 21 Hz-ről 153 ± 43 Hz-

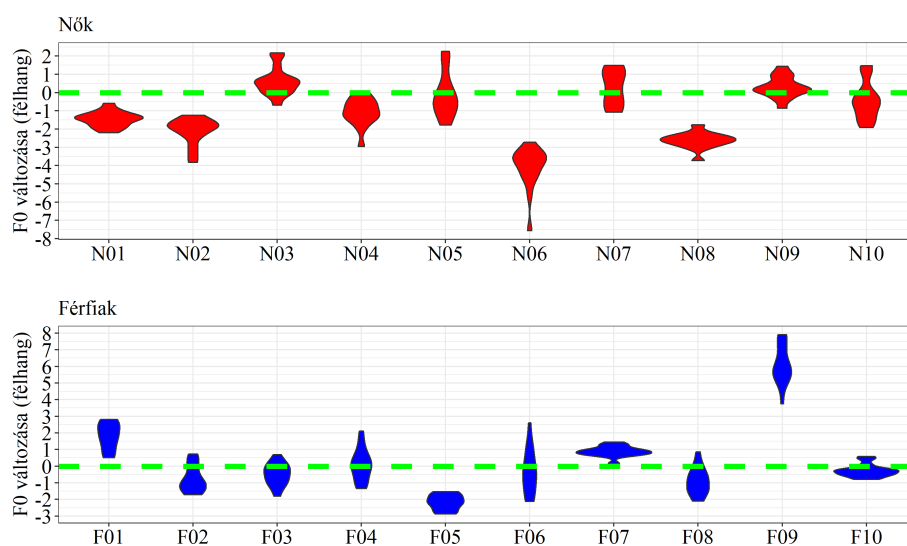
re emelkedett, azaz 6,09 félhanggal magasabb átlagos alapfrekvenciával olvasott a későbbi felvételen.



2. ábra. A beszélők átlagos f0-értékének eltérése 10 év elteltével a nők és férfiak esetében (félhang)

A 3. ábra azt mutatja meg, hogy hogyan alakult az átlagos f0 egyénenként. Míg a 2. ábrán az f0-változást a teljes felvételek átlagos alapfrekvencia-értékéből számítva ábrázoltuk, addig a 3. ábra nemcsak abban tér el ettől, hogy beszélőkre bontva látszanak az adatok, hanem abban is, hogy a mondatonkénti f0-eltérések láthatóak rajta. Az adatok minden beszélő esetén abból a 15 (illetve egy esetben 14) értékből állnak össze, ami a 15 mondat f0-átlagának különbsége a 10 év elteltével, tehát, ha a beszélőhöz tartozó összes érték a nulla y -értéknél behúzott zöld szaggatott vonal felett helyezkedik el, az azt jelenti, hogy minden mondatához tartozó érték pozitív, vagyis növekedés történt 10 év elteltével az f0-ban. Ha a szaggatott vonal alatt helyezkedik el, akkor minden mondatához tartozó érték negatív, vagyis csökkenés történt 10 év elteltével. Ha az ábrán az adott beszélőhöz tartozó értékek hegedűformáját metszi a szaggatott vonal, tehát egy része a negatív, másik része a pozitív tartományba esik, akkor az azt

jelentí, hogy a mondatok egy részénél növekedés, a másik részénél pedig csökkenés történt 10 év elteltével. Látható, hogy a változások mértéke és iránya egyénenként eltérő. Néhány esetben előfordul, hogy a változás iránya eltér a csoportra jellemző trendtől, ezek a különbségek azonban kisebb mértékűek, csoportszinten tehát eltérő tendencia figyelhető meg a két nem között. Egy-egy kiugró értéket találtunk a két nem esetében, F09-es esetében, akinek 107 ± 21 Hz-ről 153 ± 43 Hz-re emelkedett az alapfrekvencia-átlaga, és N06-ot, akinél az alapfrekvencia-csökkenés mértéke nagyobb mértékű volt a többi nőéhez képest (212 ± 25 Hz-ről 168 ± 39 Hz-re csökkent).



3. ábra. A beszélők f0-átlagának eltérése (félhang) a két felvételi időpont között (fent: nő, lent: férfi)

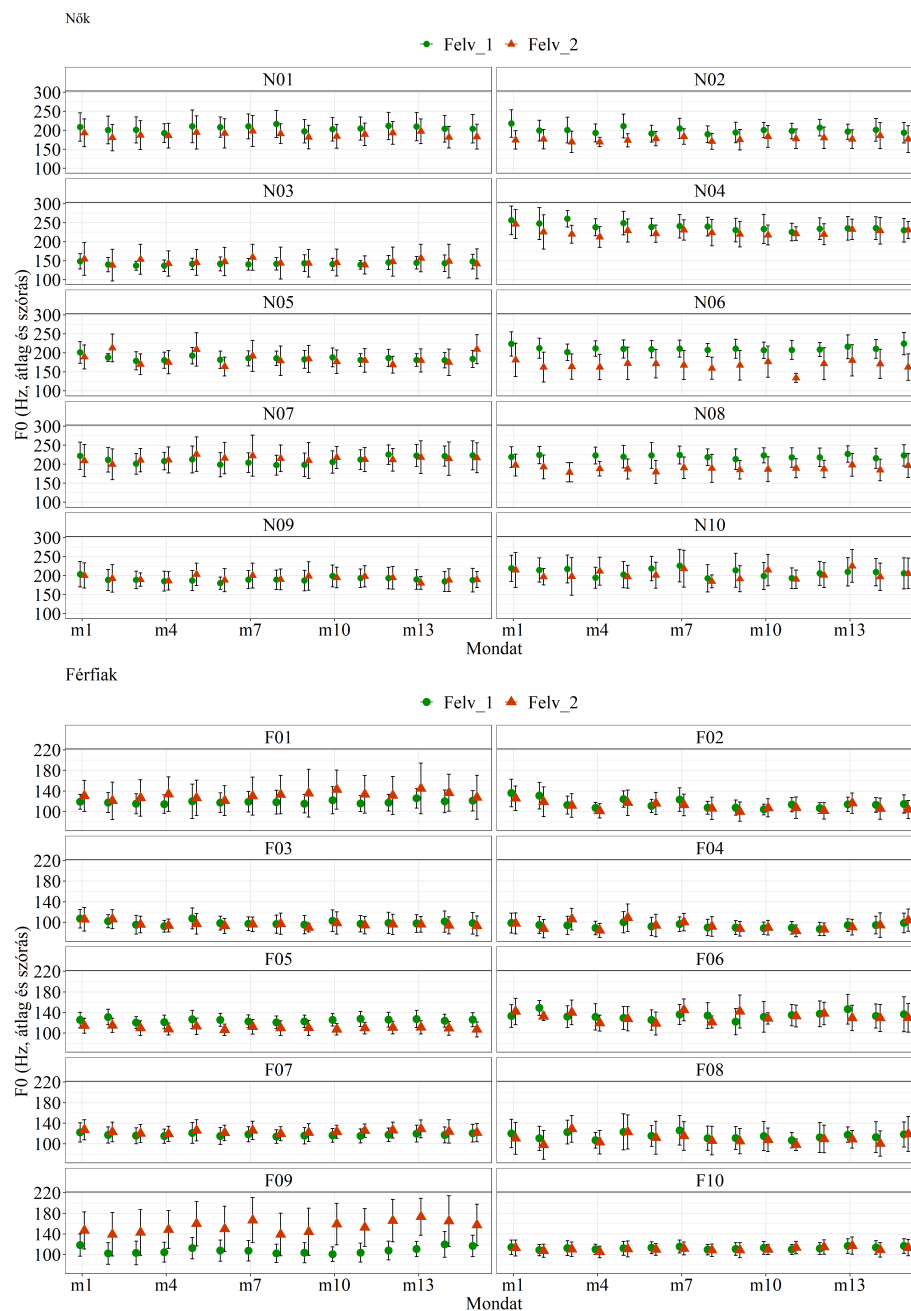
Nagyok az egyéni különbségek mind a nők, mind a férfiak tekintetében. A férfiaknál négy olyan eset van (F01, F05, F07, F09), ahol mind a 15 mondatnál ugyanabba az irányba változott az f0 10 év elteltével. Ez az irány három esetben (F01, F07, F09) pozitív, tehát nőtt az alapfrekvencia, és egy esetben (F05) negatív, vagyis csökkent az f0. A változás mértéke is különbözik az egyes beszélőknél: három beszélő esetén nem haladja meg a 3 félhangot (F01, F05,

F07), és egy esetben (F09) a változás sokkal jelentősebb: 4 és 8 félhang közötti, ez a változás pozitív irányú. Az ő esetében találtunk kiugró értéket a férfiakra jellemző teljes átlagban (vö. 2. ábra). Ez azt jelenti, hogy az ő esetében ez az emelkedés szisztematikus volt. A többi 6 férfi beszélőnél (F02, F03, F04, F06, F08, F010) vannak olyan mondatok, ahol negatív, és olyanok is, ahol pozitív irányba változott az f_0 a 10 év alatt. Közülük négy beszélőnél (F02, F03, F08, F010) az értékek jelentős része negatív tartományba esik, tehát a mondatok többségében csökkenés figyelhető meg az f_0 -ban 10 év elteltével. Ezek a csökkenések nem haladják meg a 2 félhangot, illetve egy beszélő esetén (F010) az egy félhangot sem. Két beszélő esetén (F04, F06) a mondatok kb. felében csökkent az f_0 , a másik felében pedig nőtt a 10 év alatt, vagyis nem jellemzi egységes tendencia az f_0 -változást.

A nőknél 5 olyan eset van (N01, N02, N04, N06, N08), ahol az összes mondatban csökkent az alaphérfkvencia 10 év elteltével. A csökkenés mértéke egyenként eltérő, három beszélőnél az értékek jelentősebb része két félhang alatt van (N01, N02, N04). Ennél valamivel nagyobb a változás a másik két beszélő esetén: az N08-as beszélőnél az értékek jelentős része 2 és 3 félhang közé esik, míg a legnagyobb mértékű változás az N06-os adatközlő esetén volt mérhető, az értékek jelentős része 3 és 5 félhang közé esik. A másik 5 női beszélő esetén voltak mondatok, ahol csökkenés, és voltak mondatok, ahol növekedés volt megfigyelhető az f_0 -ban 10 év elteltével. Két beszélőnél (N05, N010) az értékek jelentős része negatív tartományba esik, egy beszélőnél (N03) pedig pozitív tartományba, az előbbinél nagyobb részt csökkenés, az utóbbinál növekedés jellemző az f_0 -ra. Az N05-ös és az N010-es beszélőnél a csökkenések általában 2 félhangnál kisebbek, az N03 adatközlőnél pedig a növekedés általában 1 félhang alatti. Két beszélő van (N07, N09), akiknél a pozitív és negatív értékek nagyjából egyenlő arányban fordulnak elő, tehát nem mondható el egységes tendencia az f_0 -változásra.

A 4. ábra árnyaltabb képet ad arról, hogy az eltéréseket maga a mondat befolyásolja-e, valamint hogy a két felvételi időpont közötti változás mértéke (Hz-ben mérve) meghaladja-e az egyes felvételeken belüli variabilitást. Vannak

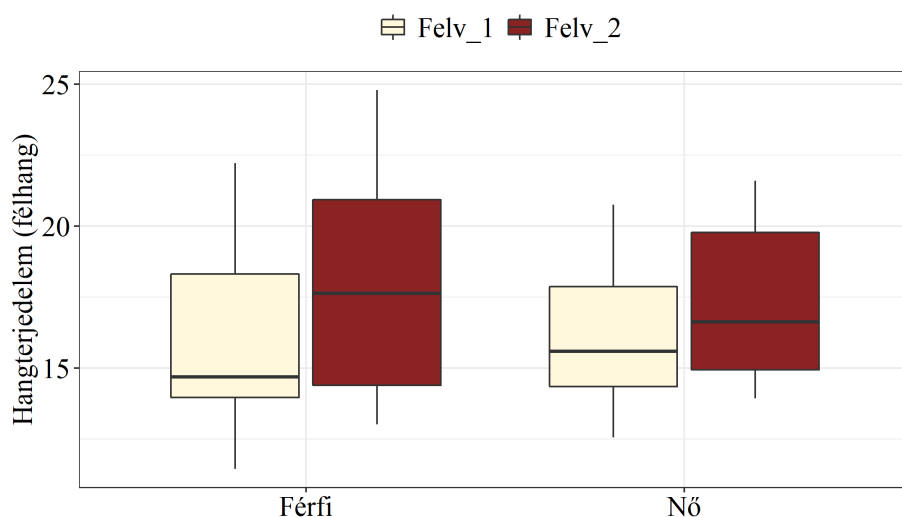
olyan adatközlők (F010, F03, N03, N07, N09), akiknél a beszélőn belüli variabilitás elhanyagolhatóan kicsi mind a felvételek, mind a mondatok szintjén. Vannak olyan adatközlők is (nők és férfiak vegyesen: N05, N010, F02, F04, F06, F08), akiknél a mondatok között már valamivel nagyobb a variabilitás egy felvételen belül, azonban a két felvétel között nem látszik szisztematikus eltérés. Ezek azok a beszélők, akiknél nem állapítható meg változás 10 év elteltével. Vannak olyan adatközlők (N06, N08, F09), akiknél a felvételek közti különbség jól elkülönül egymástól, annak ellenére, hogy a felvételen belüli variabilitás nagyobb, és mivel az eltérés szisztematikus, a mondatok többségében mérhető. Vannak olyan adatközlők is, akiknél a két felvételen mért f_0 -értékek látszólag alig térnek el egymástól, azonban ha a mondatok szerinti bontást nézzük, akkor láthatjuk, hogy szisztematikusan minden mondatnál megtalálható az azonos irányú eltérés, tehát az eltelt 10 év az ő alaphérfkvenciájukra is hatással volt, csak kis mértékben (N01, N02, N04, F01, F05, F07). A mondatonként mért átlagos alaphérfkvencia-értékeket kevert modellel vetettük össze. Ez alapján ezt a változót a nem és a felvételi időpont interakciója határozta meg szignifikáns módon ($F(1, 599) = 4,502, p = 0,048$). Ez az eredmény a fenti, teljes felvételekben mért átlagokra kapott eredményekkel együtt értelmezendő, és azt jelenti, hogy a nők esetében találtunk eltérést a két felvételi időpont között. A főhatások az adatok nagy részét magyarázzák ($R^2_{LMM(m)} = 0,792$), de a random hatások (a mondatok random konstansa, a felvétel random meredeksége beszélőnként) jelentősen erősítik a modell magyarázó erejét ($R^2_{LMM(c)} = 0,980$).



4. ábra. A vizsgálatban szereplő egyes beszélők alapfrekvenciája beszélőnként az első és második felvételen, mondatokra bontva

3.2. A hangterjedelem változása

A hangterjedelem, azaz a legmagasabb és a legalacsonyabb alapfrekvencia-érték aránya nem tért el a két nem között, és valamivel magasabb értékeket kaptunk 10 év elteltével (nők, első felvétel: $16,14 \pm 2,70$ félhang; nők, második felvétel: $17,30 \pm 2,85$ félhang; férfiak, első felvétel: $16,12 \pm 3,50$ félhang; férfiak, második felvétel: $17,90 \pm 4,22$ félhang) (5. ábra). A hangterjedelem esetében Scheirer–Ray–Hare-próbát végeztünk. Az ismételt mintás próba eredménye alapján nem találtunk szignifikáns eltérést sem a nemek, sem a felvételek időpontja alapján, sem a kettő interakciójában.

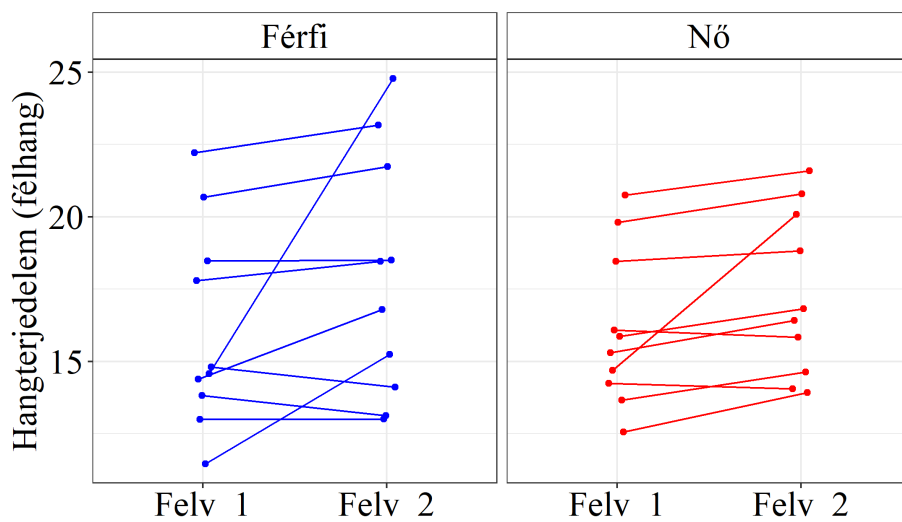


5. ábra. A hangterjedelem (félhang) a két felvételi időpontban nemek szerint

Ezt a tendenciát a 6. ábrán mutatjuk be beszélőnként, hogy a tágabb tartomány háttérét feltárjuk. Nagyon az egyéni különbségek abban a tekintetben, hogy 10 év elteltével mennyit változott a hangterjedelem: vannak beszélők, akiknél elenyésző ez a változás, de vannak pozitív, illetve negatív irányú jelentősebb változások is. A legkisebb különbség 10 év elteltével a férfiaknál $-0,02$ félhang (gyakorlatilag stagnálás), a nőknél $+0,64$ félhang. A legnagyobb változás 10 év elteltével a férfiaknál $+5,86$ félhang, a nőknél $+7,58$ félhang volt, ami szintén

arra utal, hogy a nőknél jelentősebbek a hangterjedelembeli változások, illetve hogy a jelentős változások pozitív irányúak.

A férfiak esetében négy beszélőnél (F02, F04, F07, F010) ez a változás elhanyagolható, nem haladja meg a 0,5 félhangot, tehát gyakorlatilag stagnálásról beszélhetünk, három beszélőnél (F03, F05, F06) a változás 1 félhang körül van, és három beszélő van (F01, F08, F09), akinél a változás jelentős, 5-6 félhangnyi. A kisebb változások (1 félhangnyi) fele pozitív, a másik fele negatív irányú volt, tehát nincs jellemző tendencia, viszont a jelentősebb változások (5-6 félhangnyi) minden esetben pozitív irányúak voltak, tehát ezen adatközlők esetén nőtt a hangterjedelem 10 év elteltével. Míg az alacsonyabb hangterjedelem-értékek esetében viszonylag logikus az emelkedés vagy stagnálás, addig a felsőbb tartományokban inkább csökkenés lenne várható. Ezzel szemben két nő és négy férfi kivételével emelkedés volt jellemző attól függetlenül, hogy az első felvétel idején mekkora volt a hangterjedelmük.



6. ábra. A férfiak és a nők hangterjedelmének változása 10 év elteltével

Négy beszélő esetén (N03, F03, F07, F09) a hangterjedelmi változás meghaladta a 2 félhangot, tehát náluk a változás jelentős volt 10 év elteltével, és ezek

a változások mind a tágabb hangterjedelem felé mutattak elmozdulást. A beszélők többségénél (6 nő és 5 férfi) a változás 0,5 és 2 félhang közé esett, két férfi esetében csökkent a hangterjedelem 10 év elteltével, a többi adatközlőnél növekedett. Öt beszélő esetén (N01, N04, N05, F04, F08) a változás nem haladta meg a 0,5 félhangot, tehát gyakorlatilag stagnálásról beszélhetünk.

A fenti adatok arra hívják fel a figyelmet, hogy a 10 év alatt csoportszinten az átlagos alapfrekvenciában a nők ejtésében találtunk eltérést, mégpedig csökkenést, a hangterjedelmet tekintve azonban nem volt szignifikáns változás. Ugyanakkor a beszélők közötti eltérések jelentősek voltak, mind az egyes felvételi időpontokon belül, mind pedig a két alkalom közötti eltérés alapján. Az  táblázat a hangterjedelem, az f0 minimum- és maximumértéke és az átlaga alapján egyszerre teszi összehasonlíthatóvá a beszélőket, valamint az egyes beszélőkön belül a két felvétel adatait. Kirajzolódik, hogy bizonyos beszélők esetében nagyobb mértékben tért el az átlagos f0 a két felvétel között, miközben a hangterjedelem szinte nem változott (pl. F01, F05, N01, N06, N08), mivel az alapfrekvencia minimum- és maximumértékei egy irányban és hasonló mértékben módosultak. Ugyanakkor volt olyan beszélő is, akinél a hangterjedelem változása jelentős volt a szélső értékek valamelyikének módosulása miatt, miközben az átlagos f0 szinte állandónak mondható (N03). Egyes beszélőknél (szinte) minden értékben nagyobb mértékű változás figyelhető meg (pl. F07, F09), de a legnagyobb számban olyanokat találunk, akiknek az esetében mindegyik paraméter lényegében állandó volt (pl. F02, F04, F06, F08, F010, N04, N05, N07). Mindez arra utal, hogy az alapfrekvenciából adódó különféle mérőszámok egymástól viszonylag függetlenül alakulnak, hiszen míg a (felvételen mérhető minimális és maximális f0-értékből számított) hangterjedelem nem érzékeny arra, hogy a rendelkezésre álló hangtartományból a beszélő melyik regisztert milyen gyakorisággal használja, az átlagértékeket, amelyet a legtöbb tanulmány vizsgál, ez a tényező erősen befolyásolja.

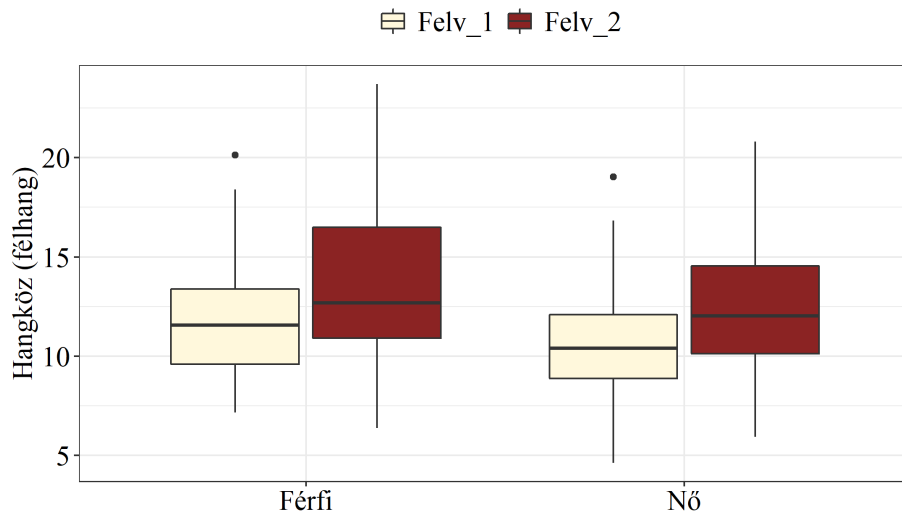
1. táblázat. Az egyes felvételek különféle alaphangfrekvencia-mutatói a beszélők függvényében

Beszélő	Első felvétel			Második felvétel		
	hangterjedelem	f0-min-f0-max	átlagos f0	hangterjedelem	f0-min-f0-max	átlagos f0
	(félhang)	(Hz)	(Hz)	(félhang)	(Hz)	(Hz)
F01	22,22	76–275	118	23,17	82–314	132
F02	17,80	77–215	116	18,46	75–219	110
F03	14,38	75–173	99	16,79	75–199	96
F04	13,00	75–160	93	13,00	75–160	94
F05	13,82	79–175	125	13,12	82–175	110
F06	20,68	79–259	134	21,74	76–267	132
F07	11,45	100–194	117	15,25	100–241	123
F08	18,47	75–219	114	18,50	75–219	108
F09	14,57	75–174	108	24,79	75–316	153
F010	14,80	82–193	112	14,11	86–194	111
N01	14,24	151–343	205	14,05	150–338	189
N02	15,29	128–310	200	16,41	125–324	176
N03	14,69	101–235	142	20,08	100–319	146
N04	16,08	150–380	238	15,83	152–379	224
N05	18,46	100–291	184	18,82	100–297	182
N06	19,81	103–323	212	20,80	100–333	168
N07	15,87	150–376	210	16,82	150–396	214
N08	13,66	136–299	221	14,63	129–300	189
N09	12,55	150–310	189	13,93	150–335	193
N010	20,75	104–345	207	21,59	100–350	202

3.3. A hangközértékek változása

Míg a hangterjedelem a teljes felvételen mérhető minimális és maximális alaphangfrekvencia-érték közötti tartományt számszerűsíti, addig a hangköz a kisebb beszédegységek f0-tartományát adja meg, a jelen esetben a vizsgált mondatokét (tehát itt az adott mondatban az f0max és az f0min közti félhangokban mért eltérést). A hangköz-eltéréseket a 7. ábra mutatja mind a nők, mind a felvételi időpontok között. Egységes tendenciának látszik, hogy a második felvételen mért hangközök tágabbak (nők, első felvétel: $10,39 \pm 2,37$ félhang; második felvétel: $12,53 \pm 3,20$ félhang; férfiak, első felvétel: $11,74 \pm 2,67$ félhang; második felvétel: $13,78 \pm 4,11$ félhang). Ebben a tekintetben a két felvételi időpont

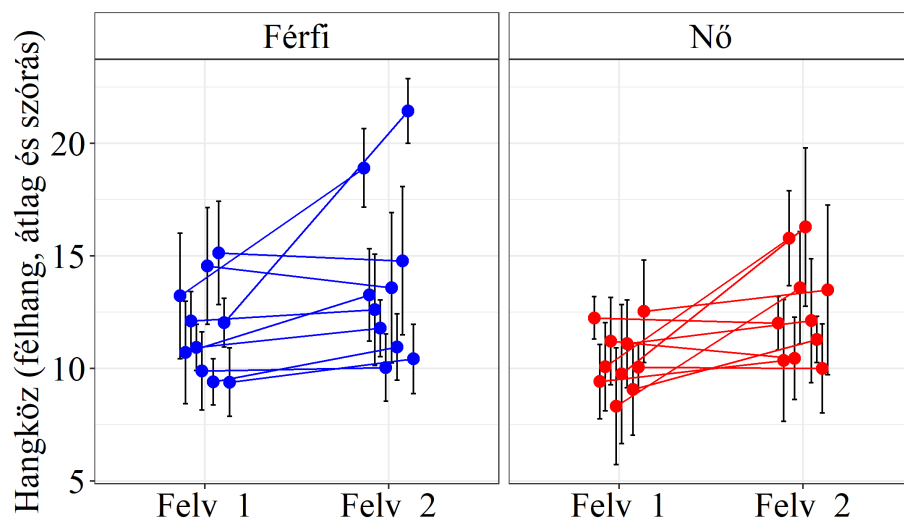
között szignifikáns eltérést kaptunk ($F(1, 599) = 102,372, p < 0,001$). A hangközök nemek közötti eltérését illetően azt látjuk, hogy a nők hangközértékei valamivel alacsonyabbak, de ez az eltérés nem volt szignifikáns. Ugyanakkor a marginális R^2 (0,096) és a kondicionális R^2 (0,440) összevetése azt mutatja, hogy bár fellelhető szignifikáns eltérés a két felvételi időpont között, a random hatások (a mondatonkénti random konstans és a felvételenkénti random meredekség a beszélőkre) jelentősen növelik a modell magyarázóerejét, de önmagában a fix hatások nem elégséges magyarázó erejűek, további, erős hatás befolyásolja az eredményeket.



7. ábra. A hangközértékek alakulása a nemek szerint a két felolvasás során

A beszélőket tekintve az rajzolódik ki, hogy vannak köztük olyanok, akiknél az eltérések a nullához közeliak, tehát nem történt jelentős hangközbeli változás a 10 év alatt, és vannak olyan beszélők is, akiknél az eltérések jelentősebben eltérnek a nullától (8. ábra). Hat férfi és hét nő esetében nőtt a mondatok átlagos hangközértéke, egy férfinál és egy nőnél csökkent. Összességében az átlagos hangközértékek tehát nőttek, de csökkenés is előfordult. Azokban az esetekben, ahol legalább 4 félhangnyi volt a különbség a két felvétel között (2 férfi és 3

nő), minden esetben növekedést mutatott a hangköz 10 év elteltével. Azokban az esetekben viszont, ahol a két felvétel közti eltérés 4 félhang alatti volt, az eltérés iránya már nem egységes: növekvő (3 férfi, 3 nő) és csökkenő (1 férfi és 1 nő) tendencia is található 10 év elteltével. Kevés beszélőnél (négy férfi, két nő) mondható stagnálás a két felvételi időpont között, az ő esetükben a két felvétel közti átlagos hangközkülönbség nem haladja meg a 0,5 félhangot. Összességében elmondható, hogy a beszélők egy részénél növekedés, egy másik részénél pedig stagnálás figyelhető meg a hangközértékekben 10 év elteltével. Ez a növekedés vagy stagnálás azonban nem függ a nemtől, de általában növekedésként jellemezhető a random hatások erőteljes szerepével, mint fentebb ismertettük.



8. ábra. A hangközértékek eltérései 10 év elteltével beszélőnként

3.4. Az egyéni variabilitást magyarázó tényezők

Az eredmények mind az f0 átlaga, mind a hangterjedelem és a hangköz esetében nagy beszélők közötti variabilitást mutattak. Érdekes elemezni, hogy ezek háttérében milyen tényezők állhatnak. Az életkor mellett számos tényező befolyásolhatja a zöngképzést, így az alaphangfrekvencia értéktartományát, a zöngé

periodicitását is, amelyek a jelen vizsgálatban is hatással lehetnek az eredményekre. A férfiak esetében egyértelműen a pubertáskori alaphfrekvencia-változás a legjelentősebb, amely nem kizárólag a szervek növekedésének, hanem a hormonháztartás átalakulásnak és a gégen belüli belső struktúra változásának is a következményei (összefoglalóan [Gugatschka et al., 2010](#)). A férfiak idős korban bekövetkező alaphfrekvencia-változása ismét részben a hormonháztartás átalakulásából fakad (összefoglalóan [Gugatschka et al., 2010](#); [Kadokia et al., 2013](#)), de emellett mindkét nem esetében meg kell említeni a szövetek, porcok változásait is. A nők esetében ugyancsak bekövetkezik a pubertásban a zöngképzésben változás, de az a férfiakéhoz képest kisebb mértékű. Az ő esetükben a serdülőkortól időskorig a menstruációs ciklus során folyamatosan változó hormonszintek ciklikusan határozzák meg a zöngképzést. A menstruáció előtt a premenstruációs dysphonia, ami az ehhez az időszakhoz köthető hormonhatások miatti nagyobb valószínűséggel előforduló reflux és ödéma következménye, valamint az ezen időszak hirtelen hormonváltozásai a gége neuromotoros kontrollját is befolyásolják (összefoglalóan: [Amir & Kishon-Rabin, 2004](#); [Kadokia et al., 2013](#)). Ugyancsak meghatározó a felnőttkorban a nők esetében a terhesség, szülés. Az átlagos alaphfrekvencia értéke a terhesség egyik trimeszterében sem mutatott eltérést a kontrollszemélyekétől Saltürk és kollégái ([2015](#)) kutatásában (a maximális fonációs idő a 3. trimeszterben rövidült). Pisanski és kollégái ([2017](#)) egy nyomonkövetéses vizsgálatban azt találták, hogy a szülés utáni egy évben alacsonyabb és monotonabb a nők alaphfrekvenciája, mint azelőtt, illetve a terhesség alatt, de az egy év elteltével a korábbi értékekhez tér vissza a beszédük. A menopauza az alaphfrekvencia mélyülését okozza a bekövetkező hormonális változások következtében (összefoglalóan: [Amir & Kishon-Rabin, 2004](#)).

A pajzsmirigy-hormonok esetében az alulműködés okozhat zöngéváltozást, mégpedig a szövetekben felgyűlt szacharidok következtében a premenstruációhoz hasonló hatást vált ki, alaphfrekvencia-mélyülés és a tartomány beszűkülése is jellemző (összefoglalóan: [Kadokia et al., 2013](#)).

Mindezek miatt szerepelt a vizsgálatban egy kérdőív, hogy a 10 év elteltével milyen olyan események merültek fel a beszélőink esetében, amelyek az

esetleges kiugró értékeiket indokolhatják. Ezek mellett persze további tényezők is szerepet játszhatnak (pl. fáradtság, aktuálisan az átlagtól eltérő hangterhelés – akár vegyszeres érintkezés), amelyeket a jelen vizsgálatban sem tudunk figyelembe venni.

A beszélők által kitöltött kérdőív alapján azt tudjuk, hogy három nő: N01, N08 és N09 esetében folyamatos pajzsmiriggyel kapcsolatos betegség és/vagy hormon tartalmazó terápia áll fenn. Négy nőnek, N02, N06, N08, N09, született gyermeke a két felvételi idő között, de csak N08 gyermeke egy év körüli. Egy nő (N04) pedig idősebb a többi adatközlőnél, 45 éves volt az első felvétel alkalmával, ami esetlegesen a menopauza felmerülését indokolhatja. Két férfi leszokott ugyan, de korábban dohányoztak (F06, F09), egy nő néha (N08), kettő pedig rendszeresen dohányzik több mint 10 éve (N03, N06). A napi beszédidőre is rákérdeztünk. Azt kellett megbecsülni a beszélőknek, hogy átlag hány órát beszélnek egy nap. E tekintetben négy beszélő adott magas értékeket (több mint 10 órát). Mindezeket a [2](#) táblázatban foglaltuk össze.

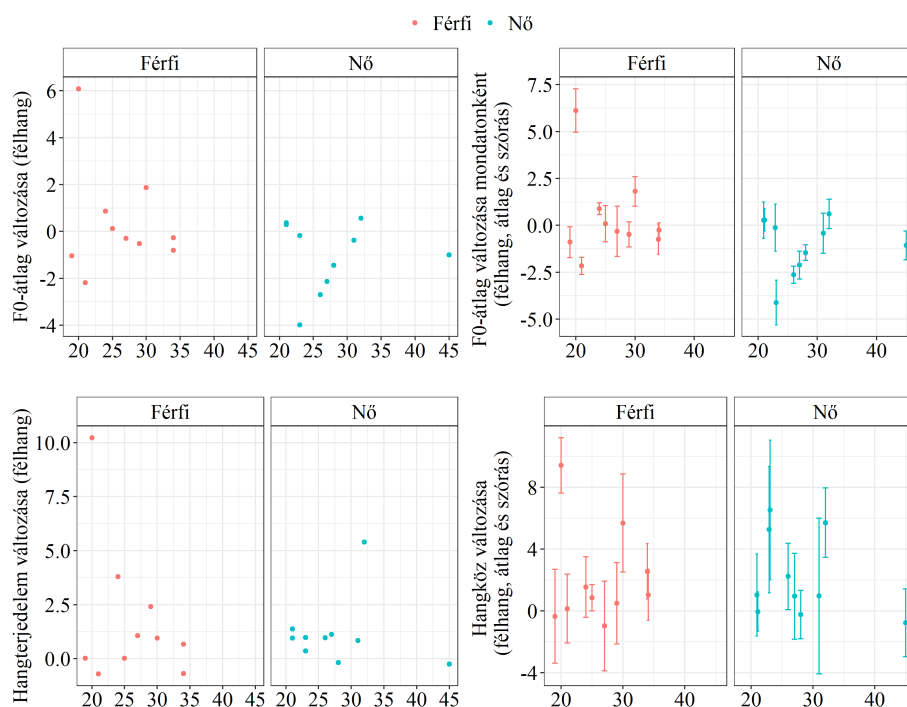
Mivel egy longitudinális vizsgálat során az adatközlők utolérése nehézkes, így nem zártuk ki ezeket a beszélőket sem.

Az életkor és az elemzett változások között nem találtunk olyan összefüggést, amely alapján jelentős eltérés lenne a beszélők között. A [9](#) ábrán a beszélők életkora szerint látjuk a teljes felvételen mért átlagos alapfrekvencia változását (bal felül), a mondatonként számított átlagalapfrekvencia-változást (jobb felül), a hangterjedelem (bal alul) és a hangközátlag változását (jobb alul). A férfiak esetében jellemzően a korábban is említett F09-es beszélő ejtésében látjuk a legnagyobb változást, amely a többi férfi adataitól jelentősen eltér minden mért változóban. Az első felvételkor mért alapfrekvencia-értékei hasonlóak a többi férfiéhoz, míg a második felvételen mért adatok azoktól jóval magasabbak. A férfiak esetében a 30 éves férfi, F01, adatait érdemes még megfigyelni, az ő esetében ugyan kisebb a többiektől számított eltérés, de az életkort is figyelembe vevő tendenciától eltér. F09 korábban dohányzott, de már leszokott, más, a beszédét befolyásoló tényezőről nem számolt be. A dohányzás ugyanakkor az alapfrekvencia csökkenését okozza a kutatások szerint (pl. [Sorensen & Horii](#)

2. táblázat. A beszélők alaphfrekvencia-értékeit esetlegesen befolyásoló adatok (pajzsm. = pajzsmirigy-betegség)

	nem	kor	dohányzás	terhesség	pajzsm.	hangterhelés (óra/nap)
F01	Férfi	30	nem		nem	6
F02	Férfi	34	nem		nem	12
F03	Férfi	29	nem		nem	6
F04	Férfi	25	nem		nem	1
F05	Férfi	21	nem		nem	2,5
F06	Férfi	27	leszokott		nem	2
F07	Férfi	24	nem		nem	2,5
F08	Férfi	19	nem		nem	5
F09	Férfi	20	leszokott		nem	3
F010	Férfi	34	nem		nem	1,5
N01	Nő	28	nem	nem	igen	7
N02	Nő	27	nem	igen	nem	12
N03	Nő	32	igen	nem	nem	17
N04	Nő	45	nem	nem	nem	7
N05	Nő	23	nem	nem	nem	5
N06	Nő	23	igen	igen	nem	10
N07	Nő	21	nem	nem	igen	8
N08	Nő	26	néha	nem	nem	4
N09	Nő	21	nem	igen	igen	8
N010	Nő	31	nem	igen	nem	4,5

1982; Ayoub et al., 2019). F01 esetében ilyen hatás sem áll fenn. Egyik beszélő esetében sem volt jellemző, hogy bármely mért változó (f0, hangterjedelem, hangköz) alacsonyabb/szűkebb lett volna a többi férfira jellemző eredménynél korábban, a második mérésakor mért értékek azonban F09 esetében magasab-
bak/tágabbak lettek azoknál.



9. ábra. A vizsgálatban elemzett hangjellemző-változások a beszélő életkora függvényében:
átlagos alaphang (bal felül), mondatonkénti átlagos alaphang (jobb felül),
hangterjedelem (bal alul) és hangköz (jobb alul)

A nők beszédében az átlagos alaphang változásában a 25–35 éves kor között egyre kisebb változás látszik kirajzolódni a 9. ábrán. Ez azonban egy-egy beszélő adatait jelenti. Összességében a huszonéves nők esetében jelentősebb alaphang-csökkenés látszik megjelenni, mint a 30–40 évesek esetében. A 45 éves nő eredményei – bár a menopauza esetleges hatása megjelenhetne – nem térnek el a 30-asoktól ebben a tényezőben. N06 f0-változása nagyobb mértékű a

többiekénél. A hangterjedelem esetében az életkor mentén egy olyan trendvonal sejthető, amely alapján az életkor növekedésével a kis mértékű növekedéstől az enyhe mértékű csökkenés irányába jelentősen eltolódást. Az illeszthető trendvonal meredeksége azonban alacsony, a beszélők száma pedig kevés. Mindenképp kiugró értéket mutatott ugyanakkor a 32 éves nő, N03. A nők között a hangköz változásában a legidősebb nő esetében valamelyest alacsonyabb értéket kaptunk a többi női beszélőhöz képest. Az eltérés nem jelentős, mégis enyhe csökkenés, míg a többiek esetében stagnálást vagy növekedést találtunk. A két 23 éves, N05, N06, és a 32 éves N03 esetében nagyobb mértékű emelkedést mértünk. N03, N05 és N06 esetében találtunk tehát a többiekétől eltérő tendenciákat a nők beszédében. N03 dohányzik, illetve napi 17 órára becsülte a hangterhelését (a legmagasabb értékre minden vizsgált személy közül). N06 esetében született egy gyermek az elmúlt 10 évben, illetve ő is dohányzik. N05 nem jelzett olyan tényezőt, amely az eltérést okozhatta. A többi olyan női beszélőnél, akinek gyermeke született, jelentős a hangterhelése, vagy pajzsmirigy-gyulladása van, nem találtunk a csoportra jellemző tendenciánál nagyobb mértékű változást.

4. Következtetések

Fiatal és középkorú nők és férfiak mondatolvasásában elemeztük az alaphangfrekvencia jellemzőinek eltérését 10 év elteltével. A teljes felvételen és a mondatonként mért átlagos alaphangfrekvenciát, a hangterjedelmet és a hangközvetített össze.

Feltételeztük, hogy a változás mértéke a fiatal és középkorú felnőttek esetében nem lesz magas, inkább az egyéni változatosság lesz jellemzőbb (i). Ez a feltételezés részben nyert alátámasztást, ugyanis a változás mértéke alacsony, de a nők és a férfiak esetében is inkább egy-két adatközlő kivételével jól megragadható csoporttendenciákat találtunk.

A korábbi szakirodalom alapján feltételeztük, hogy (ii) a nők esetében az alaphangfrekvencia csökken (ahogy mind longitudinális (Mwangi et al., 2009), mind keresztmetszeti (Stathopoulos et al., 2011; Eichhorn et al., 2018) vizsgálatok

csökkenő tendenciát találtak, 20 és 60 év között), (iii) a férfiak esetében pedig inkább az egyéni eltérések dominálnak (20 és 50 év között Schötz (2006) enyhe csökkenést, Stathopoulos és munkatársai (2011) enyhe növekedést, míg Cox és Selent (2015) enyhe növekedést, majd csökkenést talált). Ez a két feltételezésünk alátámasztást nyert. A nők esetében a teljes felvételen és a mondatonként mért átlagos alaphangfrekvenciában is csökkenést mértünk. A férfiak felolvasásában ugyanakkor vagy stagnálás vagy emelkedés volt mérhető, de a változás nem volt szignifikáns az ő esetükben.

A hangterjedelem és a hangköz esetében nem vártunk jelentős eltérést (iv), mivel a fiatal-középkorú felnőttek esetében kevésbé jelentős öregedési vagy egyéb hatást vártunk a beszédben. A hangterjedelem általában emelkedést vagy stagnálást mutatott mindkét nem esetében, de nem találtunk szignifikáns eltérést a két felvételi időpontban mért adatok között, az előzetes hipotézissel egyezően. A hangközértékek változása szignifikánsnak bizonyult, hiszen a legtöbb mondat esetében nagyobb hangközértéket kaptunk a második felvételi időpontban. Az egyes beszélőkre jellemző eltérés többnyire ugyancsak emelkedés volt. Ennek hátterében azt feltételezzük, hogy a második felvétel idején részben az életkori változásból (nagyobb tapasztalat) adódóan, részben a már ismerős helyzet és feladat miatt a beszélők komfortosabban érezték magukat a stúdiókörülmények között, rutinosabbak voltak, és ezért a felolvasáskor jobban törekedtek a kifejezőbb megformálásra, tágabb hangközű dallamíveket valósítva meg. Ez ugyanakkor nem mutatott összefüggést az alaphangfrekvencia átlagos értékének változásával, hiszen a nőknél átlagosan csökkenés volt tapasztalható ebben az értékben.

Mivel nemenként tíz fő beszédét elemeztük, a beszélők közötti variabilitás nem elhanyagolható annak ellenére sem, hogy jellemzően kimutatható csoporttendenenciákat találtunk. A fiatal és középkorú nők esetében a korábbi keresztmetszeti tanulmányokban talált tendencia, azaz az átlagos alaphangfrekvencia csökkenése az életkor előrehaladtával ebben a longitudinális vizsgálatban is megmutatkozott. A férfiak beszédében a korábbi, keresztmetszeti tanulmányok ebben az életkori szakaszban változatos képet adtak, a jelen longitudinális vizsgálatban

csak 1-1 férfi esetében találtunk változást, csoportszinten nem. Fontos kiemelni, hogy míg a hangköz esetében nagy variabilitást is tapasztaltunk egy-egy beszélő mondatai között, az átlagos alapfrekvencia esetében jellemző volt, hogy az adott beszélőre (ha jellemző volt változás), az a mondatainak nagy részére vagy mindre azonos irányú eltérésben jelent meg a felvételi időpontok között.

A vizsgált beszélők esetében elemeztük, hogy az életkorral, a dohányzással, terhességgel, hangterheléssel, pajzsmirigybetegeccsel hogyan függ össze a vizsgált jellemzőkben az adott nemre jellemző értékektől való eltérés. Azt találtuk, hogy a huszonéves és harmincas életkorok között a nők esetében nagyobb változás volt mérhető az alapfrekvencia átlagában, de a többi tényező nem befolyásolta a csoportok eredményeit. Ugyanakkor ezen tényezők pontos hatását célzott vizsgálatokkal lehet elemezni.

A vizsgálat és a longitudinális elemzésre alkalmas magyar nyelvű korpusz adatai általánosságban lehetővé teszik egy olyan referencia-adatbázis kialakítását, amelyet különféle alkalmazásokban, például az igazságügyi hangszakértői munkában alkalmazni lehet. Az itt szereplő életkori intervallum a bűnügyi szempontból releváns beszélői életkori sávot fogja át (vö. [Skarnitzl & Vaňková 2017](#)), továbbá az azonos mondatokról készült felolvasások egyénenkénti különbségei tanulsággal szolgálnak a hangszakértői munka számára is (vö. pl. [Tatár et al. 2021](#)). Mindemellett a jelen vizsgálat tárgyát is képező alapfrekvencia az igazságügyi hang-összehasonlítás egyik leggyakrabban alkalmazott paramétere, mivel a szöveg tartalmától függetlenül vizsgálható (nem kívánja meg azonos szegmentumok azonos hangkörnyezetben való előfordulását, ellentétben például a magánhangzóformánsok elemzésével), másrésztől ugyanakkor ez az egyik legváltozékonyabb sajátossága az emberi beszédnek ([Skarnitzl & Vaňková 2017](#)). Mindezek a tényezők teszik szükségessé a hasonló vizsgálatokat, amelyekben az f_0 -t befolyásoló különféle tényezők (például az életkornak a korábbi vizsgálatokhoz képest rövidebb távú) hatása megmutatkozik.

Köszönetnyilvánítás

A kutatást az NKFIH FK128814-es számú pályázat támogatta.

Hivatkozások

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.
- Amir, O., & Kishon-Rabin, L. (2004). Association Between Birth Control Pills and Voice Quality. *The Laryngoscope*, 114, 1021–1026.
- Ayoub, M. R., Larrouy-Maestri, P., & Morsomme, D. (2019). The Effect of Smoking on the Fundamental Frequency of the Speaking Voice. *Journal of Voice*, 33, 802 11–802 16. URL: <https://doi.org/10.1016/j.jvoice.2018.04.001>.
- Bartoń, K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. URL: <https://CRAN.R-project.org/package=MumIn>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48.
- Belsky, D. W., Caspi, A., Houts, R., Cohen, H. J., Corcoran, D. L., Danese, A., Harrington, H., Israel, S., Levine, M. E., Schaefer, J. D., Sugden, K., Williams, B., Yashin, A. I., Poulton, R., & Moffitt, T. E. (2015). Quantification of biological aging in young adults. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 4104–4110.
- Berger, T., Peschel, T., Vogel, M., Pietzner, D., Poulain, T., Jurkutat, A., Meuret, S., Engel, C., Kiess, W., & Fuchs, M. (2019). Speaking Voice in Children and Adolescents: Normative Data and Associations with BMI, Tanner Stage, and Singing Activity. *Journal of Voice*, 33, 580.e21–580.e30.
- Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer. (Version 6.1.30) [Computer Program]. URL: <http://www.praat.org>.

- Brown, W. S., Morris, R. J., Hollien, H., & Howell, E. (1991). Speaking fundamental frequency characteristics as a function of age and professional singing. *Journal of Voice*, 5, 310–315.
- Busse, E. W. (2002). General Theories of Aging. In J. R. M. Copeland, M. T. Abou-Saleh, & D. G. Blazer (Eds.), *Principles and Practice of Geriatric Psychiatry, Second Edition* (p. 23–28). Chichester: J. Wiley and Sons.
- Cox, V. O., & Selent, M. (2015). Acoustic and respiratory measures as a function of age in the male voice. *Journal of Phonetics and Audiology*, 1, 105. doi:[10.4172/jpay.1000105](https://doi.org/10.4172/jpay.1000105).
- Debruyne, F., & Decoster, W. (1999). Acoustic differences between sustained vowels perceived as young or old. *Logopedics Phoniatrics Vocology*, 24, 1–5.
- Decoster, W., & Debruyne, F. (2000). Longitudinal voice changes: facts and interpretation. *Journal of Voice*, 14, 184–193.
- Eichhorn, J. T., Kent, R. D., Austin, D., & Vorperian, H. K. (2018). Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice*, 32, 644 1–644 9.
- Grácz, T. E., Huszár, A., Krepsz, V., Száraz, B., Damásdi, N., & Markó, A. (2020). Longitudinális korpusz magyar felnőtt adatközlőkről. In G. Berend, G. Gosztolya, & V. Vincze (Eds.), *XVI. Magyar Számítógépes Nyelvészeti Konferencia kötete* (p. 103–114). Szeged: JATEPress.
- Gugatschka, M., Kiesler, K., Obermayer-Pietsch, B., Schoekler, B., Schmid, C., Groselj-Strele, A., & Friedrich, G. (2010). Sex Hormones and the Elderly Male Voice. *Journal of Voice*, 24, 369–373. doi:[10.1016/j.jvoice.2008.07.004](https://doi.org/10.1016/j.jvoice.2008.07.004).
- Harrington, J., Palethorpe, S., & Watson, C. I. (2007). Age-related changes in fundamental frequency and formants: A longitudinal study of four speakers. In *Proceedings of Interspeech* (p. 2753–2756).

- Hollien, H., & Ship, T. (1972). Speaking fundamental frequency and chronological age in males. *Journal of Speech and Hearing Research*, 15, 155–159.
- Kadakia, S., Carlson, D., & Sataloff, R. T. (2013). The effect of hormones on the voice. *Journal of Singing*, 69, 571–574.
- Kaur, J., & Narang, V. (2015). Variation of pitch and formants in different age group. *International Journal of Multidisciplinary Research and Modern Education*, 1, 517–521.
- Kuznetsova, A., Brockhoff, B. P., & Christensen, B. R. H. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82, 1–26. URL: <https://doi.org/10.18637/jss.v082.i13>.
- Mangiafico, S. (2021). RCompanion: Functions to Support Extension Education Program, Evaluation. URL: <https://CRAN.R-project.org/package=rcompanion>.
- Mwangi, S., Spiegl, W., Hönig, F., Haderlein, T., Maier, A., & Nöth, E. (2009). Effects of Vocal Aging on Fundamental Frequency and Formants. In *Proceedings of the International Conference on Acoustics NAG/DAGA* (p. 1761–1764). Rotterdam, Netherlands.
- Neuberger, T., Gyarmathy, D., Grácz, T. E., Horváth, V., Gósy, M., & Beke, A. (2014). Development of a large spontaneous speech database of agglutinative hungarian language. In *Proceedings of TSD, 2014* (p. 424–431). New-York – Berlin: Heidelberg–Springer.
- Nishio, M., & Niimi, S. (2008). Changes in speaking fundamental frequency characteristics with aging. *Folia Phoniatrica et Logopaedica*, 60, 120–127. doi:[10.1159/000118510](https://doi.org/10.1159/000118510).
- Pedersen, M., Agersted, A., & Jønsson, A. (2015). Aspects of Adolescence and Voice: Girls versus Boys – A Review. *Journal of Child & Adolescent Behavior*, 3, 211. doi:[10.4172/2375-4494.1000211](https://doi.org/10.4172/2375-4494.1000211).

- de Pinto, O., & Hollien, H. (1982). Speaking fundamental frequency characteristics of Australian women: Then and now. *Journal of Phonetics*, 10, 367–375.
- Pisanski, K., Bhardwaj, K., & Reby, D. (2017). Women’s voice pitch lowers after pregnancy. *Evolution and Human Behavior*, 39, 457–463. URL: <https://doi.org/10.1016/j.evolhumbehav.2018.04.002>
- Quené, H. (2014). hqmisc: Miscellaneous convenience functions and dataset. R package version 0.1-1. . URL: <http://CRAN.R-project.org/package=hqmisc>
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: <https://www.R-project.org/>
- Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on f0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication*, 52, 638–651.
- Russell, A., Penny, L., & Pemberton, C. (1995). Speaking fundamental frequency changes over time in women: A longitudinal study. *Journal of Speech, Language, and Hearing Research*, 38, 101–109.
- Saltürk, Z., Kumral, T. L., Bekiten, G., Atar, Y., Atac, E., Aydoğdu, I., Yıldırım, G., Kılıç, A., & Uyar, Y. (2015). Objective and Subjective Aspects of Voice in Pregnancy. *Journal of Voice*, 30, 70–73. URL: <http://dx.doi.org/10.1016/j.jvoice.2015.02.013>
- Schötz, S. (2006). *Perception, analysis and synthesis of speaker age*. Ph.d. thesis. University of Lund Sweden.
- Sebastian, S., Babu, S., Oommen, N. E., & Ballraj, A. (2012). Acoustic measurements of geriatric voice. *The Journal of Laryngology & Otology*, 2, 81–84.
- Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). afex: Analysis of Factorial Experiments. URL: <https://CRAN.R-project.org/package=afex>.

- Skarnitzl, R., & Vaňková, J. (2017). Fundamental frequency statistics for male speakers of common Czech. *Acta Universitatis Carolinae: Philologica*, 2017, 7–17. URL: https://karolinum.cz/data/clanek/3901/Phil_3_2017_01_Skarnitzl_Vankova.pdf.
- Sorensen, D., & Horii, Y. (1982). Cigarette smoking and voice fundamental frequency. *Journal of Communication Disorders*, 15, 135–144.
- Stathopoulos, E. T., Huber, J., & Sussman, J. E. (2011). Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *Journal of Speech, Language, and Hearing Research*, 54, 1011–1021.
- Tatár, Z., Varga, Z., & Főző, E. (2021). Beszélőazonosítás a kriminalisztikában. In A. Markó (Ed.), *Tanulmányok a beszédtudomány alkalmazásainak köréből* (pp. 229–250). Budapest: ELTE Eötvös Kiadó. URL: <https://www.eltereader.hu/media/2021/10/Marko-Alexandra-Tanulmanyok-a-beszedtudomany-v3-e-valtozat-jo.pdf>.
- Torre, P., & Barlow, J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders*, 42, 324–333.
- Traunmüller, H., & Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. Manuscript.

Hangsúlyos szótagok vizsgálata a francia nyelvben a Melodic Analysis of Speech (MAS) módszer használatával

Párizsi és québeci nyelvváltozatok összehasonlítása

Szabados Laura¹

¹*Eötvös Loránd Tudományegyetem*

Abstract

The aim of this research is to present the main differences in the intonation of stressed syllables in declarative utterances of spontaneous speech in Parisian and Quebec language variants with the Melodic Analysis of Speech method (Cantero Serena & Font-Rotchés, 2020). We cannot speak about dialectal contrast regarding our results as only six speakers' (three from Paris, three from Quebec) utterances have been analyzed. However, these results tend to show remarkable similarities with previous descriptive and contrastive studies about the intonation of the two dialects. Generally, the main intonation contour is the same; stressed syllables are always rising – except at the end of the sentence where they are falling – so it is a peaking pattern, and the first peak is consistently a stressed syllable in both variants. The main differences between Parisian and Quebec variants are as follows: in Quebec intonation, the F0 value is generally less high, stressed syllables are not as much characterized by sudden rises as in Parisian intonation, and syllable boundaries are prosodically less clear than the Parisian ones.

Keywords: intonation, stressed syllable, spontaneous speech, French, Paris, Quebec

1. Bevezetés

A jelen munkánk célja (Cantero Serena & Font-Rotchés, 2020) Melodic Analysis of Speech (MAS) módszerével vizsgált és elemzett párizsi és québeci nyelvváltozatok közti intonációs különbség bemutatása kijelentő módú megnyilatkozásokban, elsősorban hangsúlyos szótagok szempontjából. Az említett módszer kiváltképp alkalmas nyelvek közti vagy dialektális különbségek bemutatására azok intonációjának vizsgálatával. Két fő fázisból áll: (1.) akusztikus fázis,

Email address: szabadosl@student.elte.hu (Szabados Laura)

melynek során a Praat (Boersma & Weenink 2019) alkalmazás segítségével a mondat minden szótagjának lemérjük hertzértékét, majd standardizáljuk azt, (2.) percepció fázis, melynek során manipuláljuk a hangot a standardizált értékek alapján, és azt anyanyelvi beszélőknek lejátszunk ellenőrzés céljából (Cantero Serena & Font-Rotchés 2020). Jelen előtanulmányunkban csak az első fázis módszereit alkalmaztuk a párizsi és québeci csoport intonációs különbségeinek vizsgálatához. Nem feltétlenül beszélhetünk dialektális eltérésekről, hiszen összesen három-három anyanyelvi beszélő intonációs mintázatát vizsgáltuk, amely egy összetett, dialektális eltérés tanulmányozásához nem elegendő, így alkalmasabb párizsi és québeci nyelvváltozatokról beszélni. Ezen nyelvváltozatok kontrasztív elemzését mutatjuk be.

Mivel a hangsúlyos szótagok intonációs eltéréseit vizsgáljuk, fontos megemlíteni, hogy a francia kötött hangsúlyos nyelv, minden esetben a legutolsó szótagra esik a főhangsúly. A fokozott, érzelmi hangsúly azonban az első vagy második szótagra esik, ez az érzelmi töltet pedig az intonációt is megváltoztatja, melyet a későbbiekben láthatunk (Delattre 1938). Emellett – a szótagok kapcsán – fontos kitérni a francia nyelv alapvető szótagolási szabályaira is, melyek alapvetően megegyeznek a magyar szabályokkal, azonban mégis meg kell említeni néhány lényeges különbséget és sajátosságot. Például a francia nyelv hangzókészletében lévő ún. siklóhangok vagy félmagánhangzók (*semi-voyelles*) – melyeket általában digramma vagy akár trigramma jelöl – nem elválaszthatóak, hiába szerepel két (vagy három) magánhangzó az írásban (a *voir* 'látni' szó így tehát egy szótagból áll: [vwɑʁ], nem „[vo-aʁ]”). A mássalhangzókból álló di- vagy trigrammák sem elválaszthatóak: *ren-sei-gne-ment*, *be-cque-ter*. Az *x* és az *y* intervokális helyzetben maradnak: *exa-men*, *loyer*, valamint a schwát szóvégi pozícióban nem szokás elválasztani: *fa-tigue*. Ha a likvidákat egyéb mássalhangzó előzi meg (felpattanó zárhang vagy réshang), akkor ez a mássalhangzó a likvidával együtt kerül a következő szótagba: *a-gréa-ble*, a két egymás mellett lévő likvidát azonban el kell választani: *par-ler* (Bárdosi & Karakai 2017/3). A vizsgált megnyilatkozások szótagolása során ezen szabályokat alkalmaztuk.

A francia hangsúly és szótagolási szabályok mellett a két vizsgált francia nyelvváltozat intonációjának általános jellemzőit is fontos megemlítenünk. A párizsi vagy Párizs környéki intonáció rendszerint a sztenderd francia intonációs mintázatait mutatja, amelyről azt mondhatjuk, hogy a kijelentő módú, semleges mondatok esetében mindig emelkedő ív, majd ereszkedés a jellemző, vagyis domború dallamgörbe. Ahogy Pataki (1985) említi, az egyszerű mondatok esetében a dallamcsúcs a mondat első részének végén van, ehhez annyit fűzhetnénk hozzá, hogy sokszor az első ritmikus csoport – vagyis az első nominális, verbális vagy egyéb szintagmát jelölő, szintaktikailag és szemantikailag összetartozó fonetikai egység (Lambert-Drache, 1997) – végét jelenti ez a meghatározás. Az összetett mondatok esetében pedig a mondat értelme szerint helyezkedik el a dallamcsúcs (Pataki, 1985), tehát a szintaktikailag szorosan összetartozó elemek vagy egy teljes tagmondat végére esik, majd ezt követi az ereszkedés. Ezt a sztenderd mintázatot a három párizsi beszélő vizsgált mintái is jól tükrözik. Egy, a párizsi/sztenderd nyelvezettől jelentősen eltérő dialektust beszélők intonációját is meg kell említenünk: a québecit. A jelentős nyelvi különbségek azzal magyarázhatóak, hogy a 17. században Kanadába érkező francia telepek többsége az *oïl* nyelvjárás nyugati és középső dialektusait beszélte, tehát alapvetően ezen dialektális sajátosságok – valamint erőteljes angol befolyás – jellemzik a Québecben beszélt nyelvet (Bárdosi & Karakai, 2017/3). A québeci intonációról azt mondhatjuk általánosságban, hogy a dallamgörbe lényegében nagyon hasonló, de a québeci a sztenderd francia intonációnál átlagosan mélyebb, intonációs szempontból nehezebben kivehetőek a szótaghatárok, valamint több „intra-individuális” melodikus változás jellemzi, míg a sztenderd franciában több „inter-individuális” melodikus változás figyelhető meg (Kaminskaïa, 2005). Emellett perceptuális szempontból több nyelvész – mint például Boudreault (1968) vagy Léon & Nemni (1968) – „éneklőbbnek” vagy „dallamosabbnak” írja le a québeci intonációt, ám ez a jelleg vitatott (Kaminskaïa, 2005). Ez utóbbi megállapítást a három québeci beszélő vizsgált mintázatai nem is feltétlenül mutatják, a többi jellemző viszont rendre megfigyelhető az ő esetükben is.

2. Kutatásmódszertan és a korpusz bemutatása

A két említett nyelvváltozat vizsgálatához három-három anyanyelvi beszélő (két-két nő, és egy-egy férfi) tíz-tíz kijelentő módú megnyilatkozásának intonációját elemeztük a Praat (Boersma & Weenink 2019) alkalmazás segítségével, így összesen hatvan mintát vizsgáltunk Cantero elmélete (Baditzné Pálvölgyi 2011; Cantero Serena & Font-Rotchés 2020) alapján. A beszélők pontos adatait az 1. táblázat mutatja:

1. táblázat. A beszélők adatai

PÁRIZS	Beszélő 1	Beszélő 2	Beszélő 3
kor	32 éves	31 éves	70 éves
nem	nő	férfi	nő
foglalkozás	beszédterapeuta	számítógép-hálózati szakmérnök	nyugdíjazott technikus
születési hely	Villeneuve Saint-Georges	Longjumeau	Saint-Claud
QUÉBEC	Beszélő 1	Beszélő 2	Beszélő 3
kor	80 éves	24 éves	57 éves
nem	férfi	nő	nő
foglalkozás	író	énekesnő	közgazdász, ügyintéző
születési hely	Montréal	Charlemagne	Lévis

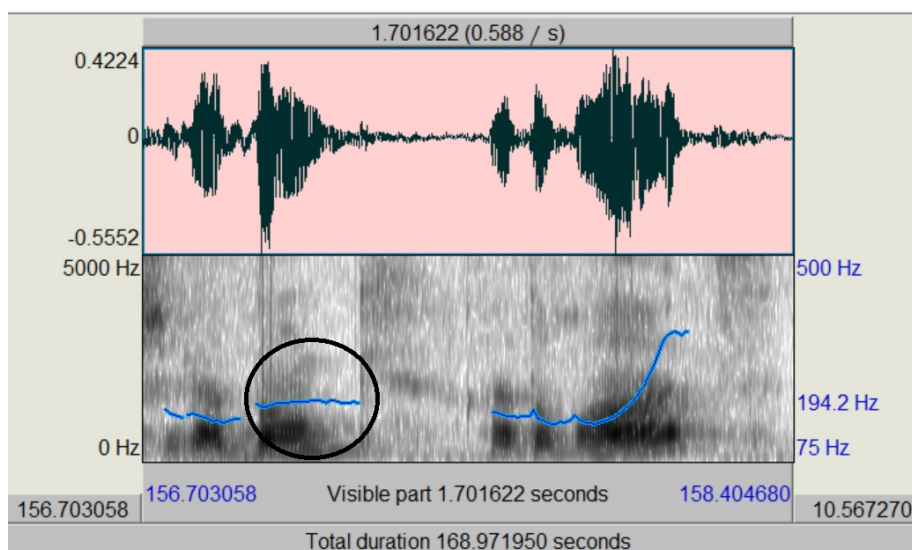
Ahogy látható, a nemek közti eloszlás egyforma, kor és foglalkozás tekintetében azonban nagyobb a szórás. A születési hely szempontjából pedig azt mondhatjuk, hogy a párizsi változatot beszélők mind Párizshoz viszonylag közeli kisebb településekről származnak (egyedül Saint-Claud található messzebb, Párizstól körülbelül 400 kilométerre dél-nyugatra), a québeci változatot beszélők közül ketten Québec déli vagy dél-nyugati részéről származnak (Montréal és a hozzá közel található Charlemagne), egyikük pedig szintén messzebb, Québec

dél-keleti részén született (Lévis körülbelül 240 kilométerre található Montréal-tól).

A párizsi beszélők esetében az *Interactive Atlas of Romance Intonation* (Prieto et al., 2010–2014) oldalon szereplő ún. „térképgyakorlatban” (*maptask*) és egy interjúban lévő megnyilatkozásokból válogattunk beszélőnként tízet, majd elemeztük ezek intonációját a Praat (Boersma & Weenink, 2019) alkalmazás segítségével, a mintákat pedig egy Excel táblázatban gyűjtöttük össze. Az említett „térképgyakorlat” abból áll, hogy a párizsi beszélők egy-egy térképet kapnak, amelyek részben eltérnek egymástól, de ezt a két beszélgetőpartner nem tudja, mivel csak a saját térképüket láthatják, és ilyen módon kell az egyiküknek elírányítani a másikat egy megadott kiindulópontból egy megadott célíg. Ez a spontán beszélgetés került rögzítésre az említett *Interactive Atlas of Romance Intonation* (Prieto et al., 2010–2014) oldalon. Fontos megjegyezni, hogy csak olyan mondatokat vagy mondatrészeket választottunk, melyek semlegesek és kijelentő módban vannak. Az ún. kiemelő (*mise en relief*) szerkezetű mondatokat, melyeknek célja a mondaton belül egy adott információ kihangsúlyozása, nem elemeztük, hiszen azoknak intonációja is eltér a semleges mondatokétól (Pataki 1985). A québeci beszélők esetében – mivel az említett *Interactive Atlas of Romance Intonation* (Prieto et al., 2010–2014) oldalon nem található semmilyen hanganyag Québecet illetően – híres vagy ismertebb québeci személyek Youtube-on lévő interjúiból válogattunk megnyilatkozásokat: Gilles Archambault [1], montreáli író, Céline Dion [2], Charlemagne-ban született énekesnő, valamint Sophie Brochu [3], lévis-i közgazdász. Ezáltal minden, az elemzés szempontjából jelentős információ birtokába juthattunk a québeci beszélőket illetően is, melyet a fentebbi táblázatban tüntettünk fel (kor, nem, foglalkozás és születési hely). Fontos megjegyezni, hogy az interjúk során csak az interjúalanyok spontán, kijelentő módú semleges megnyilatkozásait vizsgáltuk olyan mondatrészeket választva, melyeknek hangminősége nem okozott problémát a Praat (Boersma & Weenink, 2019) alkalmazásban való elemzésben. Ezáltal tehát teljesen spontán beszédben tudtuk vizsgálni a kijelentő módú megnyilatkozások intonációját mind a párizsi, mind pedig a québeci beszélők tekintetében. A pá-

rizsi és québeci változatok hangsúlyos szótagjainak elemzése, illetve pontosabb összehasonlítása érdekében pedig dobozdiagramot, illetve kétmintás t-próbát készítettünk az említett szótagokhoz rendelt százalékos arány felhasználásával. A választott megnyilatkozások 5–11 szótagot tartalmaznak, a két csoport közti szótagszám-arányokra ügyeltünk. A hangsúlyos szótagokat tartalmazó szavak egy–négy szótagszámúak, a legtöbb egyszótagú (60 darab) vagy kétszótagú (44 darab) szó. A három- és négyszótagú szavak sokkal ritkábbak (15 és 5 darab). Egyedül a háromszótagú szavak esetében van meghatározóbb eltérés a két csoport között (4 darab a párizsi mintán és 11 darab a québeci mintán). Az összes megnyilatkozást az alábbi módon elemeztük, két fázisban.

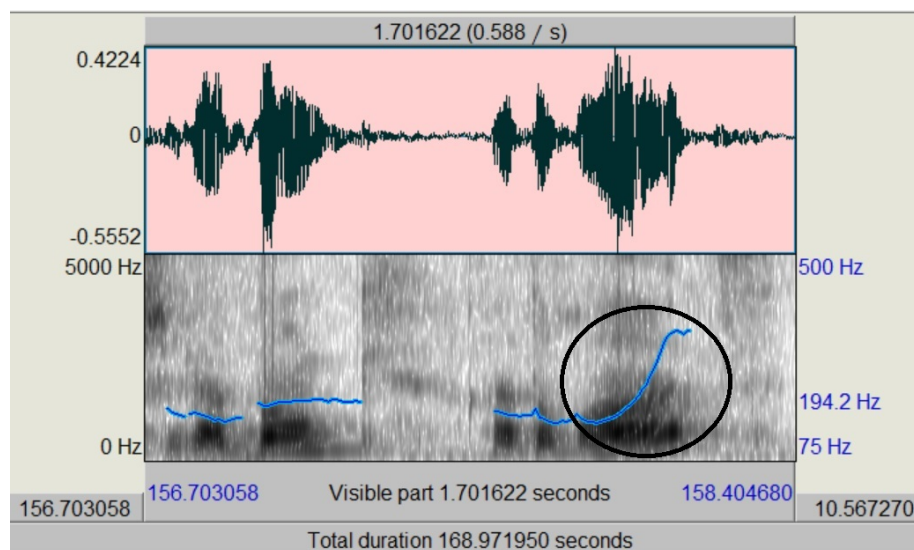
2.1. Első fázis: a szótagok értékének mérése



1. ábra. A *J'ai la banque à droite* 'Tőlem jobbra van a bank' megnyilatkozás intonációja a Praat alkalmazásban. A *banque* szótag alaphfrekvencia-kontúrját láthatjuk bekarikázva.

Az 1. ábrán látható az egyik párizsi megnyilatkozás intonációja a Praat (Boersma & Weenink, 2019) alkalmazásban. A megnyilatkozást először szótagokra bontottuk a bevezetésben említett, francia szótagolási szabályoknak megfelelően – illetve egyéb megnyilatkozások esetében figyelembe véve a beszédtempót

és az ebből következő nagyon gyakori schwa-kiesést például –, majd ezen szótagok Hertzben mért értékét használtuk fel. A képen látható bekarikázott *banque* [bã:k] szótag esetében nem látható túl nagy (10%-ot meghaladó) tautoszillabikus – vagyis egy szótagon belüli – eltérés, ezért két érték megadása helyett az átlagértéket vettük alapul. Azonban olyan esetekben, amelyekben meghaladta a 10%-ot a tautoszillabikus eltérés, a legalacsonyabb és legmagasabb értéket használtuk fel az adatok torzításának elkerülése érdekében. A 2. ábrán látható egy ilyen szótagra példa: az említett megnyilvánulásban a *droite* [dʁwat] szótag (vö. *voir* [vwaw] az első szakaszban).

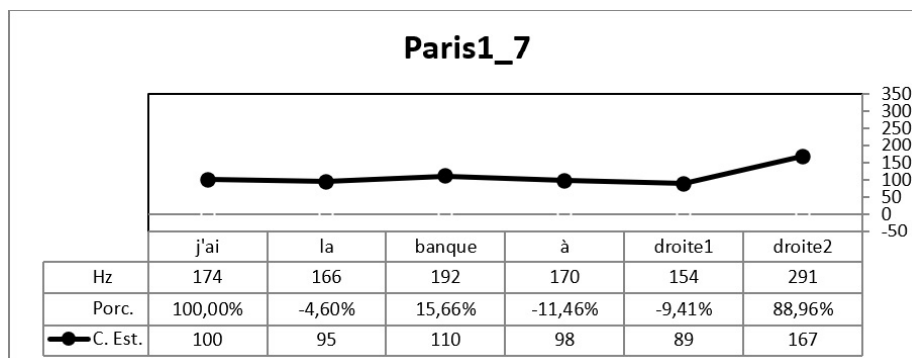


2. ábra. A *droite* szótag alapfrekvencia-kontúráját láthatjuk bekarikázva.

2.2. Második fázis: a mért szótagok értékeinek standardizálása

A mérés után a szótagok Hertzben kapott frekvenciaértékeit standardizáltuk egy Excel táblázatban, mivel a kapott F0 értékek csak az adott beszélő saját hangregiszterére vonatkoznak, így általános következtetéseket nem lehet belőlük levonni. Ezért elkerülhetetlen tehát a standardizáció az elemzéshez.

A két külön tautoszillabikus értéket az egyes és kettes alsó indexszel jelöltük, melyet a *droite* [dʁwat] szótag esetében láthatunk a 3. ábrán. A szótagok



3. ábra. A *J'ai la banque à droite* 'Tőlem jobbra van a bank' megnyilatkozás sztenderdizált intonációs mintája. Az egyes és kettes alsó index a két eltérő tautoszillabikus értéket mutatja.

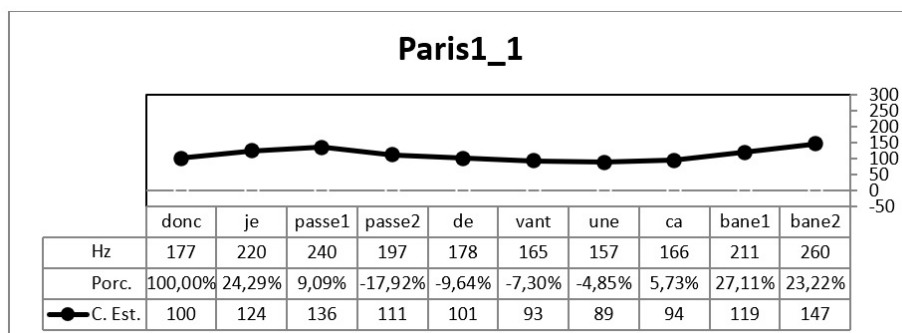
alatt szerepel a Hertzben mért frekvenciaérték, majd a százalékos arány, és végül a sztenderdizált, relatív értékek, amelyekből megkapjuk az ábrán látható intonációs mintát. A frekvenciaértékekből kapott százalékos adatok az abszolút-érték és a rögtön utána következő érték tonális korrelációját mutatják. Vagyis a szótagok egymáshoz viszonyított frekvenciaértékeit százalékban mérjük, az első szótag frekvenciaértéke 100%-ot ér, az utána következő pedig az ehhez viszonyított százalékos értéket kapja. Így tehát a pozitív százalékos érték emelkedést, míg a negatív ereszkedést jelöl. A 3. ábrán például az első szótagnak (*j'ai* [ʒe]) 174 Hz a frekvenciaértéke, a másodiknak (*la* [la]) pedig 166 Hz. Ez 4,6%-os ereszkedést mutat az első két szótag között. Az első szótag a 100-as relatív értéket kapja, a második szótaghoz pedig a 95-ös relatív értéket rendeljük. A végső intonációs mintát pedig ezek a relatív értékek adják ki, amivel elkerülhető, hogy csak az adott beszélőre vonatkozó intonációs sajátosságokat figyeljük. Így általánosabb képet kaphatunk az adott megnyilatkozás intonációjáról, illetve össze tudjuk azt hasonlítani más beszélők – szintén sztenderdizált – megnyilatkozásaival is (Baditzné Pálvölgyi 2011). Az alábbiakban bemutatjuk legfontosabb észrevételeinket a három-három, párizsi és québeci anyanyelvi beszélő kapcsán.

3. A vizsgált nyelvváltozatok

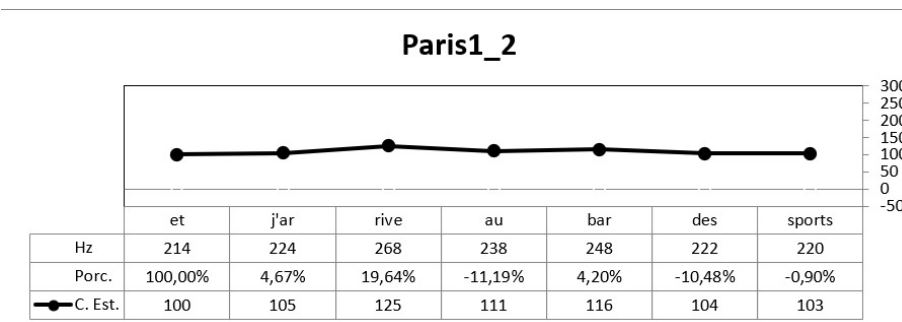
3.1. A párizsi intonáció

A három párizsi beszélő kapcsán a vizsgált eredmények a bevezetésben leírt megállapításokat tükrözik, vagyis a sztenderd francia intonációs mintázatát mutatják. A kijelentő módú megnyilatkozásokban legtöbbször a megnyilatkozás-közi ritmikus csoportok utolsó szótagja, valamint a legtöbb megnyilatkozás legutolsó ritmikus csoportjának utolsó szótagja is emelkedést mutat. Ezen emelkedések valójában rendre a hangsúlyos szótagokat jelölik. Azonban előfordulnak olyan megnyilatkozások, melyeknek legutolsó hangsúlyos szótagja ereszkedést mutat. Ez a különbség azzal magyarázható, hogy egyes megnyilatkozások vége nem egyenlő a beszélők teljes mondatának a végével, azaz egy szintaktikailag szorosan összetartozó egység, egy mondatrész vagy egy teljes tagmondat végét jelölik ezen emelkedések – és ezáltal gyakran a dallamcsúcsot (vagyis a dallamgörbe legmagasabb pontját) is. Sok esetben a mondat végére is eshet ez az emelkedés, amelyet pedig azzal magyarázhatunk, hogy a vizsgált spontán beszéd során a beszélő valószínűleg még nem fejezte be a gondolatmenetét, folytatta volna az adott mondatot, azonban beszélgetőpartnere megakadályozta azt, ezért nem jött létre az ereszkedő ív. Ezáltal azt is konstatálhatjuk, hogy azon megnyilatkozások pedig, amelyekben tapasztalható a megnyilatkozás-végi ereszkedés, a beszélő mondatának és adott gondolatának a végét is jelölik. Ezt demonstrálhatjuk két, egymást követő megnyilatkozás (egy teljes mondat) mintázatának bemutatásával: *Donc je passe devant une cabane, et j'arrive au bar des sports.* 'Tehát elmegyek egy kunyhó mellett és megérkezem a sportbárhoz.' [4] A „*Donc je passe devant une cabane*” megnyilatkozás mutatja az emelkedő ívet, dallamcsúcsa a *-bane* [ban], utolsó szótagra esik – amelyet az ábrán két külön értékkel feltüntetve láthatunk (*bane*₁ és *bane*₂), ez a megjelölés a szótag 10% feletti tonális különbségét mutatja, melyet a (2.) pontban említettünk.

Az „*et j'arrive au bar des sports*” megnyilatkozás esetében pedig a dallamcsúcs utáni ereszkedő ívet látjuk, hangsúlyos szótagja (*sports* [spɔːʁ]) enyhe ereszkedést mutat.



4. ábra. A *Donc je passe devant une cabane* 'Tehát, elmegyek egy kunyhó mellett' megnyilatkozás sztenderdizált intonációs mintája.

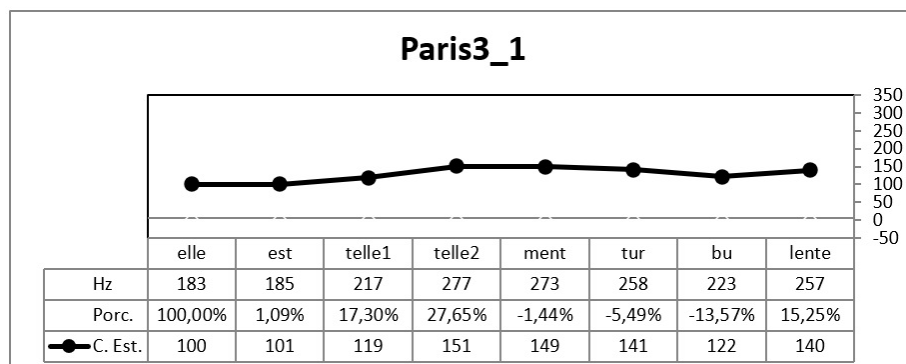


5. ábra. Az *et j'arrive au bar des sports* 'és megérkezem a sportbárhoz' megnyilatkozás sztenderdizált intonációs mintája.

Azt is láthatjuk a **4.** és **5.** ábrán, hogy a mondat eleji és mondatközi hangsúlyos szótagok is emelkedést mutatnak (*passe* [pas], *-bane* [ban] és *-rive* [ʁi:v]).

Ezenkívül azt is megfigyelhetjük, hogy egyes megnyilatkozások kissé eltérő mintázatúak annyiban, hogy nem a ritmikus csoport utolsó szótagja mutat nagyobb emelkedést, hanem az első szótagja. Ezt remekül mutatja az alábbi példa: *Elle est tellement turbulente*. 'Ő annyira izgága.' [5]

Itt nem semleges kijelentő megnyilatkozásról beszélhetünk, hanem a beszélő kihangsúlyozza a *tellement* 'annyira' szót, és egyfajta érzelmi töltetet visz a mondandójába; ilyenkor a hangsúly helye is megváltozik, hiszen – ahogy már említettük a bevezetésben – semleges kijelentő módban az utolsó szótag lenne a



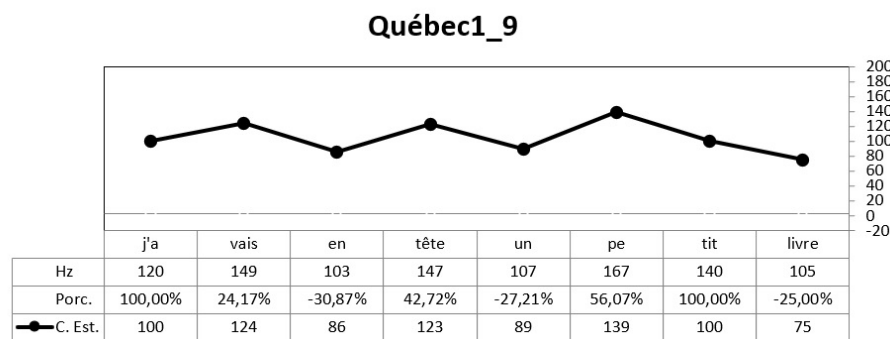
6. ábra. Az *Elle est tellement turbulente* 'Ő annyira izgága' megnyilatkozás sztenderdizált intonációs mintája

hangsúlyos: *tellement* [tɛl'mɑ̃], itt azonban érzelmi hangsúly kerül az első szótagra: *tellement* ['tɛlmɑ̃], ennek tudható be a különbség. A korábbi semleges kijelentések (4. és 5. ábra), illetve ez utóbbi megnyilatkozás (6. ábra) alapján teljes bizonyossággal állíthatjuk, hogy valóban mindig a hangsúlyos szótagok mutatnak emelkedést, ez alól kivételt képeznek a vizsgált megnyilatkozások közül azok, amelyek mondatvégiek, hiszen ott ereszkedő ív tapasztalható, így az utolsó hangsúlyos szótag is ereszkedő. Ezen mintázatokat rendre mutatja a korpuszban szereplő többi minta is. Valamint – ahogy az az első és harmadik diagramon is látható – nagyon gyakoriak a párizsi megnyilatkozásokban a hirtelen, jelentős mértékű emelkedések a hangsúlyos szótagok esetében (pl.: *cabane* [ka'ban] (4. ábra), *tellement* ['tɛlmɑ̃] (6. ábra). Az is megfigyelhető, hogy a hangsúlyos szótagok jelentős többsége akkora intonációs eltérést mutat, hogy két külön tautoszillabikus érték (a legmélyebb és legmagasabb) megadására volt szükség az átlagérték megadása helyett.

3.2. A québeci intonáció összevetése a párizsi mintázatokkal

A három québeci beszélőt illetően a vizsgált eredmények viszonylag jól mutatják az első szakaszban említett megállapítások többségét. A „dallamosabb” perceptuális hatás azonban nem feltétlenül mutatkozik meg. Egyedül a férfi be-

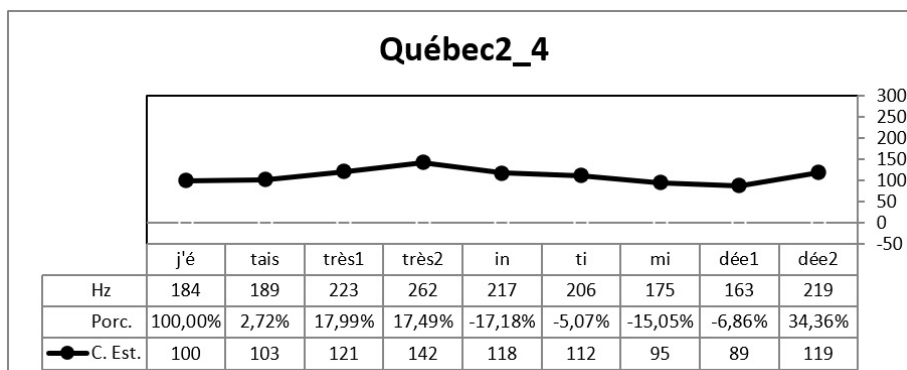
szélő esetében figyelhető meg helyenként, ezt a 7 ábrán láthatjuk: *J'avais en tête un petit livre*. 'Egy kis könyv járt a fejemben.'.[1]



7. ábra. A *J'avais en tête un petit livre* 'Egy kis könyv járt a fejemben' megnyilatkozás sztenderdizált intonációs mintája.

Itt valóban megfigyelhető, hogy a mondaton belül folyamatos az emelkedés-ereszkedés váltakozása szótagonként, nagyjából egyforma mértékben, majd a mondat lezárul a végső ereszkedéssel. Mivel a beszédtempó lassabb – amelyet a beszélő korával lehetne valószínűleg magyarázni, hiszen az idősebb frankofón generációk beszédét általánosságban lassabb beszédtempó jellemzi (Wachs, 1998) –, több ritmikus csoportot tartalmaz a megnyilatkozás, ezáltal több a hangsúlyos szótag is: *-vais* [vɛ:], *tête* [tɛ:t], *pe-* [pə] (a *petit* 'kis' szó nyomatékosítása miatt itt szintén érzelmi hangsúlyról beszélhetünk, ezért az első szótagot kell figyelembe vennünk az utolsó helyett), *livre* [li:vɛ]. A legutolsó szótag kivételével (hiszen az mondatvégi, így ereszkedő) láthatjuk, hogy itt is mindig a hangsúlyos szótagok emelkedőek, ebből a lassabb beszédtempóból és az emelkedő hangsúlyos szótagok jelenlétéből ered a „dallamosabb” jelleg. A fiatalabb, női beszélőknél azonban ez a sajátosság nem fedezhető fel. Az ő esetükben rendre a párizsi beszélőknél tapasztalt dallamgörbékhez hasonló mintázatot figyelhetünk meg, ahogy azt például a 8 ábra is mutatja: *J'étais très intimidée*.

A térképgyakorlathoz képest formálisabb beszédhelyzet ellenére itt is gyorsabb a beszédtempó – akárcsak a párizsi beszélőknél – és a két hangsúlyos szótag

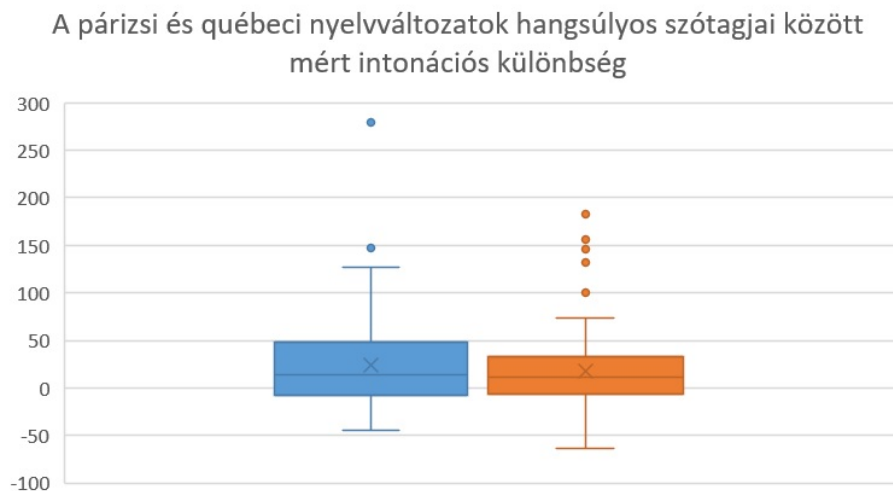


8. ábra. A *J'étais très intimidée* 'Nagyon meg voltam rémülve' megnyilatkozás sztenderdizált intonációs mintája

(*très* [tʁɛ:], -*dée* [de]) mutat emelkedést, illetve a hangsúlyos szótagok esetében szintén két külön tautoszillabikus értéket láthatunk.

Általánosságban a legmélyebb és legalacsonyabb tautoszillabikus értékek között nagyjából egyforma mértékű eltéréseket tapasztalhatunk a párizsi és a québeci beszélők mintázatain is: például majdnem 18%-os ereszkedés és 23,22%-os emelkedés (a *passe* és -*bane* szótagok esetében) az első elemzett párizsi diagramon (4. ábra), 27,65% értékű emelkedés (a *telle-* szótagban) a harmadik elemzett párizsi diagramon (6. ábra), valamint 17,49% és 34,36% értékű emelkedés (a *très* és -*dée* szótagokban) az utóbbi québeci diagramon (8. ábra). A dallamcsúcs esetében mindkét nyelvváltozatban általában nagyobb a két mért érték közti különbség, mint a többi hangsúlyos szótagban. Azonban egy eltérés is észrevehető – amely a korpuszban lévő legtöbb mintán megfigyelhető –, ugyanis míg a tautoszillabikus értékek nem feltétlenül mutatnak jelentős különbséget, a hangsúlyos szótag és az előtte lévő hangsúlytalan szótag között a québeci mintázatokon ritkábban jelennek meg „kiugrásszerű”, hirtelen emelkedések a párizsi mintázatokkal ellentétben. Például az első vizsgált párizsi ábrán (4. ábra) 27,11%-os emelkedés van a penultima (hangsúlytalan szótag) értéke és az utolsó (hangsúlyos) szótagban mért legalacsonyabb érték között (*ca-bane*₁), az utóbbi québeci ábrán (8. ábra) azonban a penultima értéke és az utolsó szótag

legmagasabb értéke között mérhető a körülbelül 34%-os emelkedés (*-mi-dée₂*). Ebből is látszik, hogy a québeci szótagok között kevesebb intonációs eltérés van, kevesebb a hirtelen nagymértékű emelkedés, a szótaghatárok is jobban összemósódnak ebből a szempontból, ahogy azt a bevezetésben is említettük. Emellett pedig az is megfigyelhető, hogy a québeci intonációban az átlagos F0 mélyebb, mint a párizsi, ahogy azt szintén jeleztük az első szakaszban is – azonban lényeges hozzáfűzni azt is, hogy az F0 érték a beszélő saját hangregiszterétől is függ. A két nyelvváltozat hangsúlyos szótagjai között lévő intonációs különbséget a 9. ábrán szemléltetjük részletesebben, amely a hangsúlyos szótagokhoz rendelt, százaléklékban mért értéket mutatja.



9. ábra. Az SPSS statisztikai program kétmintás t-próbájában vizsgált két csoport arányai.

Látható, hogy a vizsgált mintázatok jelentősebb eltérést mutatnak a párizsi (kék) és a québeci (narancssárga) hangsúlyos szótagok intonációja között. A québeci mintán a relatív alaphangmagasság-különbségek kisebbek, a legalsó és a legmagasabb mért érték is alacsonyabban van a párizsi nyelvváltozathoz viszonyítottnak, valamint kisebb szórás is jellemzi. Ellenben a québeci minták több, az átlagostól eltérő, kiugró értéket mutatnak, melyeket az ábrán lévő kisebb pontok ábrázolnak. A 9. ábra alapján is látható, hogy habár kevés beszélő

intonációs mintázatát elemeztük, ezek mégis viszonylag jól tükrözik az alapvető dialektális intonációs különbségeket, melyeket a korábbi kutatások eredményeiként láthattunk (Pataki, 1985; Kaminskaïa, 2005).

Fontos azonban megvizsgálni, hogy ezen különbségek mennyire szignifikánsak a két vizsgált csoport között. Ehhez az SPSS statisztikai programban (Norman, 1968–2021) kétmintás t-próbát alkalmaztunk. A dobozdiagramhoz hasonlóan a párizsi és québeci hangsúlyos szótagokhoz rendelt százalékos értékeket vetettük össze a t-próba során az intonációs különbségek komolyabb összehasonlítása érdekében.

2. táblázat. Az SPSS statisztikai program kétmintás t-próbájában vizsgált két csoport arányai

	Nyelvváltozat	N	Átlag	Szórás	Az átlag standard hibája
%	Párizs	91	23,8974	49,20945	5,15855
	Québec	113	17,2010	41,60999	3,91434

A 2. táblázatban látható, hogy a québeci minta több hangsúlyos szótagot tartalmaz, mint a párizsi (N oszlop). Ez elsősorban annak tudható be, hogy a 80 éves férfi beszélő lassabb beszédtempója miatt több hangsúlyos szótag volt jelen az ő megnyilatkozásaiban. Azt is fontos megjegyezni, hogy a két csoport átlaga, valamint a két csoporton belüli szórás viszonylag nagy eltérést mutatnak: a québeci értékek alacsonyabbak, mint a párizsiak. Ez azon megállapításokra enged következtetni, melyeket a 9. ábra alapján is láthattunk. Azonban az SPSS programban végzett t-próba alapján azt mondhatjuk, hogy szignifikáns eltérés nincsen a párizsi csoport mintájára ($M = 23,89$; $SD = 49,20$) és a québeci csoport mintájára ($M = 17,20$, $SD = 41,60$) kapott eredmények között; $t(202) = 1,05$, $p = 0,294$. A humán tudományokban akkor tekinthetnénk szignifikánsnak az eltérést, ha a szignifikanciaszint (p) 0,05-ös vagy az alatti értéket mutat; a kapott 0,294-es p -érték azonban még tendenciaszerűen sem tekinthető szignifikánsnak. A korábbi kutatási eredmények (Pataki, 1985;

[Kaminskaia, 2005](#) ismeretében valószínűsíthető, hogy a t-próba során kapott eredmény, amely szerint nincs szignifikáns eltérés a két vizsgált csoport között, a jelen előtanulmányunkban felhasznált adatok kis mennyiségének és a nagy szórásértékeknek tudható be.

4. Összegzés

A fentiekben bemutatott előtanulmány célja a párizsi és québeci nyelvváltozatok hangsúlyos szótagjai közti intonációs különbségek vizsgálatának próbája volt, spontán, kijelentő módú megnyilatkozásokban [Cantero Serena & Font-Rotchés \(2020\)](#) Melodic Analysis of Speech (MAS) módszerének felhasználásával. A dialektális összehasonlítást elvetettük, hiszen ahhoz túl kevés beszélő hanganyagát elemeztük, így inkább nyelvváltozatokról beszéltünk. A megnyilatkozásokat az említett Praat [\(Boersma & Weenink, 2019\)](#) alkalmazás segítségével vizsgáltuk, a 2. fejezetben részletesen tárgyalt MAS módszerrel. A mérési fázisban a megnyilvánulásokat a bevezetésben bemutatott francia szabályoknak megfelelően szótagoltuk, a szótagok Hertzben kapott frekvenciaértékét adtuk meg. A mérési fázist követően pedig standardizált, relatív értékeket használva kaptuk meg a dallamgörbéket. A 2. pontban tisztázott százalékos arányokat – és a sztenderdizált intonációs mintákat – felhasználva pedig összevetettük a korpuszban szereplő, összesen 60 megnyilatkozás (30 párizsi, 30 québeci) hangsúlyos szótagjait. A kevés beszélő és a kort, illetve foglalkozást érintő nagyobb szórás ellenére elemzéseinkből ugyanazokra az alapvető eredményekre jutottunk, melyeket korábbi tanulmányok részletes dialektális összehasonlításai is mutatnak. A három párizsi anyanyelvi beszélő kijelentő módú, spontán megnyilatkozásaibanak mintázata teljesen megfelel a sztenderd francia intonációs mintázatnak, amelyet az első szakaszban részleteztünk, vagyis a dallamgörbe mindig dombrú, a hangsúlyos szótagok pedig rendre emelkedést mutatnak – a mondatvégi pozíciót leszámítva, ahol láthattuk, hogy mindig ereszkedés tapasztalható –, illetve a dallamcsúcs is mindig hangsúlyos szótagra esik. A hangsúlyos szótagok esetében legtöbbször két tautoszillabikus frekvenciaérték megadására is szük-

ség volt, valamint ezen szótagok gyakran kiugrásszerű emelkedést mutatnak. A három québeci anyanyelvi beszélő kijelentő módú, spontán megnyilatkozásainak mintázata is többnyire jól mutatja a québeci dialektus korábbi intonációs vizsgálatainak eredményeit és a sztenderd francia intonációhoz viszonyított, korábban is elemzett eltéréseit és hasonlóságait, amelyeket Kaminskaïa (2005) is bemutatótt. Láthattuk tehát saját vizsgálataink során is, hogy a québeci intonációt a párizsi/sztenderd francia intonációhoz hasonlóan domború dallamgörbe jellemzi kijelentő módban, a hangsúlyos szótagok – szintén az ereszkedő, mondatvégi pozíciót leszámítva – mindig emelkedést mutatnak, a dallamcsúcs pedig szintén minden esetben hangsúlyos szótagra esik. Azonban ezen szótagokat hirtelen, kiugrásszerű emelkedés sokkal ritkábban jellemzi, a szótaghatárok így kevésbé különülnek el prozódiai szempontból, valamint a québeci változat összességében mélyebb is a párizsi változatnál. A tautoszillabikus, legalacsonyabb és legmagasabb értékek közti eltérés viszont jelentős különbséget nem mutat a két változat között, továbbá mindkét nyelvváltozatban legtöbbször a dallamcsúcs esetében mérhető a legnagyobb tautoszillabikus eltérés. A 80 éves, québeci férfi beszélőnél a vitatott „éneklőbb” vagy „dallamosabb” jelleg is észrevehető, hiszen – lassabb beszédtempóval és ezáltal több ritmikus csoporttal – több hangsúlyos, emelkedést mutató szótag figyelhető meg egyes mintáiban. Összességében tehát mind a párizsi, mind a québeci változatok vizsgálata hasonló tendenciákat mutat, mint az alapvető dialektális intonációs eltérések, azonban ezen eltérések a kisebb mennyiségű adatok elemzése és a nagy szórásértékek miatt nem szignifikánsak az SPSS programban (Norman, 1968–2021) végzett kétmintás t-próba alapján, így részletesebb dialektális különbségek elemzéséhez további kutatások szükségesek, homogénebb csoportokkal és természetesen több anyanyelvi beszélő spontán megnyilatkozásainak vizsgálatával.

Hivatkozások

Baditzné Pálvölgyi, K. (2011). *Magyarajkú spanyolul tanulók spanyol intonációja: eldöntendő kérdések. Tézisek.* Ph.D. thesis ELTE BTK Nyelvtu-

- dományi Doktori Iskola, Romanisztika Program. Budapest. URL: <http://doktori.btk.elte.hu/lingv/baditznepalvolgyikata/tezis.pdf>.
- Bárdosi, V., & Karakai, I. (2017/3). *A francia nyelv lexikona*. Budapest: Corvina.
- Boersma, P., & Weenink, D. (2019). Praat: Doing Phonetics by Computer (Computer program). Version 6.1.02. URL: <http://www.praat.org>. (2020. dec. 2.).
- Boudreault, M. (1968). *Rythme et mélodie de la phrase parlée en France et au Québec*. Québec/Paris: Presses de l'Université Laval/Klincksieck.
- Cantero Serena, F. J., & Font-Rotchés, D. (2020). *Melodic Analysis of Speech (MAS): Phonetics of Intonation*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Delattre, P. (1938). L'accent Final En Français: Accent D'intensité, Accent De Hauteur, Accent De Durée. *The French Review*, 12, 381502.
- Kaminskaïa, S. (2005). *Une étude comparée de l'intonation de la parole spontanée dans deux dialectes du français sur deux niveaux prosodiques*. Thèse de doctorat. The University of Western Ontario Canada.
- Lambert-Drache, M. (1997). *Sur le bout de la langue – Introduction au phonétisme du français*. Toronto: Canadian Scholars' Press Inc.
- Léon, P. R., & Nemni, M. (1968). Franco-canadien et français standard: problèmes de perception des oppositions vocaliques. In P. Léon (Ed.), *Recherches sur la structure phonique du français canadien* (Studia Phonetica 1) (p. 18–35). Ottawa: Didier volume 1.
- Norman, H. N. (1968–2021). SPSS Statistics (Computer program). Version 22.0.
- Pataki, P. (1985). *Francia nyelvtan a középiskolák számára*. Budapest: Tankönyvkiadó.

Prieto, P., Borràs-Comes, J., & Roseano, P. (Coords.). (2010–2014). Interactive atlas of romance intonation.

Wachs, S. (1998). Le relâchement de la prononciation en français parlé en Ile de France (résumé de Thèse de doctorat de IIIe cycle. *Langage et société*, 85, 126–128.

Források

1. Rencontre avec Gilles Archambault – Partie 1. (2013). Gilles Archambault, québécois íróval készített interjú. URL: https://www.youtube.com/watch?v=li7H7kYyWyY&t=413s&ab_channel=Biblioth%C3%A8queRobert-Bourassa (2020. dec. 8.).
2. Céline Dion RARE interview (1992). Céline Dion, québécois énekesnővel készített interjú. URL: https://www.youtube.com/watch?v=HZirZqnjjZ4&t=12s&ab_channel=CelineDionLT (2020. dec. 8.).
3. Entrevue avec Sophie Brochu (2020). Sophie Brochu, québécois közgazdással készített interjú. URL: https://www.youtube.com/watch?v=c_UylqJ89ro&ab_channel=Radio-CanadaInfo (2020. dec. 8.).
4. A két párizsi beszélővel készített „térképgyakorlat”: (2010). URL: [http://prosodia.upf.edu/iari/enquestes/french/paris/index.html#maptask\(2020.dec.10.\)](http://prosodia.upf.edu/iari/enquestes/french/paris/index.html#maptask(2020.dec.10.)).
5. A párizsi beszélővel készített interjú (2010). URL: [http://prosodia.upf.edu/iari/enquestes/french/paris/index.html#entrevistes\(2020.dec.10.\)](http://prosodia.upf.edu/iari/enquestes/french/paris/index.html#entrevistes(2020.dec.10.)).

Hangos olvasás közbeni hibázások, hibajavítások változása negyedik és ötödik osztály között

Bóna Judit¹, Steklács János²

¹*Eötvös Loránd Tudományegyetem BTK*

²*Pécsi Tudományegyetem BTK*

Abstract

By analyzing oral reading, we can examine the ability to read and many elements of reading comprehension, which play an important role in the diagnosis of reading disorders and development of reading techniques. In the present study, we analyze the changes in oral reading in schoolchildren between 4th and 5th grade.

In our previous research, the correlations between reading time, number of errors, frequency of self-repairs, and frequency and average duration of eye-movements were already examined in recordings of the same children. In the present study, the main question is the following: What eye-movements occur during error-repairs?

Students of average ability from an average primary school in the capital participated in the study. The experiment was performed twice, one year apart: the children were in 4th grade at the time of the first measurement and in 5th grade at the time of the second measurement. The results of oral reading of 10 children (5 boys and 5 girls) are presented in the study, for whom the loss of data was less than 10% in the recordings of the eye-movements.

The children's task was to read aloud the same text from the monitor, in both measurements. Recordings and analysis were performed in both cases with the same instrument (Tobii X120) and software. Speech analysis and evaluation of eye-movements were performed with Praat and ELAN software.

Results show that although the speech rate of oral reading accelerated to 5th grade, the frequency of errors and the time required for self-repairs did not change or increased. However, there were differences between the two measurements in eye-movements during error-repairs. The results draw attention to the fact that reading development would be needed even in the 5th grade.

Keywords: oral reading, eye-movements, reading fluency, error-repairs

Email addresses: bona.judit@btk.elte.hu (Bóna Judit), steklacs@gmail.com (Steklács János)

1. Bevezetés

1.1. Hangos olvasás

Az iskolai olvasástanítás a beszédre épül, a megértés az első tanulási szakaszban erőteljesen az akusztikai szintre támaszkodik. A hangos olvasás azonban nem csupán egy eszköz a szövegértő olvasás képességének elsajátításában, hanem célnak is kell tekintenünk, hiszen a jó olvasó irányában természetes elvárás, hogy képes legyen a hangos, pontos, tagoló, értő, értelmező olvasásra, felolvasásra. Ezen felül természetesen helytálló az a megállapítás is, amely szerint az olvasás tempója, pontossága, artikulációja, a tévesztések aránya összefügg a néma olvasás teljesítményével, a szövegértő olvasás képességével, hatékonyságával (Alt & Samuels, 2011; Kuhn, 2011).

Az olvasás történetében a 20. századot megelőző korszakokban az olvasás egyben a megértést is jelentette. Szövegértéssel kevésbé, inkább a szöveg értelmezésével foglalkoztak, az olvasási képesség alatt leginkább a szép, kifejező, értelmező olvasást értették. Jelentős változást majd csak a 20. század eleje hozott az olvasásról való gondolkozásban, ez elsősorban Luis Emile Javal és Ewald Hering szemmozgáskutatásainak, méginkább pedig Edmund Huey, 1908-ban megjelent, *The Psychology and Pedagogy of Reading* című könyvének köszönhető, ez utóbbi munkában jelent meg először a fluencia leírása is. Az első standardizált hangos olvasás teszt is csak 1917-re datálható (Pearson, 2009). Az a mára már evidenciának számító nézetünk, hogy az olvasás megértése jóval összetettebb jelenség, a folyékonyság csupán előfeltétele ennek az idő távlatában tekintve vissza, nem számít régi keletű felismerésnek.

Az olvasási folyamatot, az olvasott szöveg megértését vizsgáló kutatások, köztük a rendszerszintű nemzetközi felmérések is a néma olvasás eredményességét vizsgálják. A hangos olvasással szemben ennek primátus változása az olvasás történetében is meghatározott korszakra, a 8-12. század közötti szakaszra tehető (Adamik, 2004; Benczik, 2001; Manguel, 2002; Steklács, 2017). Ettől az időszaktól kezdve, napjainkban pedig különösen a néma olvasásé a főszerep a mindennapi életben. Az olvasás tanulásának kezdeti szakaszában viszont a szó-

beliségnek, a fonéma-graféma megfeleltetési szabályok, az összeolvasás, a szavak önálló felismerésének elsajátítási szakaszában, folyamatában a hangos olvasásnak van meghatározó szerepe. A hangoztatás később belső, néma beszéddé alakul, később pedig többnyire eltűnik a szubvolkalizációs folyamatban.

A hangos olvasás funkciója az egyén életében nem tűnik el teljesen, mindannyiunknak szüksége van a felolvasás képességére, amely az írott nyelv megszoaltatását jelenti, értést, értelmezést és interpretálást, értelmeztetést is magában hordoz. A mindennapi funkció mellett a hangos olvasás rendelkezik egy megkülönböztetett iskolai szereppel, nevezetesen, hogy ennek vizsgálatával tudjuk felismerni, majd javítani a tanulók szövegértési folyamatának alapját képező számos olyan elemet, amelyek hiánya a megértés deficitjéhez is vezet. Ilyen mutatók a tempó, megakadások, hibák száma és típusai.

1.2. Fluencia

Leegyszerűsítve a fogalom összetett meghatározásait azt mondhatjuk, hogy az olvasási fluencia, folyékonyság fogalma alatt az olvasástechnika automatizált szintjét értjük. Fluens és nem fluens olvasó között a legfőbb különbség, hogy a fluens olvasó rendelkezik azzal a szintű automatizáltsággal, amely lehetővé teszi számára, hogy figyelmét, kognitív erőforrásait a dekódolási folyamatok helyett a jelentés konstruálására, a megértésre, önmaga olvasási folyamatának a monitorozására és további metakognitív tényezők működtetésére fordítsa. Az az olvasó, aki nem rendelkezik a megfelelő szövegértési szint alatti folyamatok automatizmusával természetesen hátránnyal rendelkezik a megértés terén az automatizált dekódolást végezni képes társaival szemben. Mindezek mellett a munkamemória hatékony működése szempontjából bizonyos olvasási tempó elérésére is szüksége van a jó olvasónak, amely meglétéhez szintén automatizált dekódolási folyamatra van szükség. A megfelelő tempót nemcsak a nyelvi egységek gyors felismerése biztosítja, hanem ezzel szoros összefüggésben a tévesztések alacsony száma, illetve ezek gyors, hatékony javítása is, valamint az írott nyelv szabályaira vonatkozó tudás (Kamil et al., 2011; Kuhn, 2011; Rasinski et al., 2011). A fluencia legfontosabb meghatározóit, tényezőit tehát a következőképpen foglalhatjuk össze:

automatizált dekódolási folyamatok; fonéma-graféma azonosítás, szófelismerés, fonológiai és morfológiai tudatosság, metanyelvi tudás, nyelvi tudatosság, amely az adott nyelv írásrendszerére és az adott nyelvre általánosságban is vonatkozik, a megértéshez szükséges olvasási tempó, illetve alacsony tévesztési arány minden nyelvi szinten.

Mint már utaltunk rá, a szövegértés szempontjából kiemelkedően fontos a fluencia szintje. Leegyszerűsítve az összefüggéseket azt mondhatjuk, hogy a fluencia és az olvasás megértési szintje között erős korrelációt tapasztalunk hangos és néma olvasás esetében is, hiszen a folyamat automatizáltságának szintje meghatározza, hogy az olvasó mennyi erőforrást tud a szövegértési folyamatra koncentrálni (Hudson et al., 2015; Kamil et al., 2011; Kang & Y. Shin, 2019).

Az olvasás folyékonyága, a fluencia értelmezésében is elkülöníthetünk nyelvészeti és pszichológiai természetű nézeteket. A nyelvészeti értelmezés általában azokkal a leginkább hangos olvasást érintő tényezőkkel egészül ki, amelyek az írás kihangosítására, hangzóvá, beszéddé, beszélt nyelvvé tételére vonatkoznak. Ha hangosan olvasunk, akkor elkerülhetetlen, hogy tagoljunk, értelmezzünk is, ehhez pedig szükségünk van a szuprasegmentális tényezők alkalmazására (Baker et al., 2008; McGill-Franzen & Allington, 2011). Fontos látnunk tehát, hogy a hangos olvasást tekinthetjük és vizsgálhatjuk olvasási és beszédproduktions folyamatként is.

1.3. Az olvasási folyamat szemkamerás vizsgálata

A szemmozgás vizsgálata alapjául szolgáló tudományos munkák az 1800-as évek végén jelentek meg, Luis Emil Javal francia szemész és Ewald Hering német pszichológus nevéhez fűződnek. Mindketten a binokuláris látást, a szemmozgások típusait írták le nagyjából egy időben. Mostanra a műszeres technológia fejlődésének köszönhetően nagyon pontos leírását tudják adni a kutatások a szem információfeldolgozás, olvasási folyamat során végzett tevékenységének. A két kulcsfontosságú fogalom a szemmozgások tekintetében a szakkád és fixáció. Szakkádok alatt a szem ugrásszerű mozgásait értjük, a fixáció pedig azokat a pontokat jelenti, ahol a szem megáll, ilyenkor zajlik a vizuális informá-

ció felvétele és feldolgozása. A fixáció hosszát a látott vizuális inger minősége, feldolgozhatósága, információs telítettsége határozza meg. A fluens olvasó rövidebb, kevesebb fixáció segítségével, ritmikusabb szakkadikus mozgásokkal olvas egy szöveget. Magát a folyamatot természetesen más tényezők is befolyásolják, ilyen például az olvasó motiváltsága, ébersége, a szöveg tipográfiai sajátosságai, nyelvi szerkesztettsége, nehézsége. A jó olvasó átlagosan 7–9 karakter távolságú, negyed másodpercnyi hosszúságú szakaszokban fixál, bár nagy különbségeket tapasztalnak a vizsgálatok. A tanuló és gyenge olvasó átlagosan rövidebb szakkádokkal, hosszabb fixációkkal olvas, olvasási folyamatát arányaiban sokkal több visszafelé ható, regresszív szakkád jellemzi (Rayner, 1983; Conklin et al., 2018; Steklács, 2019).

Hangos olvasás vizsgálata tekintetében a szem-száj késleltetést fogalmát kell megemlítenünk, ami arra utal, hogy a felolvasás során a szemünk, a vizuális feldolgozás előbb jár, mint az akusztikus kivitelezés, azaz a beszédünk. Ez a késleltetés limitált ugyan, bizonyos távolságnál nem lehet nagyobb, viszont elmondhatjuk, hogy a jó, fluens olvasónál nagyobb, mint a gyenge és tanuló szakaszban lévő olvasók esetében. A nagyobb távolság természetesen több lehetőséget nyújt a kontrollra, a tévesztések gyors korrekciójára, ami következményeképpen a fluens olvasó még kevesebbet hibázhat, téveszthet a beszédprodukciós kivitelezési szakaszban. Hasonló jellemzőkről számolnak be a kutatások a hangszeres játék esetén is (Rayner et al., 2004; Duchowski, 2007; Vorstius et al., 2014; Steklács, 2014, 2019).

1.4. A hangos olvasás szimultán elemzése szemkamerás műszerrel és akusztikai-fonetikai vizsgálat segítségével

A hangos olvasás tehát vizsgálható olyan produkciós, illetve reprodukciós tevékenységként, beszédprodukciós és olvasási folyamatként is, amely egyszerre mutatja, viseli magán a beszélt és írott nyelvi kommunikáció jellemzőit. Elmondhatjuk azt is, hogy az olvasás közbeni tévesztések, temporális jellemzők elemzésekor érdemes azok természete alapján a két kategória (beszélt és írott nyelvi) mentén csoportosítani. Ugyanez érvényes a hibajavítások folyamatára

is. A mostani és ennek előzményeként tekinthető vizsgálatainkban (Bóna & Steklács, 2020; Steklács & Bóna, 2020) módszertani szempontból új perspektívából elemeztük a jelenséget. A hangos olvasást hang- és videofelvételen egyaránt rögzítjük, majd az említett jellemzőket egyszerre vizsgáljuk beszédelemző szoftverrel és szemkamerás, tekintetkövető műszerrel, így tapasztalataink szerint egyrészt az eddig ismert módszerekhez képest pontosabb képet kapunk, másrészt relevánsabb következtetéseket tudunk megállapítani a hangos olvasás közben zajló olvasási, megértési és beszéd folyamatokról. Véleményünk szerint ezeknek a jelenségeknek a megismerésével nem csupán az adott korosztály olvasási folyamatának jellemzőiről, azok változásáról tudhatunk meg többet, de az osztálytermi munkában, az olvasástanítás tervezésében is segítségünkre lehetnek az eredmények. A negyedik és ötödik osztály közötti különbségek azért is külön érdekesek számunkra, mert két iskolafok közötti átmenet, valamint az olvasástanítás, képességfejlesztés megszűnésének hatásaira is kíváncsiak vagyunk, amely közvetlen hatással van többek között a tanulási eredményekre, hiszen a felső tagozattól sokkal nagyobb jelentőséggel bír a tanuló életében az önálló tanulás, szövegfeldolgozás képessége.

Előző munkánkban (Bóna & Steklács, 2020) vizsgáltuk már a hibajavítások és a szemmozgások összefüggéseit negyedik osztályban. Azt találtuk, hogy a hangos olvasás során mért beszédtempó és a szünetek gyakorisága összefüggésben áll egymással, ugyanakkor az olvasás gyorsasága nem függ össze a hibagyakorisággal és a javítás arányával. Az olvasás tempója és fluenciája összefüggésben áll a szemmozgások gyakoriságával és a fixációk időtartamával is. A nehezebben javítható hibák javításához a beszélőnek több időre van szüksége, ami egyrészt a közlés későbbi leállításában, másrészt a hosszabb szerkesztési szakaszokban mutatkozhat meg. A szemmozgás jellemzői is jelzik a nehezebben végbemenő javítási folyamatokat: ilyenkor az adatközlő vagy hosszan fixál a szón, vagy többféle stratégiát is alkalmaz a hibajavítás közben.

Ezeket az eredményeket negyedik osztályos tanulók hangos olvasásának elemzésekor állapítottuk meg. Nincsenek adataink azonban arról, hogy a későbbi életkorokban milyen szemmozgások és hibajavítási stratégiák jellemzőek. Ezért

egy év elteltével ismét megvizsgáltuk ugyanazon iskolás gyermekek hangos olvasását, akikkel a korábbi (Bóna & Steklács, 2020; Steklács & Bóna, 2020) kísérleteket végeztük. Mostani vizsgálatunk előtt a következő kérdéseket fogalmaztuk meg: 1. Hogyan változik a hibázások előfordulása negyedikről ötödik osztályos korra? Milyen típusú hibákat vétenek és milyen gyakorisággal ugyanazon tanulók a két osztályfokon? 2. Milyen szemmozgások történnek az egyes hibajavítások szóbeli produkciója közben? 3. Változtak-e a gyermekek hibajavítási stratégiái, illetve az azokhoz kapcsolódó szemmozgások egy év alatt?

Hipotéziseink szerint 1a. ötödik osztályra a gyermekek kevesebb hibával olvassák fel ugyanazon szöveget. 1b. A hibák típusainak aránya nem változik. 2. Mindegyik osztályfokon megfigyelhető mindegyik szemmozgástípus, amelyet a korábbi vizsgálatunk során megfigyeltünk a hibajavítások közben. 3. Ötödik osztályra a gyermekek gyorsabban javítják a hibáikat, és a hibajavítások szemmozgások típusainak aránya is megváltozik.

2. Adatközlők, anyag és módszer

A jelen elemzésünk egy korábban már bemutatott, de más szempontból vizsgált kutatási anyag (Steklács & Bóna, 2020) adatait mutatja be a hibajavítások jellemzőire fókuszálva. A korábbi vizsgálat az olvasás tempóját, a szüneteket, a bizonytalansági megakadások és hibák, illetve a szemmozgások gyakoriságát és a fixációk idejét elemezte, de nem foglalkozott a hibajavítások módjával, temporális jellemzőivel, illetve a hibajavítások közbeni szemmozgások típusaival.

A vizsgálatban egy átlagos fővárosi általános iskola átlagos képességű tanulói vettek részt. Az adatközlők kiválasztása véletlenszerűen történt, a diákokat a tanítójuk választotta ki a vizsgálatra. Két alkalommal rögzítettük az adatközlők felolvasását, egy év különbséggel, a tanév ugyanazon hetének ugyanazon napján, februárban. A gyermekek az első mérés időpontjában 4. osztályba, a második méréskor 5. osztályba jártak. A második mérés időpontjában nem minden gyermek járt már az iskolába, illetve volt olyan is, aki aznap beteg volt, ezért tanulmányunkban 10 olyan tanuló (5 fiú és 5 lány) hangos olvasásának az

eredményeit mutatjuk be, akik mindkét alkalommal részt vettek a vizsgálatban. Az adatvesztés a szemkamerás felvételeken minden esetben 10% alatti volt. A gyermekek magyar anyanyelvűek és tipikus fejlődésűek voltak, látásuk megfelelő volt a képernyőről való olvasáshoz (vagy élesen láttak, vagy megfelelően korrigálta a látásukat a szemüveg).

A felvételeket mindkét esetben ugyanazzal a műszerrel (Tobii X120) és szoftverrel rögzítettük. A szemkamera a monitor aljára volt rögzítve, a gyermekek – akik egyesével vettek részt a kísérletben – mintegy fél méterre ültek a monitor-tól. A feladat megkezdése előtt a szemkamerás műszert egyénenként kalibráltuk. Ezután arra kértük az adatközlőket, hogy olvassák fel hangosan a monitoron megjelenő szöveget (ez megegyezett a BEA és a GABI adatbázisok hangos olvasási feladatában használt szöveggel, [Gósy et al. 2012](#), [Bóna et al. 2014](#)). A szöveg egy kérdő formájú címből és 12 kijelentő mondatból állt, a szövegszavak száma 234, a karakterszám 1816 volt. A feladatra vonatkozó utasítás így szólt: *Olvasd el hangosan, figyelmesen a következő szöveget! Ha végeztél, mondd el, hogy mi számodra a legfőbb üzenete!* A tartalomra vonatkozó kérésre azért volt szükség, hogy a gyermekek a szöveg tartalmára is figyeljenek a felolvasás közben. A szöveget három közvetlenül egymást követő monitorképen jelenítettük meg, a „lapozás” idejét a gyermek olvasási tempója határozta meg.

A gyermekek hangos olvasása közben a szemkamerával rögzítettük a tekintetük mozgását, a fixációkat és a szem szakkadikus tevékenységét. A szemmozgással párhuzamosan az akusztikai jel is rögzítésre került. A szemmozgás és a hang rögzítése egy videofájlba történt. Ezután a hangfelvételeket a Praat 5.0 szoftverrel ([Boersma & Weenink 2008](#)) annotáltuk, időtartamméréseket végeztünk rajtuk, illetve elemeztük az olvasás közbeni hibajavításokat. Háromféle hibatípust különítettünk el: a betoldást, a kihagyást és a cserét. Betoldáskor a felolvasó egy vagy több (nem leírt) plusz hangot, szótagot vagy szót illeszt bele a felolvasott szövegbe. Kihagyáskor a leírt szövegből kimarad valami (hang, szótag, szó vagy több elem) a felolvasás során. Csere esetén a felolvasó más hangot vagy szót ejt, mint ami a szövegben szerepel. A hibajavítások elemzésekor a következő paramétereket vizsgáltuk meg: a hibásan kimondott közlés

szótagszáma a hibás szó kezdetétől a megszakítási pontig, a hiba kezdetétől a megszakítási pontig eltelt idő, a szerkesztési szakasz időtartama, a hibás szó kezdetétől a hibajavítás kezdetéig eltelt idő. Emellett kiszámítottuk azt is, hogy milyen beszédtempóval ejtette ki az adatközlő a hibát tartalmazó közlést.

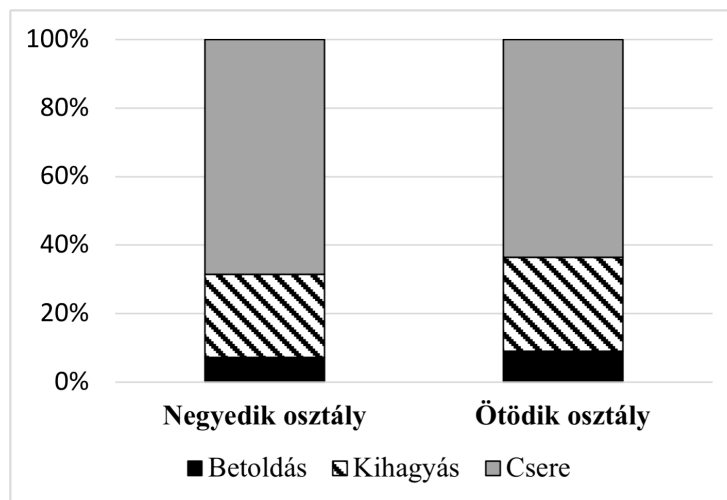
A szemkamerás felvételeket az ELAN 5.5 szoftverrel elemeztük. Azt vizsgáltuk meg, hogy milyen szemmozgások történnek akkor, amikor a hangos olvasásban hibajavítás történik. A szemmozgásokat ötféle típusba soroltuk [Bóna & Steklács (2020)] kategóriáinak megfelelően: 1. a szem szótagonként letapogatja a hibásan ejtett szót/szövegrészt (letapogatás szótagonként), 2. hosszú fixáció történik (egy hosszú fixáció), 3. egy regresszív szakkádot követően a szem ismét előrehaladó mozgást végez (regresszió), 4. bár a beszédben hibajavítás van, a szemmozgásban nem történik változás, a szem előrehaladó mozgást végez (továbbmegy), 5. a hiba javítása közben többféle szemmozgás történik (több egyszerre).

A számszerűsíthető adatokon (szótagszám és időtartamok) statisztikai elemzést (Univariate ANOVA) végeztünk az SPSS 20 szoftverrel 95%-os konfidenciaszinten.

3. Eredmények

A felolvasásokban negyedik osztályban összesen 70, ötödik osztályban 66 hibajavítást adatoltunk. Ez azt jelenti, hogy negyedik osztályban 3,9 hiba, míg ötödik osztályban 3,6 hiba fordult elő 100 karakterenként. Megfordítva, ha azt vizsgáljuk meg, hogy hány karakterenként fordult elő egy olvasási hiba, akkor ez az érték negyedikben 25,9 karakter, ötödikben 27,5 karakter volt.

Megvizsgáltuk, hogy milyen típusú hibákat javítottak az adatközlők. Mindkét osztályfokon a csere volt a leggyakoribb hibatípus, míg a betoldás a legritkább [1] ábra). A háromféle hibatípusra és a javításukra az [1] táblázatban olvashatók példák. A legtöbb nehézséget a *daganatos megbetegedést* szókapcsolat okozta, a *megbetegedést* negyedikben tízszer javították az adatközlők, míg ötödikben ötször.



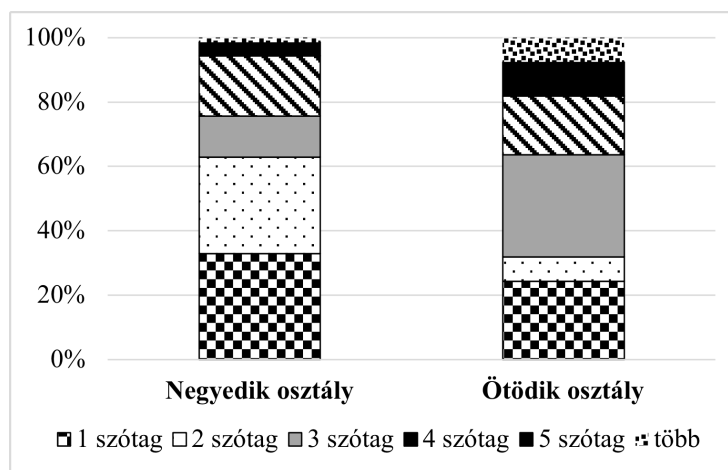
1. ábra. A javított hibák típusainak aránya a két osztályfokon

A kihagyás is hasonló arányban fordult elő negyedik és ötödik osztályban, az előbbiben 17, az utóbbiban 18 százalékban. A betoldás volt a legritkábban előforduló javított hiba, 5 és 6 százalékban találkoztunk vele.

Ezután azt elemeztük, hogy a hibás szó kezdetétől hány szótag kiejtése után állították le az adatközlők a hibát tartalmazó közlést. Az eredmények szerint az adatközlők negyedik osztályban átlagosan 2,4 szótag után, míg ötödik osztályban 3,1 szótag után állították le a hibás közlést (2. ábra). Negyedikben a közlések megszakításának 62,9%-a egy vagy két szótag kiejtése után történt meg, míg ötödikben ez az arány 31,8%-ra csökkent. A statisztikai elemzés szerint a két osztályfokon mért adatok között szignifikáns különbség volt a megszakítási pont előtti szótagszám tekintetében: $F(1, 134) = 6,499$; $p = 0,012$.

1. táblázat. Példák az egyes hibatípusokra (SIL = néma szünet, félkövérrel szedtük a hibát tartalmazó szavakat)

Hibatípus	Példa
Betoldás	<i>súlyosabb problémámat SIL problémát</i> (4. osztály)
	<i>szennyezheti SIL az SIL szennyezheti a talajt</i> (4. osztály)
	<i>permeteszer SIL a per SIL permetszerek</i> (5. osztály)
	<i>zöldségekben SIL gyümölcsökben SIL a zöltségbe SIL a gyümölcsök SIL be</i> (5. osztály)
Kihagyás	<i>daganatos megbetegés SIL megbetegedéseket okozhatnak</i> (4. osztály)
	<i>a kismamák sem sejtik hogy a bébite SIL bébi SIL ételek</i> (4. osztály)
	<i>nem keze SIL nem a kezelni</i> (5. osztály)
	<i>növények szerek SIL növényvédő szerek</i> (5. osztály)
Csere	<i>egy lakosra két kiló méter SIL mérreg jut</i> (4. osztály)
	<i>a vő SIL a vizsgálatok szerint a hazai eredetű termékekben</i> (4. osztály)
	<i>importált alakú SIL alak SIL a SIL alú SIL na SIL importált árukban</i> (5. osztály)
	<i>normális SIL hor SIL hor hormonális</i> (5. osztály)



2. ábra. A hibás közlésből kiejtett szótagok száma a megszakítási pont előtt

Megvizsgáltuk az önjavítások temporális jellemzőit is (2. táblázat). Az eredmények azt mutatják, hogy ötödik osztályban mindegyik vizsgált érték nagyobb, azaz a javítások átlagosan hosszabb időt vettek igénybe, mint negyedik osztályban. A statisztikai elemzés szerint sem a hiba kezdetétől a megszakításig eltelt időben, sem a szerkesztési szakasz időtartamában nem volt szignifikáns a különbség. Ugyanakkor a hiba kezdetétől a javítás kezdetéig eltelt idő szignifikáns különbséget mutatott a két mérési pont között: $F(1, 134) = 4,150$; $p = 0,044$.

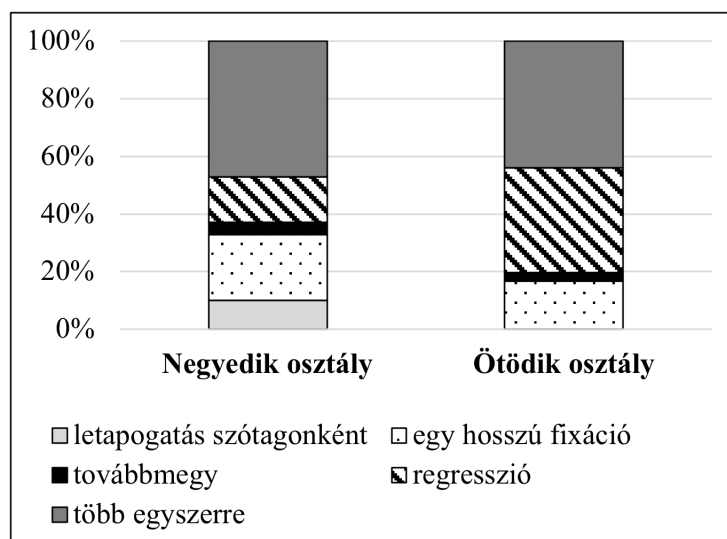
2. táblázat. A hibajavítások temporális jellemzői (ms)

	4. osztály		5. osztály	
	Átlag	Szórás	Átlag	Szórás
A hiba kezdetétől a megszakításig eltelt idő	724	515	806	505
A szerkesztési szakasz ideje	779	905	1158	1396
A hiba kezdetétől a javításig eltelt idő	1503	1061	1964	1547

Kiszámítottuk azt is, hogy a hibás közléseket milyen tempóval ejtette ki a beszélő. Negyedik osztályban az átlagos beszédtempó 3,6 (szórás:1,1) szótag/s, ötödikben 4,2 (szórás:1,0) hang/s volt. A statisztikai elemzés szerint szignifikáns különbség volt a két mérési pontban számított tempóértékek között: negyedikben az adatközlők beszédtempója szignifikánsan lassabb volt, mint ötödik osztályban ($F(1, 134) = 11,023$; $p = 0,001$).

Ezután a hibajavítások alatti szemmozgások elemzésére került sor. Megvizsgáltuk, hogy a hiba megjelenése és javítása közben milyen szemmozgások jelennek meg a korábban felállított kategóriák szerint (Bóna & Steklács 2020). Negyedik osztályban öt, ötödikben négy típusba tudtuk besorolni őket (3 ábra). Mindkét osztályfokon az volt a leggyakoribb, hogy ha tévesztett az adatközlő, akkor a javításhoz többféle szemmozgásra volt szüksége. Ezek közül negyedikben a leggyakoribb eset az volt, hogy egy hosszú fixáció és egy regresszió kombinálódott, de előfordult olyan eset is, amikor a regressziót vagy a fixációt szótagonkénti letapogatás követte, vagy akár mindegyik szükséges volt a javításhoz (azaz nem volt elég a hosszú fixáció, hanem regresszió, majd letapogatás után történt meg a javítás). Ötödikben ugyanezen adatközlők tévesztéseinél megnőtt azon esetek száma (az összes javítás 15%-a lett), amikor kettőnél többféle szemmozgásra volt szükség a hibajavításhoz. Ötödik osztályban azonban már egyáltalán nem jelent meg önállóan a szótagonkénti letapogatás. A 3 táblázat olyan példákat mutat be, amelyek javítása többféle szemmozgással járt együtt.

A szemmozgások és a hangzó beszédben történő hibajavítások időzítési jellemzői is összefüggést mutatnak. A 4 ábrán megfigyelhető, hogy az egyes szemmozgástípusok milyen hosszú javítási időt igényeltek a hibás szó kezdetétől. A leghosszabb időtartam a többféle szemmozgással jár együtt, míg azoknál a hibáknál, amelyeket a beszédben a legrövidebb idő alatt lehet javítani, a szem sem áll meg, hanem tovább olvas. Kevesebb idő kell azokhoz a javításokhoz, amelyekben a szem visszaugrik egy korábbi pontra, mint amikor megáll a hibán, és hosszan fixálva javítja azt.



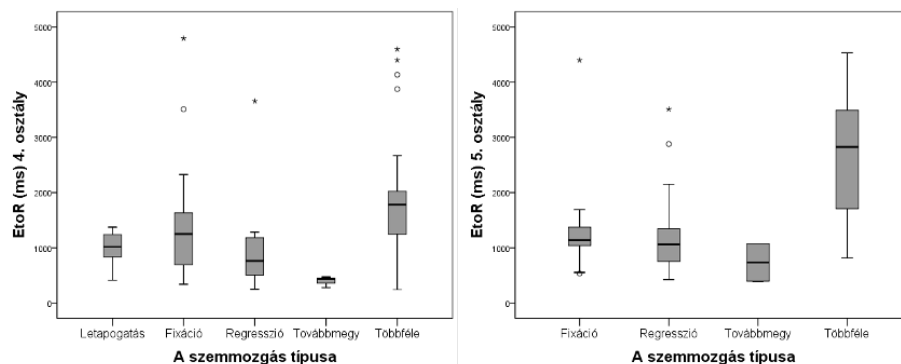
3. ábra. A szemmozgások típusainak aránya hibajavítás közben negyedik és ötödik osztályban

3. táblázat. Példák a többféle szemmozgástípussal együtt járó hibákra (SIL = néma szünet)

A szemmozgás típusa	Példa
fixáció+regresszió+fixáció+letapogatás	<i>premet SIL szerek SIL a permetezett SIL permetszerek</i>
fixáció+regresszió+fixáció+regresszió+ letapogatás	<i>az iparta SIL az importált</i>
fixáció+regresszió+fixáció+regresszió	<i>lemosható SIL lemosatlan SIL lemoshatatlan</i>
regresszió+fixáció+regresszió	<i>csatlakoz SIL család SIL családhoz</i>

4. Következtetések

Vizsgálatunkban tíz tanuló hangos olvasását elemeztük a hibajavítások szempontjából. Az előzetesen megfogalmazott hipotéziseink közül csak egy részhiptotézist sikerült alátámasztanunk, a többi nem igazolódott. Nem volt ugyanis különbség a hibák gyakoriságában a két osztályfok között (az 1.a hipotézis nem



4. ábra. A szemmozgás típusának és a hibajavítás időtartamának összefüggései negyedik és ötödik osztályban (EtoR = Error-to-repair 'hibajavítás')

igazolódott), ugyanakkor a hibák típusainak aránya nem (csak minimálisan) változott: mindkét mérési pont esetében a csere volt a leggyakoribb hibatípus (az 1b. hipotézis igazolódott).

A szemmozgások típusaira és a hibajavítások gyorsaságára vonatkozó 2. és 3. hipotéziseink sem igazolódtak. A vizsgálatunkban résztvevő tanulók alacsony száma miatt nem következtethetünk általános tendenciákra, a kapott eredmény alapján viszont érdemes néhány jellegzetességet jobban megvizsgálni.

Előző tanulmányunkban bemutattuk, hogy a negyedik osztályos teljesítményhez képest rövidült a felolvasás ideje, kevesebb, átlagosan rövidebb fixációkkal történt az olvasás (Steklács & Bóna, 2020). Ezek a folyamatok az automatizáltság, fluencia enyhe pozitív változását jelzik. Mostani munkánkban a hibajavítások számának kismértékű csökkenését, a hibatípusok állandósulását tapasztaltuk, ez utóbbi kapcsán a csere-jellegű volt a legjellemzőbb a tanulókra mindkét adatfelvétel során. A hibás szó kezdete után ötödik osztályban a tanulók később álltak meg az olvasással. Ezzel együtt azt tapasztaltuk az önjavítások temporális jellemzőivel kapcsolatban, hogy ötödik osztályban a hiba kezdetétől a javításig eltelt idő szignifikáns mértékben hosszabb volt. Mindezek háttérét vizsgálva érdemes elgondolkodnunk azon, hogy a későbbi megállás, a hosszabb javítás a felgyorsult olvasás, aktívabb antipációs és metakognitív önmonitorozási folyamatok eredménye-e, vagy más állhat a háttérben.

A hibajavítás közben tapasztalt szemmozgások elemzése azt mutatta, hogy az ötödikesek esetében megnőtt azoknak a kombinációknak a száma, ahol a kettőnél többféle szemmozgás volt jellemző, a szótagonkénti letapogatás eltűnt. Ez a változás szintén magyarázható a tempó gyorsulása, az automatizáció, a fluencia fejlődésével. Ha ezt a tendenciát megerősítené egy nagyobb számú adatközlővel végzett vizsgálat, akkor bizonyítottnak tekinthetnénk, hogy az olvasási folyamat esetében a javítások közben végzett szemmozgások is változnak, fejlődnek.

Magyarországon jelenleg egy olyan formalizált olvasásvizsgálati eljárás van, amely méri a fluens olvasás időt, valamint hibaszámot (Meixner 2000; Sipos 2019), ugyanakkor a javítások módjához kapcsolódóan nincs sem a tantervekben, sem egyéb felmérésekben olyan standardizált teljesítményérték, amelyhez a most kapott eredményeket viszonyíthatnánk. Mégis azt gondoljuk, az önálló tanuláshoz, az ötödikes tananyagban szereplő szövegek feldolgozásához kívánatos lenne, hogy az olvasás fluenciája fejlődést mutasson az előző évi eredményekhez képest abban a tekintetben is, hogy csökken a hibázások mértéke, és gyorsabb a javítás ideje. Hogy ez hatékonyan megvalósulhasson, az olvasási fluencia fejlesztése még ötödik osztályban is fontos iskolai feladat lenne, ehhez azonban a fluencia jellemzőinek mérése, feltérképezése az adott korosztályban, életkorokban szükséges előfeltétele lenne a tantervi és egyéb standardok kidolgozásában, a fejlesztés szükségleteinek, sajátosságainak meghatározása érdekében.

Köszönetnyilvánítás

A kutatást a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal NKFIH-K-120234 számú pályázata, az ET Research, a Forma Display és a Tématérületi Kiválósági Program támogatta.

Hivatkozások

Adamik, T. (2004). Az ókori retorika. In T. Adamik, A. A. Jászó, & P. Aczél (Eds.), *Retorika*. Budapest: Osiris Kiadó.

- Alt, S. J., & Samuels, S. Jay. (2011). Reading Fluency: What Is It and How Should It Be Measured? In A. McGill-Frazen, & R. L. Allington (Eds.), *Handbook of reading disability research* (p. 173–182). New York: Routledge.
- Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Beck, C. Th. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review*, 37, 18–37.
- Benczik, V. (2001). *Nyelv, írás, irodalom kommunikációelméleti megközelítésben*. Budapest: Trezor Kiadó.
- Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer (Version 5.0.1. URL: http://www.fon.hum.uva.nl/praat/download_win.html (letöltés ideje: 2013. január 10.)).
- Bóna, J., Imre, A., Markó, A., Váradi, V., & Gósy, M. (2014). GABI – Gyermeknyelvi beszédatadbázis és információtár. *Beszéd kutatás*, 22, 246–251.
- Bóna, J., & Steklács, J. (2020). A hangos olvasás hibajavításának mintázatai szemkamerás és akusztikai, fonetikai vizsgálatok tükrében. Egy 4. osztályosok körében végzett pilotvizsgálat tapasztalatai. *Anyanyelv-pedagógia*, 13, 17–29.
- Conklin, K., Pellicher-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge: Cambridge University Press.
- Duchowski, A. T. (2007). *Eye tracking methodology*. London: Springer.
- Gósy, M., Gyarmathy, D., Horváth, V., Grácsi, T. E., Beke, A., Neuberger, T., & Nikléczy, B. (2012). BEA: Beszélt nyelvi adatbázis. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* (p. 9–24). Budapest: Akadémiai Kiadó.
- Hudson, R. F., Lane, H. B., & Pullen, C. P. (2015). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, 58. (8), 702–714.
- Kamil, M. L., Pearson, P. D., Moje, E. B., & Afflerbach, P. P. (2011). *Handbook of reading research*, volume IV. New York: Routledge.

- Kang, E., & Y. Shin, M. (2019). The Contributions of Reading Fluency and Decoding to Reading Comprehension for Struggling Readers in Fourth Grade. *Reading & Writing Quarterly*, 35. (3), 179–192.
- Kuhn, M. R. (2011). Interventions to Enhance Fluency and Rate of Reading. In A. McGill-Frazen, & R. L. Allington (Eds.), *Handbook of reading disability research* (p. 307–315). New York: Routledge.
- Manguel, A. (2002). *Az olvasás története*. Budapest: Park Könyvkiadó.
- McGill-Franzen, A., & Allington, R. L. (2011). *Handbook of reading disability research*. New York, London: Routledge.
- Meixner, I. (2000). *A dyslexia prevenció és reedukáció módszere*. Budapest: Bárczi Gusztáv Gyógypedagógiai Tanárképző Főiskola.
- Pearson, P. D. (2009). The roots of reading comprehension instruction. In S. E. Israel, & G. G. Duffy (Eds.), *Handbook of Research on Reading Comprehension* (p. 3–31). New York: Taylor & Francis.
- Rasinski, T., Reutzel, R. D., Chard, D., & Linan-Thompson, S. (2011). Reading fluency. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research* (p. 286–319). New York: Routledge volume IV.
- Rayner, K. (Ed.) (1983). *Eye movement in reading. Perceptual and language processes*. New York: Academic Press.
- Rayner, K., Kennedy, A., & Radach, R. (2004). *Eye movements and information processing during reading*. New York: Psychology Press.
- Sipos, Zs. (2019). Egy, a 4. évfolyamos tanulók olvasásának vizsgálatára kidolgozott olvasólap formalizálásának folyamata és első eredményei. In E. K. Molnár, & K. Dancs (Eds.), *PÉK 2019 [CEA 2019] XVII. Pedagógiai Értékelési Konferencia [17th Conference on Educational Assessment]. Program és összefoglalók [Programme and Abstracts]*. Szeged. 87.

- Steklács, J. (2014). A szemmozgás vizsgálatának lehetőségei az olvasás és a vizuális információfeldolgozás képességének a megismerésében. *Anyanyelv-pedagógia*, 7. URL: <http://anyanyelv-pedagogia.hu/cikkek.php?id=524>.
- Steklács, J. (2017). Felolvasók és írásmagyarázók – az olvasás fogalmának értelmezése a bibliában. In Á. Fehér, & L. Mészáros (Eds.), „...et vocavit vocatione sua sancta”: *A pedagógiai hivatás a keresztény nevelésben* (p. 328–337). Vác: Apór Vilmos Katolikus Főiskola.
- Steklács, J. (2019). A szemkamerás vizsgálati módszer lehetőségei a pedagógiai szempontú kutatásokban. In J. Steklács (Ed.), *Szemkamerás vizsgálatok a pedagógiai kutatásban* (p. 5–24). Kaposvár: Kaposvári Egyetem Pedagógiai Kar.
- Steklács, J., & Bóna, J. (2020). A hangos olvasás jellemzőinek változása 4. és 5. osztályos korban – egy követéses pilotvizsgálat eredményei. In J. Bóna, & V. Krepsz (Eds.), *Nyelvfejlődés csecsemőkortól kamaszkorig* (p. 233–252). Budapest: ELTE Eötvös Kiadó.
- Vorstius, Ch., Radach, R., & Lonigan, Ch. J. (2014). Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition*, 22, 458–485.

A nemlexikális *öö* hang mint diskurzusjelölő magyar nyelvű társalgásokban

Németh Zsuzsanna^{1,2}

¹SZTE BTK Általános Nyelvészeti Tanszék

²MTA-DE-SZTE Elméleti Nyelvészeti Kutatócsoport, Eötvös Loránd Kutatási Hálózat (ELKH)

Abstract

This paper explores the possible role of the schwa-like sound *öö* in turn-taking in Hungarian conversations within a novel, conversation analytic and pragmatic framework. It is attempted to distinguish between a hesitation and a discourse organising function of the phenomenon, the latter related to the structural positions of *öö* in the conversation. With an inductive qualitative method all the occurrences of *öö* are examined in the corpus, and five characteristic structural positions are identified where by using *öö* speakers can influence the operation of turn-taking, or express their attitudes towards its rules. These positions can be at or before transition-relevance places. The analysis shows that *öö* can have a role in the application of all the three turn-taking rules (cf. Sacks et al., 1974). Discourse markers independent of the syntactic structure of the utterance can be non-lexical and may concern the sequential structure of the conversation (Fischer, 2006). This group of discourse markers involves the items that control the turn-taking system or other aspects of conversation management. Based on this, it is argued that the *öö*-s influencing turn-taking or express an attitude towards the rules of turn-taking can be regarded as discourse markers.

Keywords: *öö*, discourse marker, hesitation marker, turn-taking, conversation analysis

1. Bevezetés: A vizsgálat célja és elméleti háttere

A nemlexikális, svához hasonló *öö* hangot a magyarban széleskörűen kutatja, illetve kutatja a pszicholingvisztikai és fonetikai szakirodalom (l. pl. Gósy, 1993, 2006; Deme & Markó, 2013; Gósy et al., 2013; Horváth, 2014), és kitöltött szünetként a hezitációs jelenségek körébe sorolja (Gósy, 1993; Horváth, 2014). A hezitációs jelenségek átfogó pszicholingvisztikai vizsgálatát nyújtja Horváth (2014). A kitöltött szünet mind megjelenési formáját, mind használatát tekintve nyelvspecifikus jelenség (Crible et al., 2017), az angolban az *uh(m)*, a

Email address: nemethzs7@gmail.com (Németh Zsuzsanna)

franciában az *eah*, a spanyolban a *pues*, a japánban pedig az *eeto* sorolható ide (Crible et al. 2017). Fox (2010) az angol *uh(m)*-ot és a héber *e*-t nemlexikális, mindazonáltal konvencionalizálódott hangoknak tekinti, és a közbevetett hezitálók (*interjective hesitators*) csoportjába sorolja. Schegloff (2010) hangsúlyozza, hogy az *uh(m)* elemzése során nem elég a jelenség funkcióját figyelembe venni, ugyanannyira fontos az is, hogy a társalgás mely szerkezeti pozíciójában helyezkedik el. Például az *And uh(m)* / *But uh(m)* / *So uh(m)* + csend technika szolgálhat arra, hogy a beszélő egy korábbi sikertelen próbálkozás után lezárjon egy szekvenciát (koherens társalgási egységet), vagy jelezze, hogy diszpreferált válaszreakció következik utána (Schegloff 2010). Ezen kívül egy telefonhívás jellegzetes pozíciójában bevezetheti a hívás indoklását, illetve azt is jelezheti, hogy a hívás indoka kényes (Schegloff 2010; cf. Lerner 2013).

Kosmala & Crible (2021) megjegyzi, hogy a polifunkcionalitás tekintetében a kitöltött szünetek hasonlóak a diskurzusjelölőkhöz. A szerzőpáros a diskurzusjelölőket olyan kifejezéseként definiálja, amelyek a diskurzus szerkezetének és az interakciónak a kezelésére szolgálnak, és számos funkcióval rendelkezhetnek (pl. diskurzusrelációk jelzése vagy a beszélői attitűd kifejezése). Crible (2018) diskurzusjelölő-meghatározása szintén elsődlegesen funkcionális alapokon nyugszik, követve ezzel Schiffrin (1987) definícióját, mely szerint a diskurzusjelölők a beszélgetés egységeit összekapcsoló szekvenciálisan függő elemek.

A diskurzusjelölők meghatározásának kérdésében Diewald (2013) két iskolát említ. Az első a szintaktikailag függő elemeket, konnektívumokat tekinti diskurzusjelölőnek (pl. Fraser 1999). A második megközelítés már olyan elemeket is besorol a diskurzusjelölők kategóriájába, amelyek a szintaktikai szerkezettől függetlenek, és inkább a megnyilatkozás szerkezetéhez kapcsolódnak (Diewald 2006, 2013; Fischer 2006; Hansen 2006). Hansen (2006) szerint a diskurzusjelölők instrukciókkal látják el a hallgatót a tekintetben, hogy hogyan integrálja a DJ-t hordozó megnyilatkozást a diskurzus kialakuló mentális modelljébe olyan módon, hogy az optimálisan koherens megnyilatkozásként jelenjen meg. A második iskola esetében a diskurzusjelölő cím viselésének feltétele, hogy az elem irányítsa a diskurzust és olyan kommunikatív helyzetek alkotóelemeit kös-

se össze, amelyek nem propozicionálisak. Ez diskurzusszervező funkció, amely érintheti például a társalgás szekvenciális szerkezetét is. Ide tartoznak például a beszélőváltási rendszert szabályozó diskurzusjelölők. A második iskola által képviselt nézőpont szerint a diskurzusjelölők bármely pozícióban állhatnak, és lehetnek nemlexikális elemek is.

A beszéd folytonosságának jelenségét Götz (2013) a fluenszéma („fluence-me”) fogalmával ragadja meg. A fluenszéma a beszéd absztrakt tulajdonsága, mely hozzájárul ahhoz, hogy azt folytonosnak érzékeljük és produkáljuk (Götz 2013). Crible (2018) a fluenszémának mint a beszéd absztrakt tulajdonságának megvalósulásait fluenszémáknak nevezi, és olyan diszkrét eszközökként definiálja őket, amelyek a beszédprodukció és -percepció kognitív folyamatainak jelzéseként funkcionálnak (Crible 2018). A fluenszémának számos megvalósulása lehet, Crible et al. és munkatársai (2016) tipológiája szerint a kitöltött szünetek és a diskurzusjelölők a fluenszémák között két külön csoportot alkotnak. E szerint a felfogás szerint tehát a kitöltött szünetek és a diskurzusjelölők nem megakadásjelenségek, hanem éppen a beszéd folytonosságához járulnak hozzá, azaz ahhoz, hogy az percepció szempontból jelöletlen legyen (Crible 2018). Mindkét jelenségcsoport a beszéd produkcióját és feldolgozását segíti (Crible et al. 2017). Crible (2018) kiemeli, hogy a többi fluenszémához képest a diskurzusjelölők magas információs értékkel bírnak. Ez a felfogás tehát a kitöltött szüneteket és a diskurzusjelölőket egymástól külön, diszkrét kategóriákként kezeli. A diskurzusjelölők a kitöltött szünetekhez hasonlóan nyelvspecifikus funkciókkal rendelkeznek (Crible et al. 2017). Crible (2018) meghatározása szerint a diskurzusjelölők a pragmatikai jelölők egy típusát alkotják. Ez a típus grammatikailag heterogén, szintaktikailag opcionális, valamint polifunkcionális. A DJ-k procedurális vezérlőként metadiskurzív szinten funkcionálnak. Szerepük az, hogy megszorítsák a hordozó egység értelmezését a folyamatban lévő diskurzus egy közösen felépített reprezentációjában. Az értelmezésnek ez a megszorítása kétféle módon történhet. Az egyik, hogy a DJ-k diskurzusrelációt jeleznek a hordozó egység és a kontextusa között explicitté téve a diskurzusszegmentumok szerkezeti elrendezését, valamint kifejezve a beszélőnek a saját megfogalmazásá-

hoz fűződő meta-kommentárját. Ezen kívül a diskurzusjelölők a hordozó egység értelmezését úgy is megszoríthatják, hogy a beszélő és a hallgató viszonyához járulnak hozzá (Crible (2018)). Crible (2018) szerint a diskurzusjelölőknek döntő szerepük van a beszédtervezési folyamatokban.

Jelen tanulmányban a diskurzusjelölő fogalmának definiálásában a Diewald (2013) munkájában bemutatott második iskola diskurzusjelölő-értelmezését követem. Ennek az az oka, hogy Crible (2018) a kitöltött szüneteket a diskurzusjelölőkkel egy szinten lévő diszkrét kategóriának tekinti, így ez a taxonómia nem engedi meg, hogy egy kitöltött szünetet adott kontextusban diskurzusjelölőnek tekintsünk.

Jelen kutatásban nem használom a kitöltött szünet fogalmát. Ezt a döntést az indokolja, hogy a diskurzusjelölők meghatározásaival ellentétben a kitöltött szünet terminus nem funkcionális alapon nevezi meg a vizsgált jelenséget, hanem bizonyos nemlexikális formákhoz rendeli a *kitöltött szünet* címkét (ilyen pl. az angolban az *uh(m)* vagy a magyarban az *öö*). Ez zavart okozhat akkor, amikor a kitöltött szünet funkcióit elemezzük, hiszen minden előfordulása rendelkezik valamilyen funkcióval, amely kellő mennyiségű információ birtokában meghatározható, és amely alapján besorolható egy funkcionális kategóriába. Az is előfordulhat, hogy egy kitöltött szünet egyszerre több funkcionális alapon meghatározott kategóriába is besorolható lesz (pl. hezitációs jelölő és/vagy diskurzusjelölő). Ezért az elemzések során az *öö*-t mindig meg fogom nevezni.

A hezitáció fogalmát a jelen tanulmányban a konverzációelemzés elméleti keretét felhasználva a következőképpen definiálom. Hezitáció a fordulókonstruációs egység (a beszélő által aktuálisan végrehajtott cselekvés) előrehaladásának felfüggesztése. Ha a felfüggesztést a beszélő nem szünettel éri el, akkor a felfüggesztéshez használt elemet hezitációs jelölőnek tekintem. A hezitációs jelölők és a diskurzusjelölők viszonyát illetően többféle nézet alakult ki a szakirodalomban. Davis & Maclagan (2010) például a hezitációs jelölőket a diskurzusjelölők egy alcsoportjának tekinti. E mögött a nézet mögött az a háttérfeltevés áll, hogy a hezitáció diskurzussszervező funkcióval bír. Fox (2010) szerint vannak diskurzusjelölők, amelyek rendelkezhetnek hezitáló funkcióval, ilyen például a

y'know vagy a *like* az angolban. Ugyanezt a megfigyelést tette Schirm (2011) (2021) a *hát* és a *szóval* diskurzusjelölőről a magyarban. Podlesskaya (2010) az ún. *placeholder* jelenségét vizsgálva arra a következtetésre jut, hogy ezek a helyettesítő elemek, amelyeket a spontán diskurzusokban egy elhalasztott konszituens helyettesítésére használnak, nemcsak a diskurzusjelölők, hanem a hezitációs jelölők egy típusát is alkotják. Hayashi és Yoon (2010) felfogása szerint pedig ugyanazt a jelenséget lehet diskurzusjelölőként és hezitációs jelölőként is használni, a jelenség diskurzuskontextusban betöltött funkciójától függ, hogy mikor melyiknek tekinthetjük. A szerzőpáros megfigyelése szerint ugyanazok a mutató névmások (az *ano* a japánban, a *na-ge/nei-ge* a mandarinban és a *ce* a koreaiiban) különböző kontextusokban használhatók hezitációs jelölőkként és diskurzusjelölőkként is: hezitációs jelölőknek tekinthetők akkor, ha halasztó funkcióval rendelkeznek, és diskurzusjelölőnek akkor, ha kezdő társalgási cselekvést vezetnek be.

Jelen munkában a hezitációt nem tekintem diskurzusszervező (és ezáltal diskurzusjelölő) funkciónak. Ebből az következik, hogy a hezitációs jelölőket nem tekintem diskurzusjelölőnek, csak akkor, ha befolyásolják a társalgás szekvenciális (a fordulók egymásutánját érintő) szerkezetét, vagyis a fordulónak a résztvevők közötti megoszlását, illetve a beszélőváltás szabályaihoz való attitűdöt fejeznek ki. Következésképpen a hezitációs jelölő funkciót és a diskurzusjelölő funkciót külön kezelem, egy jelenség lehet DJ, HJ, de egyidejűleg mindkettő is. Ebből pedig az következik, hogy az *őő* minden egyes előfordulása egyedi vizsgálatot igényel. Kutatásom hipotézise, hogy az *őő* részt vehet a beszélőváltás szabályozásában, és ilyenkor diskurzusszervező funkcióval rendelkezik. Így ezek az előfordulásai értelmezhetők diskurzusjelölőként. A jelen tanulmányban az *őő* beszélőváltásban betöltött lehetséges szerepét veszem górcső alá magyar nyelvű társalgásokban.

A konverzációelemzés (társalgáselemzés) a társas interakciót annak természetes szerveződésében tanulmányozza, természetes, verbális interakciókat vizsgál hang- és videófelvételek segítségével (Schegloff, 1996; Mondada, 2013). A beszélőváltás normatív szabályait Sacks, Schegloff és Jefferson (1974) tárta fel és

írta le. Eszerint a beszélőváltást a résztvevők maguk irányítják. A váltásreleváns helyek a forduló (azaz az egy résztvevő által egyszerre elmondott egység) azon pontjai, ahol a résztvevők egy potenciálisan lezárható cselekvést ismernek fel (Clayman, 2013), ezért ezek a beszélőváltásra alkalmas helyek a társalgásban. Egy potenciálisan lezárható cselekvés egy fordulókonstrukciós egységet alkot. Ezek az egységek a forduló építőkövei (Schegloff, 2007). Minden beszélőváltásra alkalmas helyen ugyanaz a szabálysor lép életbe ugyanabban a sorrendben. Ha a beszélő kiválasztja a következő beszélőt, akkor a jelenlegi beszélőnek be kell fejeznie a beszédet, a kijelölt résztvevőnek pedig meg kell szólalnia (1a szabály). Ha a beszélő nem választja ki a következő beszélőt, akkor bármely más résztvevő átveheti a szót; az első megszólaló szerzi meg a jogot a következő fordulóhoz (1b szabály). Ha a beszélő nem választja ki a következő beszélőt, és más résztvevő sem szólal meg a b) lehetőség szerint, akkor az eredeti beszélő folytathatja (de nem köteles folytatni) a fordulót, azaz ő szerzi meg a jogot a forduló egy további szerkezeti egységéhez (1c szabály). Nagyon fontos tehát, hogy a váltásreleváns helyeken nem feltétlenül következik be beszélőváltás, az adott fordulókonstrukciós egység a váltásreleváns helyeken túlra is kiterjeszthető, ill. új cselekvés(ek) (új fordulókonstrukciós egységek) is illeszthetők hozzá még ugyanabban a fordulóban, amennyiben a váltásreleváns helyen az 1c szabály lép életbe. Ezek a szabályok minden soron következő váltásreleváns helyen ugyanebben a sorrendben lefutnak, azaz működésük rekurzív. Vagyis, ha az 1a vagy az 1b szabály szerint mégis beszélőváltás történik, akkor a soron következő forduló első váltásreleváns helyén futnak le újra.

A szakirodalomban mind a kitöltött szünetek, mind a diskurzusjelölők beszélőváltásban betöltött lehetséges szerepét kimutatták. Bortfeld és munkatársai (2001) például a tervezési nehézségek és a megakadásjelenségek (pl. újraindítások, ismétlések, kitöltött szünetek) kapcsolatát vizsgálták olyan beszélgetésekben, amikor ismerős, illetve ismeretlen témákról beszélgetnek a társalgás résztvevői. Úgy találták, hogy a kitöltött szünetek másképp viselkedtek, mint a legtöbb megakadásjelenség, ugyanis nem az ismeretlen, hanem az ismert témáknál voltak gyakoribbak. Ebből arra következtettek, hogy a fordulóeleji kitöltött

szünetek nem problémára utalnak, hanem interperszonális funkcióval rendelkeznek: a forduló átvételének szándékát jelezhetik.

Degand és van Bergen (2018) a beszélőváltási mechanizmusokat tanulmányozta spontán, a résztvevők személyes jelenlétével zajló társalgásokban és valós idejű, nem megtervezett, számítógépen keresztül történő szöveges üzenetváltásokban. Úgy találták, hogy mindkét interakció-típusban a megnyilatkozások végén elhelyezkedő diskurzusjelölők a beszédjog átengedését jelzik.

Dér (2012) a diskurzusjelölők beszélőváltásban betöltött lehetséges funkcióit elemezte a magyarban. Eredményei szerint a magyar társalgásokban jellemzőbb a diskurzusjelölők használata a beszédjog megszerzésekor (azaz forduló elején), mint fordulóvégi pozícióban, ahol viszont a beszédjog átadásának szándékát jelezheti. Fontos azonban, hogy a vizsgálat eredményei szerint egyik fordulóeleji diskurzusjelölőnek sem csupán a fordulóátvétel jelzése volt a funkciója, hanem más szerepeket is betöltöttek, például a beszélgetés korábbi részéhez vagy a tágabb kontextushoz kapcsolták a fordulót, vagy beszélői attitűdöt fejeztek ki. A leggyakoribb diskurzusjelölő ebben a pozícióban a hát volt.

Németh (2020) az *őő* társalgásban betöltött funkcióit a társalgáselemzés elméleti keretében kísérelte meg leírni. Arra a következtetésre jutott, hogy az *őő* hang a társalgásokban gyakran halasztó, időnyerő funkcióval rendelkezik. Használatával a beszélők szavakat, frázisokat, de leggyakrabban tagmondatnyi egységeket halasztanak el (gyakori például a *hogyőő* technika), és az *őő* használata a fordulóalkotás természetes részét képezi. Németh (2020) az *őő* néhány, beszélőváltásban betöltött lehetséges szerepét is áttekintette. A jelen tanulmányban ezeket a funkciókat kibővíttem, és a konverzációelemzés keretét pragmatikai kerettel kiegészítve megvizsgálom, hogy ezek a szerepek tekinthetők-e diskurzusszervező funkcióknak, és így az adott funkciót betöltő *őő* tekinthető-e diskurzusjelölőnek.

A tanulmány felépítése a következő: A 2. részben bemutatom a kutatás korpuszát és módszerét. Ezután következik az eredmények ismertetése a 3. részben. Először arra térek ki, hogy a hezitációs jelölők és a diskurzusjelölők megkülönböztetése milyen felismerésekhez vezetett az *őő* esetében, majd sor-

ra veszem azokat a szerkezeti pozíciókat, amelyekben az *őő* diskurzusszervező funkciót tölthet be. A kutatás eredményeit a 4. részben összegzem.

2. A vizsgálat korpusza és módszere

A kutatás korpusza magyar nyelvű, hétköznapi, nem előre megtervezett társalgásokat tartalmaz, melyeket hang- illetve videófelvételeken rögzítettek. A korpusz két részre osztható. Az egyik rész 9 db, egyenként kb. 20 perces, 3 résztvevős társalgást tartalmaz. Ezeket a Szegedi Tudományegyetem Pszichológiai Intézetének Noldus laboratóriumában rögzítették (SZTEPSZI-példák). A társalgások résztvevői egyetemisták voltak, akik azt az instrukciót kapták, hogy beszélgessenek bármiről 20 percig. A társalgás megkönnyítése érdekében egy-egy papírlapon kaptak három kérdést, melyek a következők voltak: (i) Egy fontos vizsgálja előtt a barátod megkér, hogy a kidolgozott tételekkel menj be az egyetem mosdójába. Azt tervezi, hogy miután kihúzta a tételt a vizsgán, kimegy a mosdóba, és becsempészi a szükséges tételeket. Segítenél-e neki? (ii) Szükség van-e még Magyarországon tehetségkutató műsorokra? (iii) Ha Magyarországon legalizálnák a könnyűdrogokat, vajon csökkenne-e az alkoholfogyasztás? A korpusz további 8, egyenként kb. 15 perces, szintén 3 fős beszélgetése a BEA beszélt nyelvi adatbázis részét képezi (Gósy, 2012). Ezek olyan társalgások, amelyek elindítása egy kérdéssel történt, amely meghatározta a beszélgetés témáját, de a beszélők egyik esetben sem tervezték meg előre, hogy mit fognak mondani. A jelenlegi vizsgálat szempontjából fontos, hogy a beszélőváltások (a BEA esetében a kezdeti kérdéstől eltekintve) spontán történtek, nem volt előre megszabva, hogy mikor és mennyit beszéljen egy-egy résztvevő. A résztvevők anonimitásának megőrzése érdekében a példák átírása kitalált neveket tartalmaz.

Noha az *őő* valamennyi előfordulását egyenként megvizsgáltam a korpuszban, kutatásom szigorúan kvalitatív jellegű, kvantitatív elemzést nem tartalmaz. Ennek két oka van. Egyrészt vannak a korpuszban az *őő*-nek olyan előfordulásai, amelyek „ellenállnak az elemzésnek” (vö. Schegloff, 2013), azaz a rendelkezésre

álló információink nem teszik lehetővé, hogy interakciós funkciójukat megállapítsuk. Ezen előfordulások számbavétele nélkül félrevezető lenne számadatokban megadni a különböző funkciókban megjelenő *őő*-ket. A tisztán kvalitatív analízist másrészt az indokolja, hogy a konverzációelemzés módszertani alapelve az előfordulásról előfordulásra történő induktív kvalitatív elemzés (Stivers & Sidnell, 2013). E megközelítés szerint csak ezzel a módszerrel írható le és magyarázható a társas interakció szerkezete. E mellett azzal érvelnek, hogy az induktív általánosításon túlmutató kvantitatív vizsgálatok torzíthatják az egyedi esetek észlelését, mert az előre meghatározott változók alapján a kutató „beleerőlteti” őket egy előre meghatározott kategóriába. A túl nagy tömegű adat pedig lehetetlenné teszi az egyedi előfordulásokra való újbóli és újbóli visszatérést, ami pedig szintén kulcsfontosságú a társalgáselemzésben (Schegloff, 2009). A következőkben ismertetem a kvalitatív vizsgálatból származó legfontosabb megfigyeléseimet.

3. A vizsgálat eredményei

3.1. Az *őő* hezitációs funkciójának és diskurzusjelölő funkciójának elkülönítése

A korpusz példáinak elemzése során úgy találtam, hogy az *őő* általában függeszti a folyamatban levő cselekvést, azaz legtöbbször rendelkezik hezitációs funkcióval. A szakirodalom eddigi eredményei szerint is ritka, hogy az *őő* része legyen a folyamatban levő cselekvésnek hezitáló funkció nélkül, de azért előfordul ilyen eset is. Lipóczy (2020) példájában (amely szintén az SZTEPSZI korpuszból származik), Fiú beszélő egy drog által befolyásolt személy hangját utánozza. Lipóczy szerint az imitáció során használt, elmélyített és hosszan tartott *őő* (Lipóczy átírásában *~ö:*) a drogos befolyásoltság által okozott lassabb felfogóképességet jelzi (Lipóczy, 2020).

- (1) Lipóczi, 2020:30 (A példák átírásában előforduló jelek magyarázatát lásd a Függelékben)

1 Fiú: *itt meg (bizonygatta) a rendőrnek hogy ~ö: nekem nincs lá:bam, nemtok=*
2 *=[menni ~° >be volt tép]=*
3 Lány2: [*hehhehehehe*]
4 Lány1: [*úristen.hehhehh*]
5 Fiú: *=ve mint az állat<*

Mivel néhány kivételtől eltekintve az *öö* legtöbbször felfüggeszti a folyamatban levő cselekvés előrehaladását, szinte mindig rendelkezik hezitációs funkcióval a korpuszban. Ez felveti a kérdést, hogy ténylegesen elválasztható-e a hezitációs funkciótól a diskurzusszervező funkció. Tanulmányomban amellet érvelek, hogy igen, hiszen azokban a példákban, amelyekben tulajdonítható az *öö*-nek diskurzusszervező funkció, ez sohasem a hezitáló funkciójából adódik. Miért? Mert nem a cselekvés felfüggesztésén keresztül szervezi az *öö* a diskurzust, hanem a pozíciója révén. Nézzük meg, hogy a társalgás lehetséges szerkezeti pozícióiban milyen diskurzusszervező funkciókkal rendelkezhet az *öö*! A pozíciókat a konverzációelemzés elméleti keretében fogom definiálni. Öt szerkezeti pozíciót különíték el. Az első a fordulókonstruktív egység belseje (váltásreleváns hely előtti pozíció). A következő három strukturális pozíció váltásreleváns helyre esik. Ezek a következők: a) forduló-kiterjesztés indítása, b) egy olyan forduló indítása, amely nem párszekvencia második párrésze, c) párszekvencia második párrésze indítása. A párszekvencia vagy szomszédsági pár (adjacency pair) két olyan cselekvésnek a sorozata, melyeket két különböző beszélő produkál, és amelyek közül az első elhangzása elvárttá és relevánssá teszi a második elhangzását. Az első a konverzációelemzés szakirodalma első párrésznek (first-pair part), a másodikat második párrésznek (second-pair part) nevezi (Schegloff 1968; Schegloff & Sacks, 1973; Sacks, 1992). Végül az *öö* ötödik pozíciójaként azokat az előfordulásait figyeljük meg, amikor a résztvevők egyszerre beszélése közben, azaz átfedésben levő beszédrészben jelenik meg.

3.2. Az *öö* strukturális pozíciókhoz köthető diskurzusszervező funkciói

3.2.1. Az *öö* előfordulása még a váltásreleváns hely előtt

A fordulókonstruációs egység belsejében előforduló *öö*-ről, mivel nem beszélőváltásra alkalmas ponton jelenik meg, azt feltételezhetnénk, hogy csupán hezitációs szerepe lehet a társalgásban. A beszélőváltási rendszerben azonban azt a szabályszerűséget is megfigyelték, hogy ha egy résztvevő már megszólalt és fordulót indított, akkor nemcsak joga, hanem kötelessége is legalább egy fordulókonstruációs egységet, azaz egy cselekvést produkálni (Schegloff 2007). Ebből belátható, hogy amennyiben valaki új cselekvésbe, új fordulókonstruációs egységbe kezd, vagy az előző cselekvést terjeszti ki, a folyamatban levő egységet potenciálisan lezárhatónak kell tennie. Ha a mondandójában társai még nem ismernek fel egy ilyen lehetségesen lezárható egységet, akkor a még folyamatban levő cselekvés felfüggesztéséről számot kell adnia. Ezt a számadást fejezheti ki az *öö*. Az *öö* használatával a beszélő biztosítja partnereit arról, hogy az ideiglenes felfüggesztés ellenére be szándékozik fejezni a cselekvést, eleget fog tenni kötelezettségének. Ez magyarázhatja, hogy az *öö*-nél gazdaságosabbnak tűnő csend helyett (vö. Németh 2021) miért az *öö*-t választjuk a halasztásra akkor, amikor új beszélő belépése nem „fenyeget” (a váltásreleváns hely még nem érkezett el). Németh (2021) a javítási műveletek preferenciahierarchiáját feltételezi. Ebből a hierarchiából következik, hogy a csend preferáltabb, mint a felfüggesztő funkciójú *öö*, mert a csend kevesebb szempontból sérti meg az előrehaladás preferenciáját: míg csend esetén a beszélő nem produkál a cselekvéshez nem tartozó elemet, a hezitáló funkciójú *öö* használatával igen, hiszen az nem része a cselekvésnek. Ezt az esetet illusztrálják a következő példák (2, 3, 4).

(2) (SZTEPSZI6, 11:00)

Johanna: *nem ilyen szempontból volt érdekes a sztori, hanem öö Mukival meg Csucsival utaztam és tömött volt a busz*

(3) (SZTEPSZI5, 01:06)

Mária: *nekem az a véleményem a puskázás kapcsán hogy **öö** (0.5) minden tanár kábé tudja hogyha a diák puskázik*

(4) (SZTEPSZI5, 17:32)

Martin: *jó hát láttam én ilyen embereket akik **ööö** az egyetemet is úgy csinálták végig .h minthogyha munkatáborban lennének*

A [2](#), [3](#) és [4](#) példában az *öö* tagmondatbevezető kötőszó után következik, azaz fordulókonstruációs egység belsejében használja őket a beszélő. Ezeken a helyeken a többi résztvevő számára egyértelmű, hogy a beszélő egy új tagmondatot kezdett, amely hozzátartozik a megkezdett fordulókonstruációs egységéhez. Ebben egy cselekvést felismerhetővé és lezárhatóvá kell tennie ahhoz, hogy beszélőváltásra alkalmas hely keletkezzen. Ezért aztán az *öö* használata a kötőszó után nem tűnik diskurzusszervező erejűnek, nem akadályoz meg például egy esetleges beszélőváltást. Ha azonban tekintetbe vesszük, hogy ezen a ponton már elvárt a beszélőtől a megkezdett egység befejezése, de láthatóan ehhez időre van szüksége, az *öö* használatát tekinthetjük egy ígéretnek a társak felé, miszerint „tisztában vagyok a rám vonatkozó elvárással, hogy megkezdett cselekvésemet lezárhatóvá tegyem, és ezt ki is kívánom elégíteni, de még időre van szükségem”. Ez pedig már tekinthető diskurzusjelölő funkciónak, mert a beszélőnek egy, a beszélőváltási rendszert érintő szabályhoz való attitűdjét fejezi ki. Ezeknek az *öö*-knek a diskurzusjelölőként való elemzését az is támogatja, hogy a beszélői attitűd kifejezését a diskurzusjelölők egy lehetséges funkciójaként tartja számon a szakirodalom (pl. [Furkó, 2011](#), [Schirm, 2011, 2021](#), [Kosmala & Crible, 2021](#)). Ez a funkció lehetséges magyarázatul szolgálhat arra, hogy miért nem választja a beszélő ezeken a helyeken az *öö* helyett a hezitáció más formáit. [Németh \(2020\)](#) megfigyelései arra utalnak, hogy közvetlenül az *öö* után gyakrabban fordul elő csend, mint *öö* nélkül (ld. [3](#) példa). Ez egybeesik [Sacks \(1992\)](#) és [Lerner \(2013\)](#) az angol *uh(m)*+szünet mintázatot érintő megfigyeléseivel. Eszerint a beszélő az *uh(m)*-mal jelzi, hogy annak ellenére, hogy most szünet következik, folytatni

fogja a megkezdett fordulókonstruációs egységet, vagyis az $uh(m)$ mintegy „legitimálja” az utána következő csendet. Horváth (2014) szintén úgy találta, hogy a magyarban is leggyakrabban a hezitálást követően fordul elő néma szünet a beszédben. A fentebb említett érv, miszerint a csend „produkciója” gazdaságosabbnak tűnik, mint egy testes elemé (vö. Németh, 2021), szintén az ilyen *öö*-k diskurzusjelölő funkciója mellett szól. Máskülönben miért ne tarthatna a beszélő *öö* nélküli szünetet egy olyan ponton, ahol még nincs váltásreleváns hely, azaz a beszélőváltási rendszer szerint új beszélő még nem léphet be?

A következő részben áttekintem azokat az eseteket, amikor az *öö*-t éppen a váltásreleváns helyen használják a beszélgetés résztvevői.

3.2.2. Az *öö* előfordulása beszélőváltásra alkalmas helyen

3.2.2.1. Forduló-kiterjesztés indítása *öö*-vel.

Az 5. példában három lány Juli közelgő esküvőjéről beszélget.

(5) (bea006n002, 06:16)

1. Juli: *há:t tesóm valószínűleg a tanúm lesz.*
2. Anna: *mhm.*
3. Juli: *ööö mer öö (0.7) így a családban is ennek volt előzménye*

Az 1. sor végén váltásreleváns hely van és Juli végintonációt alkalmaz, ezért arra következtethetünk, hogy Anna *mhm*-je inkább episztemikus állapotváltózást jelöl (vö. Heritage, 1984), mint continuer. Azaz Juli folytatásra való biztatása helyett inkább azt jelzi, hogy Anna a nemtudás állapotából a tudás állapotába került, vagyis „vette” az információt, amit Juli közölni akart vele. Ez azt jelenti, hogy Juli *ööö*-je a 3. sor elején olyan helyen fordul elő, ahol bármelyik másik résztvevő is elkezdhetne beszélni. Ezért funkciója lehet egyrészt annak a jelzése, hogy Juli az 1. sor cselekvését ki kívánja terjeszteni. Ha Anna *mhm*-jét a 2. sorban *continuer*-ként elemezzük, vagyis olyan elemként, amelynek használatával Anna arra bátorítja Julit, hogy folytassa a beszédet, akkor is

lehet Juli *ööö*-jének a 3. sor elején a diskurzus szerveződésében játszott szerepe. Juli jelezheti vele, hogy elfogadja a bátorítást és kész folytatni a beszédet, de nem tudja azonnal megkezdeni a kiterjesztés első lexikális elemét. Ezért itt az *öö* a beszélőváltásban játszott szerepe mellett a beszélő és a hallgató viszonyához is hozzájárulhat, így interperszonalitást kifejező diskurzusjelölőként is értelmezhető (vö. Crible, 2018).

A forduló-kiterjesztés egy olyan egység, amely nyelvtani szerkezetében függ az előtte levő fordulókstruktúrák egységtől, annak cselekvését terjeszti ki, viszont az előtte levő fordulókstruktúra egység már nélküle is lezárható lenne. A kiterjesztés tehát mindig váltásreleváns hely után következik (Schegloff, 1996; Lerner, 2004; Couper-Kuhlen & Tsuyoshi, 2007; Sidnell, 2012). Jelen esetben Juli a váltásreleváns helyen (1. sor vége) nem jelöl ki következő beszélőt. A kijelölés (1a) hiányában lefuthat az 1b szabály: bármelyik másik résztvevő kijelölheti magát következő beszélőnek. A példában ez sem következik be. Így a harmadik lehetőségként megjelenő 1c szabályban foglaltak lépnek életbe: az eredeti beszélő (jelen esetben Juli) folytathatja a fordulóját a váltásreleváns hely után. Juli él is ezzel a lehetőséggel, kiterjeszti a fordulóját. A kiterjesztési szándék *ööö*-vel való jelzése (3. sor eleje) biztosíthatja Juli számára, hogy ne léphessen be új beszélő, mielőtt kibővítené az első sorban megkezdett cselekvését. Ez a funkció befolyásolja a társalgás szekvenciális felépítését, vagyis a társalgást felépítő fordulólánc alakulását, hiszen befolyásolja a fordulók megoszlását a résztvevők között. Ezért mondhatjuk, hogy az *ööö* itt rendelkezik diskurzusjelölő funkcióval. A kötőszó+*öö*+szünet mintázat pontosan ebben a sorrendben gyakori halasztó technika a korpuszban (Németh, 2020). Az *ööö* után megjelenő *mer öö* (0.7) ezért arra utal, hogy már a 3. sor elején megjelenő *ööö* is rendelkezik hezitációs szereppel: diskurzus-szervező funkciója mellett szerepe lehet a kiterjesztés produkciójának elhalasztásában is.

A 6. példa egy olyan szekvenciából származik, amelyben a résztvevők arról beszélgetnek, hogy milyen sportokat űztek már életükben.

(6) (SZTEPSZI1, 628)

1. Ica: *Jazzbalettoztam*
2. Zsuzsi: *úúú*.
3. Ica: *ööö kilenc évig*.

Ica az 1. sorban látható fordulókonstruktív egységgel egy közlést hajt végre. A sportág megemlékezésének vége váltásreleváns hely, itt már felismerhető és lezárható Ica cselekvése. A lány záró intonációt alkalmaz, és nem jelöl ki következő beszélőt, így bármelyik másik résztvevő megszólalhat az 1b szabály szerint. Zsuzsi reagál a sportág megemlézésére (2. sor). Ezután Ica kiterjeszti a sportág megnevezésének cselekvését egy időhatározói bővítménnyel, mely nyelvtanilag kapcsolódik az 1. sorban látható igéhez, de annak nem kötelező vonzata. Így a hozzátoldás opcionális, az előzmények alapján nem várható. Ica *ööö*-jének (3. sor eleje) egyik funkciója nagy valószínűséggel időnyerés a jazzbalettel eltöltött idő kiszámítására. Ez azonban ismét nem magyarázza, hogy miért nem történik a halasztás csönddel. Mivel a hozzátoldás opcionális, az *ööö* produkciójával Ica jelezheti a forduló kiterjesztésének szándékát társai felé: „noha még időre van szükségem, hozzá szeretnék tenni valamit az általam eddig elmondottakhoz.” Azért is tűnik ez a jelzés indokoltnak, mert ezen a ponton még az sem látszik, hogy Zsuzsi szeretné-e folytatni a beszédet az *úúú* után, de az *ööö* hiányában a harmadik résztvevő is elkezdhetne beszélni. Ez az *ööö* is befolyásolja tehát, hogy hogyan alakul a beszédhez való jog a társalgásban. Produkciójával Ica jelzi, hogy új egységet indított, és egyúttal kötelezi magát olyan lexikális elem(ek) produkciójára, amelyekkel ez az új egység lezárhatóvá tehető. Ez diskurzuszervező funkció, ami arra utal, hogy a 3. sor elején található *ööö* nemcsak hezitációs jelölő, hanem diskurzusjelölő is. A [7], [8] [9] példákban szintén olyan *ööö*-ket figyelhetünk meg, amelyek forduló-kiterjesztést indítanak, azaz a beszélő használatukkal biztosítja, hogy a váltásreleváns hely ellenére senki ne vegye át tőle a szót, mielőtt egy újabb egységgel bővíthetné a fordulóját.

(7) (bea006n002, 00:54)

Juli: *persze a Cilit egyébként is meg[hívna(há)m, nemcsak azért mer(.)voltam =*
Cili: [((nevetés)) =
Juli: *= az esküvőjén]* **öö:** *de hogy (.) hogy igen és és ezér elég (.) nagy létszám=*
Cili: *=((nevetés))]*
Juli: *= jött ki*

(8) (bea007f005, 00:51)

Jani: *nyelvvizsgáim még nincsenek,*
Anett: ((nevetés))
Jani: **öö** *ami ugye megintcsak egy óriási hátrány*

(9) (bea006n002, 07:14)

Juli: *és az is jól áll öö: és és még az: tehát ezek a típusok vannak ver[senyben,*
Cili: [Mhm

A következő részben figyeljük meg azokat az *öö*-ket, amelyek olyan fordulót indítanak, amely nem párszekvencia második párrésze!

3.2.2.2. *Nem második párrészként funkcionáló forduló indítása öö-vel.*

Horváth (2014) úgy találta, hogy a beszédforduló elején megjelenő hezitálás beszédszándékot is jelezhet olyan esetekben, amikor a beszélő még a nyelvi tervezés fázisában van. Ha a beszélőváltási rendszer **Sacks et al. (1974)** szempontjából vizsgáljuk meg a fordulóeleji *öö*-ket, hasonló eredményre jutunk, sőt, a konverzációelemzés kerete lehetővé teszi, hogy a fordulóeleji *öö*-k diskurzus-szervező szerepét tovább differenciáljuk a társalgás szekvenciális kontextusát is figyelembe véve. Eszerint elkülöníthetjük egymástól azokat az *öö*-ket, amelyek egy párszekvencia második párrészenek elején állnak, és azokat, amelyek egy nem második párrészként funkcionáló fordulót indítanak. Nézzük először az utóbbi strukturális helyzetet! Ebben a pozícióban az *öö*-t egy olyan megszólaló

használja, aki a váltásreleváns helyen saját magát jelöli ki új beszélőnek (önkijelölés az 1b szabály alapján). Ez olyankor következhet be, ha az előző beszélő senkit nem jelölt ki a váltásreleváns helyig a beszélőváltás 1a szabálya alapján. A következő példában (10) Tündi, Feri és Kitti Feri korábbi kosárfogadásáról beszélget.

(10) (SZTEPSZI4, 02:55)

1. Tündi: *jól látom, hogy így kosár?*
2. Feri: *Mhm. De az bejött. Azt megraktam de az mos volt hajnalban.*
3. (0.5)
4. Feri: *Az bejött.*
5. (1.0)
6. Tündi: **Öö** *Lakers vagy [mire?]*
7. Feri: *[Mhm.]*

A 6. sorban Tündi a váltásreleváns helyen kijelöli magát következő beszélőnek (1b szabály), és új fordulót kezd *öö*-vel. Kvalitatív elemzésem szerint az *öö* itt több funkciót is betölt. Egyrészt elhalasztja az új fordulókonstruktív egység első lexikális elemét (HJ funkció). A halasztás oka lehet a kosárcsapat nevének keresése. A halasztó funkcióval azonban nem magyarázható, hogy Tündi miért választ hangzó elemet a lexikális elem produkciója előtt, hiszen megszólalása előtt egyébként is van egy 1 másodperces csönd, miért ne tarthatna addig ez a szünet, amíg el nem kezdi a *Lakers* szó produkcióját? Talán azért, mert a társalgásban mindez egy váltásreleváns helyen történik, és amint Tündiben felmerül a megszólalási szándék, ezt azonnal jeleznie kell társai felé, máskülönben lehet, hogy Kitti fog élni az önkijelölés lehetőségével (vagy Feri fűz új egységet a fordulójához). Az *öö* használatával tehát Tündi jelzi, hogy igényt tart egy új fordulóra, noha annak első lexikális elemét még nem tudja elkezdeni. Ahogyan egy korábbi példánál utaltunk rá, Schegloff (2007) megállapítja, hogy ha egy résztvevő megszólal egy társalgásban, akkor nemcsak joga, hanem kötelessége is legalább egy fordulókonstruktív egységet produkálni. Példánkban

ezért Tündi az *öö* használatával kötelezi magát legalább egy fordulókonstruációs egység produkciójára. Mivel ez a funkció érinti a fordulók megoszlását a beszélgetés résztvevői között, diskurzusszervező funkciónak tekinthető, ezért ez az *öö* diskurzusjelölő is. Ha használatának a beszélőváltást befolyásoló hatását nézzük, megállapíthatjuk, hogy Tündi az *öö*-vel Kitty esetleges önkijelölését és Feri forduló-kiterjesztésének lehetőségét akadályozza meg. Ezeknek az *öö*-eknek a DJ-ként való elemzését megerősíti Dér (2012) és Bortfeld et al. (2001) megfigyelése is. Míg Dér (2012) a fordulóeleji diskurzusjelölőknél mutatta ki a fordulóátvétel jelzésének funkcióját, Bortfeld és munkatársai (2001) ugyanezt a szerepet a forduló elején lévő kitöltött szünetek esetében figyelték meg.

A következő példában (11) Barbi, Heni és Letti a tizenévesekről beszélgetnek. Heni éppen azt meséli, hogy egy alkalommal, amikor a buszon utazott, egy tizenéves fiatal a saját társainak megjegyezte, hogy ha négy hely van fenntartva a kerekesszékesek számára, akkor miért nem utaznak kerekesszékesek a buszon. Heni a példa első részében a fiataalt idézi.

(11) (SZTEPSZI6, 11:40)

Heni: *Kerekesszék? Négy kerekesszéknyi hely van, de hát akkó hol vannak a kerekesszékek?*

(1.0)

Barbi: *Öö tehát az értelmiségi színvonalat próbálsz prezentálni.*

Heni kérdése nem számít párszekvencia első párrészenek, hiszen nem támaszt elvárást válasz produkciójára, csupán idézi a buszon elhangzott kérdést. Így Barbi fordulója nem második párrész, a lány megszólalása nem elvárt a társalgásban. Az *öö*-t Barbi fordulójának elején többféleképpen is értelmezhetjük. Ha csupán hezitációs funkciót tulajdonítunk neki, ismét nem tudunk számot adni arról, hogy az újonnan megszólaló miért töri meg az *öö*-vel a csendet, hiszen hezitációra a csend is alkalmas lenne. A forduló *öö* utáni része a Heni által végrehajtott cselekvést értelmezi. A konverzációelemzés szakirodalma javításkézdeményező formulaként tartja számon a megértés-ellenőrzést (Kitzinger

(2013). Ha Barbi fordulóját megértés-ellenőrzésként elemezzük, azaz úgy, hogy a lány explicitté teszi, és ezzel ellenőrzi, hogy helyesen értette-e, hogy Heni mit tett az előző fordulóban, akkor az *öö* lehet annak jelzése, hogy a javításkezdeményezés miatt Barbi megszakítja a folyamatban levő szekvencia (cselekvéslánc) folytonosságát. Az *öö* használatával a váltásreleváns helyen ugyanakkor azt is jelzi Barbi, hogy új fordulót kezdett, így blokkolja más résztvevők esetleges megszólalását, ezzel befolyásolva a beszédhez való jogok alakulását a társalgásban. Ezért ez az *öö* is tekinthető diskurzussszervező erővel bíró diskurzusjelölőnek. Ha még tovább vizsgálódunk, egy további elemzési lehetőség is felmerül. Barbi *öö*-je hasonlít azokhoz a fordulókonstrukciós halasztó technikákhoz, amelyeket Lerner (2013) figyelt meg az angolban (pl. *uh*) ún. kényes kifejezések, cselekvések produkciója előtt. Többek között ilyen cselekvésnek tekinti Lerner a kritika megfogalmazását. Ha figyelembe vesszük, hogy példánkban Barbi erős kritikát fogalmaz meg a fiatalok irányába, az *öö*-jét a kritikus megjegyzés előrevetítésének is tekinthetjük. E szerint az elemzés szerint a halasztás arra utal, hogy Barbi vonakodva mondja ki saját kritikus megjegyzését, a vonakodva kimondással pedig azt fejezi ki, hogy interperszonális szempontból kényesnek tartja a megjegyzést. Ez az interperszonális funkció szintén DJ-funkció, a beszélő és a hallgató viszonyában játszik szerepet (Crible 2018).

Amint arról már korábban esett szó, a szakirodalomban több olyan elemzéssel találkozunk, amely arról számol be, hogy a fordulóindító kitöltött szünetek, illetve diskurzusjelölők funkciója lehet a szóátvételi szándék jelzése (Bortfeld et al., 2001; Dér, 2012; Horváth, 2014). Fontos azonban, hogy ez a szerep csak olyan elemekre lehet érvényes, amelyek nem második párrész elején állnak, hiszen a második párrész produkálása elvárt az arra kijelölt résztvevőtől, annak nem szükséges a beszédsszándékát külön jeleznie.

A következő részben olyan *öö*-ket figyelünk meg, amelyek szintén váltásreleváns helyen jelennek meg, ám párszekvencia első párrésze után, azaz olyan fordulót indítanak velük a beszélők, amelyek egy párszekvencia elvárt második párrészeként funkcionálnak.

3.2.2.3. Párszekvencia második párrészenek indítása öö-vel.

Fontos különbség a párszekvencia első és második párrésze között, hogy míg egy első párrész produkciója általában opcionális a résztvevők számára, addig egy első párrész elhangzása után elvárt, hogy arra egy megfelelő (típusában az első párrész cselekvéséhez igazodó) második párrész érkezzen (Schegloff 1968; Schegloff & Sacks, 1973; Sacks, 1992). Ez azt jelenti, hogy az első esetben a megszólalóra nem nehezedik semmilyen elvárás, megszólalása önkéntes, és csak a megszólalás pillanata után köteles addig beszélni, amíg részéről a társai felismernek egy lezárható cselekvést (vö. Schegloff, 2007). A második esetben viszont (amikor már elhangzott egy első párrész), egy típuskonform válaszreakciónak érkeznie kell, annak produkciójára általában az első párrészt produkáló beszélő kijelöl egy résztvevőt, és ez az elvárás már a kijelölt résztvevő megszólalása előtt létrejön. Ez az interakciós „nyomás” kiválthatja egy olyan elem használatát a kijelölt résztvevő részéről, amely jelzi, hogy noha nem tudja válaszcselekvését azonnal elkezdeni, szándékában áll eleget tenni az őt érintő elvárásnak. A 12 példában Anett egy kérdést intéz Julihoz, aki válaszát öö-vel indítja.

(12) (bea006n002, 06:35)

1. Anett: *És a ruhád milyen lesz?*
2. Juli: **Ö** [*ö ö ö*]
3. Anett: [*Nehogy azt*] *mondd hogy fehér! ((nevetés))*
4. Juli: *Dehát fe[hér lesz, igen!*
5. Anett: [*((nevetés)) Igen ((nevetés))*
6. Juli: **Öö** *Nézegettem, két típus áll nekem igazán jól,*

Az 1. sorban Anett egy párszekvencia első párrésztét produkálja, és kijelöli Julit következő beszélőnek (1a szabály). Julinak ezért az első váltásreleváns helyen meg kell szólalnia: egy típuskonform 2. párrészt kell produkálnia (Sacks 1992). A 2. sor elején látható ööö-vel jelezheti, hogy noha még nem kezdi el a forduló cselekvését, tisztában van a rá vonatkozó normatív elvárással, és ki

kívánja azt elégíteni. Az *öööö*-nek ez a használata azért tekinthető diskurzusjelölőnek, mert Julinak a beszélőváltás 1a szabályához való viszonyulását, attitűdjét fejezi ki (vö. [Furkó, 2011](#), [Schirm, 2011, 2021](#), [Kosmala & Crible, 2021](#)). Mivel az elem egyúttal el is halasztja a válaszcselekvést, így hezitációs funkciója is van. Így ez az előfordulás is tekinthető hezitációs jelölőnek és diskurzusjelölőnek egyszerre. Az *öö* elnyújtását nagy valószínűséggel Anett közbevetett vice váltja ki (*Nehogy azt mondd hogy fehér!*). Amikor Juli a 6. sor elején újrakezdi a választ, szintén *öö*-vel indítja a cselekvést. Azt az elvárást tehát, amit az 1. sor kérdése támaszt számára, csak a 6. sorban elégíti ki, és a késői reakciót szintén *öö*-vel indítja. [Németh \(2020\)](#) megfigyelte azt a jelenséget, amikor az *öö* egy válaszjelölő *hát* diskurzusjelölőhöz kapcsolódik (vö. [13](#) példa; a *hát* válaszjelölő funkciójáról l. pl. [Kiefer, 1988](#); [Németh T., 1998](#); [Schirm, 2011, 2021](#); [Markó & Dér, 2011](#); [Kondacs, 2016](#)). Ez összhangban van [Dér \(2012\)](#) eredményeivel, aki úgy találta, hogy fordulóeleji pozícióban a leggyakoribb diskurzusjelölő a *hát*.

(13) (SZTEPSZI5, 00:52)

1. Dániel: *puskázás.*
2. (0.7)
3. Zsuzsi: *mmm*
4. (3.0)
5. Zsuzsi: *mit teszünk?*
6. (1.1)
7. Martin: ***hát öö*** (0.3) ()
8. Zsuzsi: *hát én mmm (0.5) szerintem tutira segítenék neki.*

Horváth (2014, p. 36) példái között szintén megtalálható a jelenség, amikor egy kérdésre adott hosszabb választ kezd így a beszélő: „*hát öö én úgy általában nem nagyon szoktam inni*”. [Schirm \(2018\)](#) a diskurzusjelölő-társulásokat elemezte a Szögedi Szociolingvisztikai Interjúban. A kilenc leggyakoribb társulásból hétben az egyik DJ a *hát* volt. A *hát öö* olyan esetekben is előfordult a korpuszomban,

amikor egy késői választ vezetett be, azaz a válasz nem közvetlenül az első párrész után következett ([14] l. [Németh, 2020]).

(14) (SZTEPSZI1, 01:52)

1. Kati: *és más?*
2. (0.3)
3. Kati: *mást sportoltatok?*
4. Linda: *ige:n, há a Lil[lával kip-]*
5. Kati: [*te vol*|*tál akkor [izé sportna*|*pon?*
6. Boglárka: [*nem voltam.*]
7. Kati: *nem.*
8. Linda: [()]
9. Boglárka: [*tavaly se idén se.*]
10. Linda: *igen? tavaly se voltál?*
11. (0.7)
12. Linda: *mondjuk én tavaly nem voltam. úgyhogy viszonyítási alpom*
13. *nin>csen de< ÉN nagyon élvez>tem. sze<rintem nagyon jó volt*
14. *.h hát öö eleinte pingpongoztunk, .h utána Lillával ö szumóztunk*

Ilyenkor a *hát öö* esetleges válaszjelölő funkciója még hangsúlyosabb, hiszen a beszélőnek jeleznie kell a társak felé, hogy a korábbi kérdésre fog válaszolni. A válaszjelölő diskurzusjelölő használatát még inkább indokolja ebben a pozícióban, hogy a kijelölt beszélő már egy régebbi „adósságát” törleszti, hiszen a válaszra vonatkozó normatív elvárást preferált lett volna minél korábban kielégíteni (vö. [Sacks, 1987]). Lehetséges, de természetesen még további megerősítő példákat igényel az a hipotézis, hogy a válaszjelölő *hát*-hoz kapcsolódó *öö* a nyelvhasználatban leválhat a *hát*-ról, és önállóan is betölthet válaszjelölő funkciót, mint a [9] példa késői válaszának indításakor a 6. sor elején (amennyiben ott az *öö*-t válaszjelölő diskurzusjelölőnek elemezzük). Viszont ha minden második párrészt bevezető *öö*-t úgy tekintünk, hogy az – az esetleges egyéb funkció(i) mellett – a beszélő elköteleződését és ígéretét is kifejezi a válaszra

vonatkozó elvárás teljesítésére, akkor ez összhangban van azzal, hogy ezeket az *öö*-ket a válaszjelölő *hát*-hoz hasonlóan válaszjelölő diskurzusjelölőnek tekintjük, az ebben a pozícióban megjelenő *hátöö*-t pedig diskurzusjelölő-társulásnak. Itt ismét felhozható az az érv, hogy ha ezeknek az elemeknek csupán hezitációs funkciója van, akkor miért nem választja a beszélő a csendet a hezitálásra. Az is lehetséges, hogy mivel mind a *hát*, mind az *öö* betölthet hezitáló funkciót is, a válaszjelölő *hát*-ok és *öö*-k azt fejezik ki, hogy „noha még időre van szükségem hozzá, de tisztában vagyok vele és ki is kívánom elégíteni azt a rám vonatkozó elvárást, hogy egy válaszcselekvést produkáljak”. Ez a feltevés természetesen további vizsgálatokat igényel.

További érdekes, és a beszélőváltási rendszer szempontjából különösen fontos jelenség az egyszerre beszélés, amelyben érdemes az előforduló *öö*-k lehetséges szerepét feltárnunk.

3.2.3. Az *öö* lehetséges átfedéskezelő szerepe

Figyeljük meg a [15](#) példában található egyszerre beszélést!

(15) (SZTEPSZI8, 11:46)

Gabi: *Igen tehát hogy az szerintem nem* (.) [*öööööööö*] (.) *csökken*[*tené*] *hanem*
 Viki: [*Hát azöö*] [*Mhm*]

Gabi még a váltásreleváns hely előtt megáll, és egy rövid szünet (.) következik. Meglepő, hogy annak ellenére, hogy nincs váltásreleváns hely, Viki megszólal. Ezt valószínűleg Gabi megállása váltja ki. Érdekes megfigyelni, hogy hiába nincs még váltásreleváns hely, ha a hezitáció hang nélküli formáját választja a beszélő, egy másik résztvevő megszólalhat. Ez alátámasztja a váltásreleváns hely előtti *öö*-k korábbi elemzését, miszerint ezeknek is szerepük lehet a beszélőváltásban: kifejezik, hogy a beszélőnek szándékában áll a megkezdett, ám még nem lezárható cselekvését folytatni. Hogyan elemezhetjük Gabi hosszan elnyújtott *öö*-jét? Mondhatjuk, hogy a lány azt kívánja jelezni vele, hogy még nem akarja átadni a szót, folytatni akarja a megkezdett fordulókonstrukciós egységet.

Ezt Viki is így értelmezi, mert visszalép a megszólalásától, és hagyja, hogy Gabi folytassa a megkezdett egységet. Gabi *öö*-jének átfedéskezelő funkcióját az is alátámasztja, hogy pontosan addig tart, amíg Viki vissza nem lép. Az *öö* hosszú elnyújtása arra is szolgálhat, hogy elhalassza a tartalmas beszédrész produkcióját az átfedés utánra, vagyis akkorra, amikor már az egyszerre beszélés nem veszélyezteti a megértést (vö. Schegloff, 2013). Erre utal az átfedés utáni szünet is: Gabi csak akkor szólal meg újra, miután meggyőződött arról, hogy Viki visszalépett, és ő „nyert” a beszédjogért való versengésben. Gabi *öö*-je a beszélőváltást szabályozza, így diskurzusszervező funkciója van. Mivel megakadályozza a cselekvés továbbhaladását egy lehetséges lezárás felé, hezitációs funkcióval is rendelkezik. Viki *öö*-je az átfedés vége felé abból fakadhat, hogy észleli Gabi *öö*-jét, és nem folytatja addig a beszédet, amíg az átfedés véget nem ér, de az *öö*-vel rövid ideig még „küzd” a beszédjogért: fenntartja szóátvételi szándékát, végül pedig visszalép. A 16 példában a beszédjogért való versengés kimene-tele fordított: az újonnan megszólaló résztvevő, Anita szerzi meg a következő fordulóhoz való jogot úgy, hogy Bea fordulókonstrukciós egységének végét már nem lehet érteni, a lány csak maga elé „dünnögi” azt. Érdekes, hogy Anita úgy veszi át a szót, hogy még csak nem is váltásreleváns helyen szólal meg. Ezt a beszélgetésben észlelhető dominanciájával is magyarázhatjuk, a három résztvevő közül ő beszél a legtöbbet a társalgás folyamán.

(16) (SZTEPSZI7, 00:12)

Bea: *De a marihuánával akko egyátal[án ()]*

Anita: [*öööö*] *szerintem szerintem ezzel a () val kezdjük*

A 17 példa némileg eltér az előző kettőtől.

(17) (SZTEPSZI8, 15:06)

Viki: *de de több visszaélés lenne: (.) [ööö ez ez ez olyan hogy] .h hogy működik*

Sári: [*úgyis találának valamit*]

Sári megszólalását itt is valószínűleg a csönd váltja ki, ám az előző esetekkel ellentétben a csönd most egy váltásreleváns helyen áll elő. Sári megszólalását tehát most a beálló rövid szünet mellett az is indokolhatja, hogy szóátvétele „szabályos”, a beszélőváltási rendszer 1b szabályát, az önkijelölést alkalmazza (a szóátvétel „szabályossága” miatt lehet az, hogy most egy teljes fordulókonstruktív egységet produkál, nem lép vissza olyan hamar, mint Viki a [15] példában). Sári megszólalása miatt Viki az *öö*-vel valószínűleg jelezni kívánja, hogy még nem kész átadni a beszédjogot, további grammatikai egysége(ke)t szeretne fűzni a fordulókonstruktív egységéhez, azaz ki szeretné terjeszteni a fordulót. Viki nemcsak az *öö* elnyújtásának technikáját választja ahhoz, hogy mondandójának lényeges részét az átfedés utánra halassza, hanem az *ez* mutató névmás kétszeri megismétlését is. Ezután egy hallható levegővételt és a *hogy* kötőszó megismétlését követően, miután meggyőződött arról, hogy Sári nem folytatja a beszédet, folytatja a megkezdett cselekvést. Az ismétlés ilyen átfedéskezelő funkcióját a szakirodalomban javítási funkcióként tartják számon, az ismétlést pedig javítási műveletként (vö. Schegloff, 2009, 2013). Ezt azzal indokolják, hogy az ilyen ismétlés az átfedés miatt fellépő esetleges észlelési problémákat kezeli. Ennek megfelelően az azonos funkciójú, átfedéskezelő *öö*-t is tekinthetjük javítási műveletnek. A példában látható *öö* olyan egyszerre beszélést kezel, amely a beszélőváltási rendszerben fellépett zavarra utal: a váltásreleváns helyen az eredeti beszélő is folytatni akarja a beszédet, és egy új beszélő is kijelöli magát, vagyis az 1c és 1b szabályt egyszerre alkalmazzák a résztvevők (vö. Hayashi, 2013).

Az átfedéskezelő *öö* használata tehát egyértelműen befolyásolja a beszélőváltást, így diskurzusjelölőnek is tekinthető, a cselekvést elhalasztó funkciója miatt pedig hezitációs jelölő is.

A következőkben összegzem megfigyeléseimet, amelyeket az *öö* beszélőváltásban betöltött szerepéről tettem.

4. Konklúzió: Az *őő* diskurzusszervező szerepe

Tanulmányomban kísérletet tettem arra, hogy a hagyományosan hezitációs jelölőnek tekintett, svához hasonló *őő* hang esetleges diskurzusszervező funkcióit kimutassam magyar nyelvű társalgásokban. Először elkülönítettem az *őő* hezitációs és diskurzusszervező funkcióját. E distinkció mellett azzal érveltem, hogy az *őő* diskurzusszervező funkciója nem a hezitációs funkciójából következik, hanem a társalgás szekvenciális szerkezetében betöltött pozíciójából. Az *őő*-t öt strukturális pozícióban figyeltem meg:

- (i) A váltásreleváns hely előtt az *őő* használatával a beszélő biztosíthatja partnereit arról, hogy az ideiglenes felfüggesztés ellenére eleget fog tenni kötelezettségének, miszerint a társalgásban minden megkezdett új egységet potenciálisan lezárhatóvá kell tenni. Az *őő* ilyen használatával a beszélő a beszélőváltási rendszer egy szabályához való viszonyulását fejezi ki.
- (ii) Az *őő*-t egy kijelölt új beszélő használja annak jelzésére, hogy tisztában van az őt érintő normatív elvárással, hogy egy második párrészt produkáljon, és ezt az elvárást ki is kívánja elégíteni (a beszélőváltás 1a szabályához való attitűd kifejezése).
- (iii) Egy önmagát kijelölő új beszélő az *őő* használatával jelzi, hogy jogot formál a következő fordulóra, és ezzel megakadályozza, hogy az eredeti beszélő folytassa a fordulót vagy más résztvevők lépjenek be (az *őő* egy technika az 1b szabály alkalmazásához).
- (iv) Az aktuális beszélő arra használja az *őő*-t, hogy a szóátvételt megakadályozza, és további egység(ek)et fűzhessen a fordulójához (az *őő* egy technika az 1c szabály alkalmazásához).
- (v) Az átfedésben megjelenő *őő* kifejezi, hogy a beszélő igényt tart egy további beszédrészre.

Ahogyan a bevezetőben említettük, az *őő* lehetséges hezitációs funkciója régóta ismeretes a szakirodalomban. Azt is láthattuk, hogy egyes megközelíté-

sek (pl. [Davis & Maclagan, 2010](#)) a hezitációs funkciót eleve diskurzusjelölő-funkciónak tartják. Ebben a keretben az *őő* legtöbb előfordulása – amikor a beszélő hezitálását jelzi – eleve diskurzusjelölőnek minősülne. Jelen vizsgálat azonban a hezitációs funkciót nem tekintette diskurzussszervező funkciónak, ezért ahhoz, hogy az *őő* DJ-címkét kaphasson, nem volt elég a hezitációs funkciót kimutatni. Az elemzések azonban megmutatták, hogy az *őő* mind az öt felsorolt strukturális helyzetben alkalmas arra, hogy befolyásolja a fordulók megoszlását a résztvevők között, szerepet játszva a beszélőváltási rendszer szabályos működésében. Ez indokoltá teszi, hogy ezekben a szerepekben az *őő*-t olyan diskurzusjelölőnek tekintsük, amely a megnyilatkozás szintaktikai szerkezetétől független, a kommunikációs helyzetek összekapcsolásában játszik szerepet, nemlexikális, és a társalgás szekvenciális szerkezetét érinti azáltal, hogy szabályozza a beszélőváltási rendszer működését (vö. [Fischer, 2006](#), [Diewald, 2013](#), [Dér, 2012](#)). A jelenség mindazonáltal további vizsgálatokat igényel. A kutatás további iránya lehet az *őő* funkcióinak különböző műfajú szövegekben való vizsgálata (vö. [Crible, 2018](#)).

Köszönetnyilvánítás

Köszönöm Dér Csillának, Furkó Péternek és Schirm Anitának a szakirodalom tekintetében nyújtott segítségüket, Nagy C. Katalinnak az értékes tanácsait, valamint a tanulmány anonim lektorainak az építő véleményüket. Hálás vagyok továbbá az SZTE Pszichológiai Intézetének, hogy a korpusz 9 beszélgetésének felvételét lehetővé tette, Lerch Ágnesnek, akivel a felvételeket elkészítettük, valamint Gósy Máriának és az MTA Kempelen Farkas Beszédkutató Laboratóriumának, hogy a BEA 8 beszélgetését rendelkezésemre bocsátotta.

Hivatkozások

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44, 123–147.

- Clayman, S. E. (2013). Turn-constructive units and the transition-relevance place. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (p. 150–166). Oxford: Wiley-Blackwell.
- Couper-Kuhlen, E., & Tsuyoshi, O. (2007). „Incrementing” in conversation: A comparison of practices in English, German and Japanese. *Pragmatics*, 17, 513–552.
- Crible, L. (2018). *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers*. Amsterdam, Philadelphia: 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.
- Crible, L., Dumont, A., Grosman, I., & Notarrigo, I. (2016). *Annotation Manual of Fluency and Disfluency Markers in Multilingual, Multimodal, Native and Learner Corpora. Version 2.0*. Technical report, Université Catholique de Louvain and Université de Namur.
- Davis, B. H., & Maclagan, M. (2010). Pauses, fillers, placeholders and formulaicity in Alzheimer’s discourse: Gluing relationships as impairment increases. In N. Amiridze, B. H. Davis, & M. Maclagan (Eds.), *Fillers, Pauses and Placeholders* (p. 189–215). Amsterdam/Philadelphia: John Benjamins.
- Degand, L., & Bergen, G. (2018). Discourse markers as turn-transition devices: Evidence from speech and instant messaging. *Discourse Processes*, 55, 47–71.
- Deme, A., & Markó, A. (2013). Lengthenings and filled pauses in Hungarian adults’ and children’s speech. In R. Eklund (Ed.), *Proceedings of DISS 2013. The 6th Workshop of Disfluency in Spontaneous Speech* (p. 21–24). Stockholm: KTH Royal Institute of Technology.
- Dér, Cs. I. (2012). Beszélőváltások során használt diskurzuszjelölők a magyar spontán beszédben. *Beszédkutató*, 21, 130–141.

- Diewald, G. (2006). Discourse particles and modal particles as grammatical elements. In K. Fischer (Ed.), *Approaches to Discourse Particles* (p. 403–425). Amsterdam: Elsevier.
- Diewald, G. (2013). „Same same but different” – Modal particles, discourse markers and the art (and purpose) of categorization. In L. Degand, B. Cornillie, & P. Pietrandrea (Eds.), *Discourse Markers and Modal Particles: Categorization and Description* (p. 19–46). Amsterdam: John Benjamins.
- Fischer, K. (2006). Towards an understanding of the spectrum of approaches to discourse particles: Introduction to the volume. In K. Fischer (Ed.), *Approaches to Discourse Particles* (p. 1–20). Amsterdam: Elsevier.
- Fox, B. (2010). Introduction. In N. Amiridze, B. H. Davis, & M. MacLagan (Eds.), *Fillers, Pauses and Placeholders* (p. 1–9). Amsterdam/Philadelphia: John Benjamins.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31, 931–952.
- Furkó, B. P. (2011). Diskurzusjelölők és szövegösszefüggés: a kohézió vagy a koherencia eszközei? – Az implikáció mint a szövegösszefüggés eszköze. In E. Dobi (Ed.), *A szövegösszefüggés elméleti és gyakorlati megközelítési módjai. Poliglott terminológiai és fogalmi áttekintés. (Petőfi S. János 80. születésnapjára). Officina Textologica 16* (p. 37–57). Debrecen: Debreceni Egyetemi Kiadó.
- Gósy, M. (1993). A lexikális hozzáférés: Szófelismerési stratégiák. *Beszédkutatás*, 2, 14–32.
- Gósy, M. (2006). A semleges magánhangzó nyelvi funkciói. *Beszédkutatás*, 15, 8–22.
- Gósy, M. (2012). Multifunkcionális beszélt nyelvi adatbázis – BEA. *Általános Nyelvészeti Tanulmányok XXIV. Nyelvtechnológiai Kutatások*, 24, 329–349.

- Gósy, M., Bóna, J., Beke, A., & Horváth, V. (2013). A kitöltött szünetek fonetikai sajátosságai az életkor függvényében. *Beszéd kutatás*, 22, 121–143.
- Götz, S. (2013). *Fluency in Native and Nonnative English Speech*. Amsterdam: John Benjamins.
- Hansen, M.-B. M. (2006). A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French *toujours*). In K. Fischer (Ed.), *Approaches to Discourse Particles* (p. 21–41). Amsterdam: Elsevier.
- Hayashi, M. (2013). Turn allocation and turn sharing. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (p. 167–190). Oxford: Wiley-Blackwell.
- Hayashi, M., & Yoon, K.-E. (2010). A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. In N. Amiridze, B. H. Davis, & M. MacLagan (Eds.), *Fillers, Pauses and Placeholders* (p. 33–66). Amsterdam/Philadelphia: John Benjamins.
- Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (p. 299–345). Cambridge: Cambridge University Press.
- Horváth, V. (2014). *Hezitációs jelenségek a magyar beszédben*. Budapest: ELTE Eötvös Kiadó.
- Kiefer, F. (1988). Modal particles as discourse markers in questions. *Acta Linguistica Hungarica*, 38, 107–125.
- Kitzinger, C. (2013). Repair. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (p. 229–256). Oxford: Wiley-Blackwell.

- Kondacs, F. (2016). A hát diskurzusjelölőről az óvodások diskurzusaiban. In T. Váradi (Ed.), *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2016* (p. 45–58). Budapest: MTA Nyelvtudományi Intézet.
- Kosmala, L., & Crible, L. (2021). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech, SAGE Publications (UK and US)*, 2021. doi:[10.1177/00238309211010862](https://doi.org/10.1177/00238309211010862).
- Lerner, G. (2013). On the place of hesitating in delicate formulations: a turn-constructural infrastructure for collaborative indiscretion. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational Repair and Human Understanding* (p. 95–134). Cambridge: Cambridge University Press.
- Lerner, G. H. (2004). On the place of linguistic resources in the organization of talk-in-interaction: grammar as action in prompting a speaker to elaborate. *Research on Language and Social Interaction*, 37, 151–184.
- Lipóczi, B. (2020). *Történetmesélés magyar nyelvű társalgásokban. Szakdolgozat*. Szeged: SZTE BTK Általános Nyelvészeti Tanszék.
- Markó, A., & Dér, Cs. I. (2011). A diskurzusjelölők használatának életkori sajátosságai. In J. Navracsics, & Zs. Lengyel (Eds.), *Lexikai folyamatok egy- és kétnyelvű közegben. Pszicholingvisztikai tanulmányok II* (p. 49–61). Budapest: Tinta Könyvkiadó.
- Mondada, L. (2013). The conversation analytic approach to data collection. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (p. 32–56). Oxford: Wiley-Blackwell.
- Németh, Zs. (2020). A nemlexikális öö hang interakciós szerepének elemzése magyar nyelvű társalgásokban. *Jelentés és Nyelvhasználat*, 7, 23–50.
- Németh, Zs. (2021). *A Preference Hierarchy Model of Same-Turn Repair Operations in Talk-in-Interaction*. Sheffield: Equinox.

- Németh T., E. (1998). A *hát, így, tehát, mert* kötőszók pragmatikai funkciójának vizsgálata. *Magyar Nyelv*, 94, 324–331.
- Podlesskaya, V. I. (2010). Parameters for typological variation of placeholders. In N. Amiridze, B. H. Davis, & M. MacLagan (Eds.), *Fillers, Pauses and Placeholders* (p. 11–32). Amsterdam/Philadelphia: John Benjamins.
- Sacks, H. (1987). On the preferences for agreement and contiguity in sequences in conversation. In G. Button, & J. R. E. Lee (Eds.), *Talk and Social Organisation* (p. 54–69). Clevedon, UK: Multilingual Matters.
- Sacks, H. (1992). *Lectures on Conversation*, 2 Vols. (Fall 1964 – Spring 1972). Oxford: Blackwell.
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735.
- Schegloff, E. (1968). Sequencing in conversational openings. *American Anthropologist*, 70, 1075–1095.
- Schegloff, E. (1996). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, 102, 161–216.
- Schegloff, E. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Schegloff, E. (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. (2010). Some other uh(m)s. *Discourse Processes*, 47, 130–174.
- Schegloff, E. (2013). Ten operations in self-initiated, same-turn repair. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational Repair and Human Understanding* (p. 41–70). Cambridge: Cambridge University Press.
- Schegloff, E., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8, 289–327.

- Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.
- Schirm, A. (2011). *A diskurzusjelölők funkciói: a hát, az -e és a vajon elemek története és jelenkori szinkrón státusa alapján*. Doktori disszertáció. Szeged: University of Szeged. URL: http://doktori.bibl.u-szeged.hu/759/1/schirm_anita_doktori_disszertacio.pdf.
- Schirm, A. (2018). Diskurzusjelölő-társulások a Szögedi Szociolingvisztikai Interjúban. *Alkalmazott Nyelvtudomány*, 18, 1–16.
- Schirm, A. (2021). *Diskurzusjelölők szövegeken innen és túl*. Budapest: Loisir.
- Sidnell, J. (2012). Turn-continuation by self and by other. *Discourse Processes*, 49, 314–337.
- Stivers, T., & Sidnell, J. (2013). Introduction. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (p. 1–8). Oxford: Wiley-Blackwell.

Függelék

A példák átírása során használt jelek és feloldásuk

- | | |
|---------|--|
| (.) | mikropauza; nagyon rövid szünet, körülbelül egy szótagnyi hosszúságú |
| (2.0) | megmért szünet (másodperc.tizedmásodperc) |
| [nem | a bal oldali zárójel az átfedés kezdetét jelzi |
| [persze | |
| nem] | a jobb oldali zárójel az átfedés végét jelzi |
| persze] | |
| = | Egyenlőségjel kapcsolja össze ugyanazon beszélő folyamatos megnyilatkozásának részeit, ha azokat a másik beszélő közbeszólása miatt az átírásban el kell szakítani egymástól; egyenlőségjel jelöli azt is, ha két megnyilatkozás között nincsen intervallum, de átfedés sincs. |

.h	lélegzetvétel
.hh	hangosabb, hosszabb lélegzetvétel
()	érthetetlen közlést jelöl
aha,	a vessző a szó végén emelkedő intonációt jelez
aha.	a pont a szó végén eső intonációt jelez, amely nem minden esetben esik egybe a megnyilatkozás végével
AHA	a nagybetűvel való kiemelés a hangerő növekedését jelzi
a:ha-	a kettőspont a magánhangzó megnyúlását jelzi, a kötőjel a megszakítás jele
((torokköszörülés))	a kettős zárójel a nemnyelvi hangjelenségeket jelzi
> <	a jel két eleme közé eső beszédszakasz tempója gyorsabb
< >	a jel két eleme közé eső beszédszakasz tempója lassabb
°aha°	a szó érezhetően halkabb, mint az őt körülvevő beszédrész
^	nagyobb hangmagasság
~	kisebb hangmagasság

Anyai dajkanyelvi narratívák lexikai és szintaktikai jellemzőinek longitudinális vizsgálata

Harmati-Pap Veronika¹, Vadász Noémi², Kas Bence^{1,3}, Tóth Ildikó⁴

¹Nyelvtudományi Kutatóközpont, Általános és Magyar Nyelvészeti Intézet

²Nyelvtudományi Kutatóközpont, Nyelvtechnológiai és Alkalmazott Nyelvészeti Intézet

³ELTE Bárczi Gusztáv Gyógypedagógiai Kar

⁴Természettudományi Kutatóközpont, Kognitív Idegtudományi és Pszichológiai Intézet

Abstract

Infant-directed speech (IDS) is distinguished by specific acoustic properties and lexical/syntactical simplification from the adult-directed speech (ADS). The lexical and morphosyntactic features of IDS are not well described, especially in Hungarian. Previous studies have conflicting conclusions about how the components of IDS become more complex according to development of the children, whether the advancement is linear or non-linear. We presumed a minimum complexity after the 6 months of age. In this paper we are studying the speech quantity (number of utterances, number of words), lexical features (type-token ratio, verb-noun ratio,) and morphosyntactic features (mean length of utterance, number of clauses) in maternal language during story-telling sessions in a longitudinal research. 22 mother-infant dyads participated at 1 day, 6 months and 18 months of age. We recorded both an IDS and an ADS storytelling in each session. We found the maximum of lexical and morphosyntactical simplification for IDS to the 18 months olds. For the oldest children in our study the mothers told the fairytale with more utterances, applying fewer vocabulary with more repetitions. Their utterances were significantly shorter. Overall, the maternal IDS is characterized by greater verbosity in smaller units. Our results conclude non-linear progression of maternal IDS to the child's age.

Keywords: infant-directed speech, narrative, morphosyntax

1. Bevezetés

A csecsemőkhöz szóló beszéd (*infant directed speech*, IDS) a spontán beszéd egy sajátos formája, amit a felnőttek a csecsemőkkel, kisgyermekkel való kommunikáció során használnak, és ami több jellemzőjében eltér a felnőttekhez szóló beszédétől (*adult directed speech*, ADS). A még beszélni nem tudó csecsemő

Email addresses: `harmati.pap.veronika@nytud.hu` (Harmati-Pap Veronika),
`vadasz.noemi@nytud.hu` (Vadász Noémi), `kas.bence@nytud.hu` (Kas Bence),
`toth.ildiko@ttk.hu` (Tóth Ildikó)

legnagyobb részt az elsődleges gondozójával, - tipikus esetben az édesanyjával - való interakcióban fér hozzá az anyanyelvéhez, így az IDS sajátosságait a klasszikus leírásoktól kezdve a gyermeki nyelvelsajátítás társas támogatását célzó stratégiaként értékeli a szakirodalom (vö. [Snow 1977](#)). Az IDS (a klasszikus hazai szakirodalomban *dajkanyelv*) tipikusan leírt jellemzői a prozódiai és szegmentális fonológiai jelöltség, illetve a lexikai és szintaktikai egyszerűsítés különböző formái (vö. [Snow & Ferguson 1977](#)). A dajkanyelv különböző nyelvi szintekhez kötött sajátosságainak kutatása azonban nagyrészt az angol nyelvre szorítkozott ([Genovese et al. 2019](#)), illetve az IDS-nek a gyermeki fejlődés során mutatkozó, fokozatos lexikai-szintaktikai adaptációját is jobbra az első szavak megjelenését követően vizsgálták. A magyar nyelvben eddig tudtunkkal nem végeztek ilyen irányú kísérleti kutatásokat. Ezt a hiányt igyekszik pótolni a jelen tanulmány, melyben a rögtönzött anyai mesemondás szintaktikai és lexikai jellemzőit vizsgáljuk. Vizsgálatunk újszerűségét egyrészt a vizsgált nyelv, másrészt a kötött tartalmú történetmesélésen alapuló sztenderd – az interakció témájának, szerepviszonyainak esetlegességét kiküszöbölő, minden vizsgált szituációban azonos tartalmú és felépítésű – vizsgálati helyzet adja, melyben a gyerekek különböző életkorában felvett, ADS és IDS helyzetben rögzített felvételeket hasonlítottunk össze. Tudomásunk szerint ilyen kísérleti helyzetre még nem volt példa a dajkanyelv szintaktikai és lexikai vizsgálataiban.

Kutatásunkban a következő jellemzőket vizsgáltuk anyák sztenderd képsorozaton alapuló rögtönzött mesemondásában IDS és ADS helyzetben, a gyermekük újszülött, 6 hónapos és 18 hónapos korában egy longitudinális kutatás keretein belül (a mérési időpontok számos más tényezőt figyelembe véve lettek kijelölve):

- Megnyilatkozásszám: Terjedelmi beszédbőrség mutató, mely kifejezi, hogy az anyák ugyanazt a mesét milyen terjedelemben mesélik el, így mérhetővé válik az anyai beszédbőrség ADS és IDS helyzetben.
- Szószám: Egyrészt terjedelmi beszédbőrség mutató, másrészt a lexikai változatosságot definiáló type-token aránnyal összevetve az anyai beszédben található ismétlések mennyiségének mérésére szolgál.

- Átlagos megnyilatkozáshossz (*mean length of utterance*, rövid.: MLU): Szintaktikai komplexitást mérő szám. A megnyilatkozások hossza mérhető az azokban megjelenő morfémák, szavak vagy szótagok átlagos számával. Jelen tanulmányban a megnyilatkozások hosszát szószámban mérjük.
- Tagmondatszám: A szintaktikai komplexitással összefüggő érték. Az alárendelő összetett mondatokat komplexebbnek tekintjük, mint az egyszerű vagy a mellérendelő összetett mondatokat.
- Type-token arány: Lexikai komplexitást mutató érték. Egy adott szöveg lexikai sokszínűségét mutatja, illetve azt, hogy egy mesében az anya milyen differenciált szókincset alkalmaz. Ez az arányszám a szövegben lévő összes egyedi szó (*type*, ‘szótári szó’) száma elosztva a szöveg teljes szószámával (*token*, ‘szövegszó’).
- Szófaji eloszlás: Lexikai komplexitást mérő mutató. Jelen tanulmányban az igék és főnevek eloszlását vizsgáltuk meg.

E tényezők vizsgálatával arra próbáltunk választ találni, hogy az anyai dajkanyelv mely lexikai és szintaktikai jellemzőkben tér el szisztematikusan a felnőttetekhez szóló beszédétől egy narratív szövegben, illetve hogy ez az IDS-ADS eltérés milyen fejlődésmenetet mutat a gyermek életkora függvényében.

Egyes elképzelések (Snow, 1972; Phillips, 1973) szerint az anyai dajkanyelv a gyermek életkorával párhuzamosan válik viszonylag egyszerű regiszterből egyre komplexebbé, egy lineáris fejlődési folyamatot feltételezve. Genovese et al. (2019) munkája azonban e tényezők változásának menetét nem lineárisnak vizsgálja fel. Feltételezéseik szerint a dajkanyelvben a gyermek első életévének első felében még egy komplexebb szakaszt látunk, majd az első életév második felében szerkezetileg és tartalmilag egyszerűsödik. A gyermek 1 éves kora után az anyai beszéd eléri a legegyszerűbb formáját, amelyet **komplexitási padlónak** nevezhetünk. A komplexitási padló itt tehát azt jelenti, hogy ebben az életkori szakaszban az anyák beszéde a legegyszerűbb mind szintaktikai, mind lexikai szempontból. Ezután a dajkanyelv a gyermek életkorának növekedésé-

vel szintaktikailag és lexikailag ismét egyre komplexebbé válik. Ezt nevezzük **komplexitási u-görbének**.

A kutatásunk célja tehát megvizsgálni, hogy az anyai dajkanyelv komplexitása lineárisan vagy nem lineárisan változik, alkalmazkodva a gyermek aktuális nyelvi képességeihez.

2. A dajkanyelv

2.1. A dajkanyelv mint ösztönös cselekvés

A dajkanyelvi beszédet több klasszikus beszámoló szerint nagymértékű egyszerűsítés és sajátos jelzések jellemzik, melyekkel feltehetően segíti a csecsemők és kisgyerekek nyelvfeldolgozását (Réger, 2002; Phillips, 1973). Ez a törekvés a felnőtt részéről ösztönös, nem tudatos. Egyes kutatások azt mutatják, hogy a dajkanyelvi beszédmód már a prenatális szakaszban, a magzathoz szóló beszédben is jelen van (Chládková et al., 2019). Általában véve megfigyelhető a felnőttek gyermekekhez címzett beszédében, de a legmarkánsabban az anyai beszédet jellemzi. Gergely et al. (2017) szerint habár a dajkanyelvi beszédforma ugyanúgy megjelenik az édesapa beszédében - illetve természetesen minden ember beszédében, aki kisgyermekkel kommunikál - a legnagyobb mérhető eltérés az ADS-hez képest az anyai beszédmódban található. A dajkanyelvi beszéd sajátosságai olyan osztenzív-kommunikatív jelzések a gyermek irányába, amelyek segítségével a) a gyerekek képesek azonosítani, hogy adott kommunikációs aktus feléjük irányul, illetve b) a felnőtt demonstrálja, hogy cselekvése kommunikatív szándékkal történik (Csibra & Gergely, 2009).

Csibra & Gergely (2009) Természetes pedagógia elmélete szerint ilyen elsődleges osztenzív-kommunikatív jelzés a dajkanyelvi beszéd sajátos prozódiaja. Ebben az elméleti keretben a tudás átadása spontán, a pedagógiai szándék ösztönös. A pedagógiai tudásátadást sajátos jegyek váltják ki, mint például a dajkanyelv akusztikai sajátosságai (Gergely et al., 2007). Értelmezésünkben ehhez egy komplex, implicit pedagógiai stratégia részeként szervesen kapcsolódhatnak az IDS lexikai és szintaktikai szinten megjelenő vonásai: a fokozottabb

tagoltság, a szakaszosság, alacsonyabb lexikai sokszínűség és a szintaktikai repertoár alacsonyabb komplexitása, az ismétlések gyakorisága. A dajkanyelvben az anyai kommunikáció alkalmazkodik a gyermek aktuális fejlettségi szintjéhez (Phillips, 1973). A nyelv elemeit a dajkanyelvi környezetben megtapasztaló gyerekek számára a prozódiai és egyéb nem verbális jelzések felhívják a figyelmet a felnőtt pedagógiai szándékára (Csibra, 2010). A lexikai és szintaktikai egyszerűsítések pedig a nyelv elemeinek kiemelhetőségét, megragadását teszik lehetővé a nyelvi bemenet hozzáférhetőségének támogatása útján. Több kutatás igazolja, hogy a dajkanyelvi beszéd szavakra tagolhatósága (Thiessen et al., 2005), illetve a dajkanyelvi tapasztalat mennyisége befolyásolja a későbbi szókincsfejlődést (Weisleder & Fernald, 2013).

2.2. A dajkanyelv akusztikai sajátosságai

A gyermekkel való interakcióban a felnőtt elsősorban beszédének a hangzását tudja módosítani azért, hogy a gyermek figyelmét felkeltse. A dajkanyelvi kutatások elsősorban e regiszter akusztikai tulajdonságaira összpontosítanak, – e terület magyar nyelven alapuló irodalma is egyre számottevőbb. Vizsgálták többek között a prozódiai tulajdonságait (Mády et al., 2018), a rövid–hosszú magánhangzópárok realizációit (Deme et al., 2019) és a frázisvégi és megnyilatkozásvégi nyújtásokat (Szalontai et al., 2018; Kohári et al., 2019), illetve a beszédritmust (Kohári et al., 2020). Eredményeik szerint az ADS-hez képest a csecsemőkhöz szóló beszéd eltérést mutat mind a beszéd temporális szerkezetében, mind az akusztikai tulajdonságokban (pl. zöngéesség). Azonban e kutatások nem mutatták ki, hogy a kisgyermek életkora hatással lenne a vizsgált tényezőkre az anyai beszédben.

2.3. A dajkanyelv szintaktikai és lexikai tulajdonságai

Az anyai dajkanyelv lexikai-szintaktikai komplexitása eltér a felnőttekhez irányított beszédétől. Az anyák – és kisebb mértékben mások is – leegyszerűsítik megnyilatkozásait, amikor csecsemőkhöz, kisgyerekekhez beszélnek. A legtöbb beszámoló szerint az anyai beszéd alkalmazkodik a gyermek fejlettségi szintjéhez, ez az illeszkedés azonban többféle modellt követhet:

- Lineáris változás: A legegyszerűbb elképzelés szerint a dajkanyelv a gyermek életkorának előrehaladtával egyre komplexebbé és komplexebbé válik, majd egyszer csak eléri az ADS komplexitási szintjét (Snow 1972; Phillips 1973). Egyes kutatók szerint az anyai beszéd mennyisége a gyermek életkorával párhuzamosan emelkedik (Henning et al. 2005; Bornstein et al. 1992). Phillips (1973) a dajkanyelvi szintaxisát vizsgálva (átlagos megnyilatkozáshossz) is a linearitás mellett érvel, de vizsgálataiban a legfiatalabb gyermekek 8 hónaposak voltak. Lexikai szempontból is hasonló eredmények születtek. Egyes kutatások szerint a lexikai változatosság mértéke a gyermek életkorával nő (Phillips 1973; Snow 1972; Broen 1972).
- Nem-lineáris változás: A dajkanyelv nem egyenletesen válik egyre hasonlóbbá az ADS-hez a gyermek életkorával párhuzamban, hanem a gyermek bizonyos életkorában éri el szintaktikai és lexikai komplexitásának padlóját, így a fejlődési menet komplexitás tekintetében egy u-görbében halad. Egyes kutatók szerint ez a komplexitási padló a gyermekek 1 éves kora körül figyelhető meg. Stern et al. (1983) adatai szerint 1 éves kor körül van egy csökkenés a szintaktikai komplexitásban. Genovese et al. (2019) olasz nyelvű kutatása szerint ez a legegyszerűbb szakasz, azaz a padló 9 hónapos korban tapasztalható (vö. Phillips 1973 8 hónapos fókuszcsoportjával), és az ennél fiatalabb és idősebb gyermekek esetén az anyák beszéde komplexebb. A dajkanyelvi lexikonban (Genovese et al. 2019) a gyermek első életévének második felében van egy egyszerűsödési szakasz. Longitudinális kutatásukban 80 anya-gyermek páros vett részt a gyermek 3, 6, 9, és 12 hónapos korában. Az anyai dajkanyelvi beszédet spontán, szabad játék helyzetben vizsgálták.

A komplexitás egyik összetevője a *terjedelem*. Egyes kutatások szerint már a beszéd mennyisége is kisebb a gyermekekhez szóló beszédben, az anyák kevesebb megnyilatkozást használnak, az átlagos megnyilatkozáshosszuk rövidebb (Snow 1972; Phillips 1973; Genovese et al. 2019). Néhány kutató szerint a mennyiségi

növekedés párhuzamos a gyermek életkorával (Henning et al. 2005, Bornstein et al. 1992).

Phillips (1973) szerint a dajkanyelv *szintaktikai* szempontból is egyszerűbb – az angol nyelvet tekintve egy felnőtt beszélgetőpartnerhez szólva az átlagos megnyilatkozáshossz 8 szó, míg egy gyermekhez ennek a fele, csupán 4 szó. Mindemellett a felnőttekhez irányított beszédhez képest tagoltabb, jelentéstani és szerkezeti egyszerűsítés jellemzi (Snow, 1972). A *finomhangolás hipotézise* szerint (*fine-tune hypothesis* – Bohannon et al. 1982) a felnőtt beszélők a gyermek verbális visszajelzései alapján módosítják megnyilatkozásaik átlagos hosszát. Két éves gyermekek esetén szignifikánsan rövidebb MLU jellemzi beszédüket történetmesélési szituációban. A felnőtt beszélők rövidebb megnyilatkozásai a gyermek nyelvi fejlődését támogatják, a gyermek szókincsében újonnan megjelenő szavakat a felnőttek is rövidebb mondatokban használják a szó felbukkanása körüli időszakban (Odijk & Gillis 2020). Tomasello & Farrar (1986) azonban felhívja arra a figyelmet, hogy a közös figyelmi helyzetnek itt is kiemelkedő szerepe van. Anya-gyermek párosok kommunikációját 15 és 21 hónapos korban vizsgáló kutatásában arra talált adatokat, hogy az anyák több megnyilatkozást produkálnak gyermekük felé közös figyelmi helyzeten belül, mint azon kívül.

A *lexikai* diverzitást kifejező type-token arány – ami megmutatja, hogy a szótári töveknek (type) hány előfordulása (token) szerepel a korpuszban – egyértelmű dajkanyelvi eltérést mutat ADS-hez képest, azaz az anyák csökkentett szókincsset alkalmaznak a gyermekükkel való kommunikáció során (Broen 1972, Phillips, 1973, Snow, 1972). Ez az arányszám jól érzékelteti az ismétlésre való hajlamot is: minél kisebb type-token arány párosul ugyanazon szövegmennyiséghez, annál nagyobb mértékben szerepelnek ugyanazok a szavak újra és újra ismételve az adott szövegben. Általános megfigyelés, hogy a főnevek perceptuálisan és konceptuálisan egyszerűbbek az igéknél. Adi-Bensaid és munkatársai (2015) ezzel hozzák összefüggésbe, hogy a héber adatokban ADS-ben igei főlényt találtak, míg az IDS-ben az ige-főnév arány kiegyenlítetté vált. A szerzők az ADS-beli igei főlényt a héber nyelv pro-drop tulajdonságával magyarázták,

míg a főnévi “felzárkózás” IDS-ben a gyermeki képességekhez való alkalmazkodás következménye lehet.

Genovese és munkatársai (2019) olasz nyelvi vizsgálata hívja fel a figyelmet arra, hogy az dajkanyelv kutatásoknak vannak fehér foltjai több szempontból is. Kevés kutatás irányult az angolon túl más nyelvekre lexikai-szintaktikai szempontból. Az angol nyelv morfoszintaktikai jellemzői nyilvánvalóan meghatározzák a komplexitás mutatóit, tipológiailag eltérő nyelvekben azonban más nyelvtani területeken fejeződhet ki a komplexitás csökkentése vagy növelése. Ilyen, a komplexitást befolyásoló nyelvtani sajátosság lehet a szóalakok morfológiai összetettsége vagy a névmás-elhagyás jelensége. Mivel az olasz – és a magyar is – ún. *pro-drop* nyelvek, – amelyekben adott névmások elhagyhatók – elengedhetetlen e nyelvek vizsgálata is a dajkanyelv jobb megismeréséhez. Ezen felül, bár a szintaktikai és lexikai egyszerűsítések fejlődési pályáját a klasszikus angol nyelvi vizsgálatok feltérképezték, ezek jobbára a holofrasztikus szakasztól indultak, és kevésbé vizsgálták a csecsemők preverbális szakaszára jellemző IDS tulajdonságait.

Feltételezésünk szerint, a finomhangolási hipotézisnek (Bohannon et al., 1982) megfelelően a dajkanyelvi regiszter alkalmazkodik a gyermek életkorához a gyermek verbális visszacsatolásai alapján. A dajkanyelvi alkalmazkodás azt jelenti, hogy a fiatalabb csecsemőknél – ahol a kommunikatív visszacsatolás még nem olyan erős – az anyák kevésbé egyszerűsítik a beszédüket. Majd valahol 1 éves kor után, ahogy a gyermekek nyelvi képességei is fejlődni kezdenek, ismételt komplexebb beszéd jellemzi őket. A kettő közötti időszakban, a gyermek első életvének második felében, ahol már kommunikatív visszacsatolás van, de a gyermekek nyelvi képességei még csekélyek, a dajkanyelv a legkevésbé komplex, hiszen az anyai beszéd így támogatja legjobban a gyermek kibontakozó produktív nyelvi képességeit. Tehát a dajkanyelvi alkalmazkodás nem statikus, hanem egy dinamikus folyamat, amely a gyermek életkorával párhuzamban folyamatosan változik (Odijk & Gillis, 2020). Fontos felhívni arra a figyelmet, hogy ebben az életszakaszban, 9 hónapos kor körül következik be az ún. *kognitív forradalom* is a kisgyermekek fejlődésében (Tomasello, 2019), amely magával hozza az első

gesztusokat, a közös figyelem képességét és az aktívabb diskurzusbeli részvétel képességét, ami kiválthatja a szülők erősebb alkalmazkodását a gyerek kommunikációs képességeihez.

Véleményünk szerint a lineáris és nem lineáris fejlődési menetet feltételező kutatások összeegyeztethetők. Az előbbi kutatások elsősorban a beszédértés első jeleitől kezdték vizsgálni az anyai dajkanyelv alakulását (ld. [Snow 1972](#) 8 hónapos kor), így nincs módjuk rögzíteni a csecsemőhöz szóló beszéd korai, 6–8 hónapos kort megelőző jellemzőit. A dajkanyelvi komplexitás nem lineáris fejlődésmenete tehát azokban a kutatásokban rajzolódhat ki, melyek a gyermekek újszülött korától vizsgálják azt. Ez abból fakadhat, hogy a felnőtt beszédnek a gyermek nyelvi képességeihez való igazítása nem azonnal váltódik ki az újszülött csecsemő jelenlétében, hanem feltehetően a csecsemő bizonyos kommunikációs jelzései válthatják ki a fejlődés későbbi szakaszában. Várakozásunk szerint tehát eredményeink a nem-lineáris fejlődési menetet fogják mutatni, tehát az újszülött korhoz képest a dajkanyelvi komplexitás csökkenése, majd növekedése egyfajta hullámot, U alakú görbét mutat a legtöbb komplexitási mutató tekintetében, így a szövegmenyiség (terjedelem), a megnyilatkozások átlagos hossza és a lexikai sokszínűség terén. A szófaji arányok tekintetében a magyarban – lévén pro-drop nyelv – [Genovese et al. \(2019\)](#) és [Adi-Bensaid et al. \(2015\)](#) nyomán alapvetően igei túlsúlyt várunk, ugyanakkor feltételezzük, hogy az anyai beszédben IDS helyzetben a főnevek aránya nagyobb lesz az igékhez képest, mint ADS helyzetben. Lényeges tényező az elemzések alapjául szolgáló vizsgálati helyzet módszertana is. A korábbi kutatások túlnyomórészt szabad játék helyzetet alkalmaztak ([Phillips 1973](#); [Bornstein et al. 1992](#); [Genovese et al. 2019](#); [Fernald & Morikawa 1993](#); [Snow 1972](#)) vagy hétköznapi szituációkon alapultak, mint fürdetés, étkezés ([Adi-Bensaid et al. 2015](#); [Tomasello & Farrar 1986](#)). Történetmeselési szituációkra is van példa: [Bohannon et al. \(1982\)](#) kutatásában az alanyok mesét meséltek ugyan, azonban az alanyok gyermektelen egyetemi hallgatók voltak, és gyermekekről készült fotóknak meséltek. [Chládková et al. \(2019\)](#) kutatásában várandós nők képekről meséltek magzatuknak, azonban ADS-sel való összehasonlítást nem találunk ezekben a vizsgálatokban. [Broen \(1972\)](#) tanulmányában

találunk arra adatokat, hogy a szabad játék és a történetmesélés szituáció nem mutat szignifikáns különbséget a dajkanyelvi beszédmodok között. Az ADS-sel való összevetés miatt azonban véleményünk szerint a történetmesélési szituáció az alkalmas, hiszen így tudunk több összehasonlítható, különböző életkorban felvett dajkanyelvi és felnőtthez irányított szövegeket létrehozni.

Azokban az esetekben, amikor a gyermekek életkora szerint egynél több csoport adatait vizsgálták, általánosan anya-gyermek párok különböző csoportjait hasonlították össze (Phillips, 1973; Adi-Bensaid et al., 2015; Bornstein et al., 1992; Fernald & Morikawa, 1993), míg Genovese et al. (2019) kutatása a dajkanyelv időbeli változásaira összpontosít longitudinális vizsgálati keretben. Véleményünk szerint, mivel az anyák személyisége, karakterjegyei, iskolázottsága befolyásolhatja beszédbőségüket, ezért a csoportközi vizsgálatok helyett ugyanazon anyák beszédének ismételt mérését tekintjük megfelelő módszernek a dajkanyelvi változások mérésére, mivel így ugyanazon személyek beszédének változását követhetjük nyomon.

Ennek ellenőrzése érdekében fontosnak tartottuk minél inkább kontrollálni a szituáció körülményeit a meseolvasási szituáció alkalmazásával. A kötött tartalmú történetmesélésen alapuló sztenderd vizsgálati helyzet tehát új azokból a szempontokból, hogy AD és ID helyzetben összehasonlítható szövegeket eredményez, és a gyermek különböző életkoraiban felvett ID szövegek összehasonlítására is alkalmasabb.

3. A vizsgálat körülményei

3.1. Résztvevők

Az anya-csecsemő párokat Budapesten, a Magyar Honvédség Egészségügyi Központ Szülészeti-Nőgyógyászati és PIC Osztályán toboroztuk a kutatás céljainak megfelelő bevonási kritériumok alapján: i) az édesanya elsőszülő, ii) az újszülött érett és egészséges, iii) az édesanya magyar anyanyelvű. Minden édesanya teljes körű tájékoztatást követően írásos beleegyezését adta a részvételbe az adatfelvétel megkezdése előtt. A kutatás etikai engedélyét az ETT-TUKEB

(1942-12/2016) és az EPKEB (77/2015) adta. A teljes longitudinális mintából 22 anya-gyermek pár hangfelvételei álltak rendelkezése mindhárom vizsgált életkorban és mindkét mesemondási (ADS/IDS) helyzetben. A 22 édesanya közül 81,8% rendelkezett főiskolai/egyetemi diplomával, életkoruk átlaga születéskor 31,7 év ($\pm 5,8$) volt. A csecsemők átlagosan 3360 gramm (± 362) súllyal, 39,6 gesztációs hétre ($\pm 0,9$) születtek. Úgy adódott, hogy jelen mintában a csecsemők többsége lány volt ($n=16$, 72,7%). Az újszülöttkori vizsgálatokra a kórházban, míg a későbbi életkorok adatfelvételi hullámaira a Természettudományi Kutatóközpont Kognitív Idegtudományi és Pszichológiai Intézetének humán viselkedéslaborjában került sor. A kutatás helyszíne és követéses jellege miatt nagyrészt iskolázott, Budapesten és vonzáskörzetében élő családok alkották a mintát.

3.2. Adatfelvétel módja

Az anyákkal három alkalommal készítettünk hangfelvételt, gyermekük újszülött, 6 és 18 hónapos korában. Minden alkalommal 2 meseolvasási felvétel készült, egy felnőtthez irányított beszéd (ADS) és egy gyermekéhez irányított beszéd (IDS) helyzetben (II. kép). Azért volt szükség minden egyes alkalommal új ADS felvételre, hogy ezzel kiszűrjessük az anyai beszédet befolyásoló egyéb, szituatív tényezőket. A gyermek újszülött korában ADS és IDS felvétel egyaránt a kórházi kórteremben készült a kísérletvezető jelenlétében. A gyermek 6 és 18 hónapos korában a felvétel az ELKH TTK KPI laborban készült. Az anya elsőként a kísérletvezetőnek olvasta fel a mesét, miközben gyermekével ugyanazon helyiségben egy másik kísérletvezető foglalkozott, minimalizálva az anya-gyermek páros közötti kontaktust – mivel a teljes szeparáció ebben az életkorban nem lehetséges. Ezután a kísérletvezetők kijöttek a helyiségből, ahol így az anya ismét elmesélte a mesét gyermekének, immár kettesben.



1. kép. Adatfelvételi helyzet ADS (bal) és IDS (jobb) helyzetben a TTK KPI laborjában. A képeket az édesanya frásos beleegyezése alapján közöljük.

3.3. Instrukciók

Az édesanyák egy *Manóbújócska* című képeskönyvet kaptak (2. kép), az instrukció pedig az volt, hogy meséljék el a mesét a képek alapján saját szavaikkal, szó szerint beleszöve az a pár mondatot, amelyeket szöveg-buborékban, vagy az oldal alján leírva látnak. Ez utóbbiak a kutatás egy másik, jelen közleményben nem tárgyalt célját szolgálták. A képeskönyv 12 képből állt. A képek alapján az anyukáknak egy 6 főbb eseményből álló cselekményt kellett elmesélniük.



2. kép. Példák a *Manóbújócska* című képeskönyvből.

A meseolvasás alkalmas arra, hogy meghatározott kereteken belüli, egymással összehasonlítható szövegek keletkezzenek, mivel a képes történet vezeti a

beszéd témáját, meghatározza az átadandó információmennyiséget, biztosítja a beszédfolytonosságot (vö. megfigyelés esetén az anya megszólalása nem feltétlenül biztosított).

3.4. Adatelemzés

Az anyai mesékről minden helyzetben hangfelvétel készült, amely hanganyagokat CLAN programban chat formátumban (MacWhinney, 2017) jegyeztünk le, majd minden egyes megnyilatkozáshoz hozzárendeltük a hozzá tartozó hangsáv-szakaszt. Ezekből a lejegyzésekből *chat2text* paranccsal a nyelvfeldolgozó rendszer számára alkalmas flo.cex kiterjesztésű szövegfájlokat hoztunk létre. A szövegek nyelvi elemzését az e-magyar digitális nyelvfeldolgozó rendszer (Váradi et al., 2017) legújabb, moduláris pipeline verziójával, az emtsv-vel (Indig et al., 2019; Simon et al., 2020) végeztük. Az ún. mondatra-bontó és tokenizáló (Mittelholcz, 2017), a morfológiai elemző (Novák, 2014; Novák et al., 2016; Novák, 2003) a szófaji egyértelműsítő (Orosz & Novák, 2012, 2013), valamint a függőségi- és összetevős mondatelemzést (Zsibrita et al., 2013) végző modulok kimenete egy tabulátorral szeparált fájl volt, ahol egy token egy sort foglalt el, az oszlopokban pedig az említett elemző modulok eredménye szerepel. Ezek a nemzetközi trendekhez illeszkedő, magyar fejlesztésű nyelvfeldolgozó eszközök jó minőségű elemzést nyújtanak (a részletes kiértékelésekhez lásd: Simon et al., 2020), ezért utólagos javításra nem volt szükség. A mondatokat üres sor választja el egymástól. E kimeneti adattáblák alapján végeztük el a statisztikai elemzéseinket. Az így kapott adatokat nyolc szempont szerint elemeztük, komplexitás-mérés céljából.

- terjedelem

1. **megnyilatkozásszám/mese:** Egy megnyilatkozásnak tekintettük az egyszerű mondatokat, illetve az alárendelt összetett mondatokat. A mellérendelt összetett mondatok tagmondatait külön megnyilatkozásként kezeltük. A megnyilatkozások száma alapján láthatjuk azt,

hogyan az anyák ugyanazt a mesét milyen terjedelemben mesélik el, így mérhetővé válik az anyai *beszédbőség* ADS és IDS helyzetben.

2. **szószám/mese:** Amellett, hogy a beszéd terjedelmét mérhetjük, a type-token aránnyal párhuzamban kezelendő (ld. lexikai komplexitás, 3. pont), amely által az ismétlések gyakoriságára kaphatunk adatot. Ha megegyező szövegmennyiség esetén alacsonyabb type-token arányt kapunk, akkor az azt jelenti, hogy az anyák többször használják ugyanazon szótípusokat, azaz többször ismételnék.

- lexikai komplexitás

3. **type-token arány:** A szótípusok (type) számát elosztottuk az összes előfordulási számukkal (token). Ez az arányszám megmutatja, hogy a szöveg milyen mértékben áll különböző lexémákból, vagyis az anyák mekkora szókincset alkalmaz a mese elmondásához, ami a lexikai változatosság mérőszáma. Ez az arányszám a szövegben lévő összes egyedi szó (type, 'szótári szó') száma elosztva a szöveg teljes szószámával (token, 'szövegszó'). Az így kapott arányszám 0 és 1 között helyezkedik el, ahol kerekített 0 esetén a teljes szöveg ugyanabból az egy szóból állna, 1 esetén pedig nem szerepelne kétszer ugyanaz a szó a szövegben. Tehát minél jobban közelít ez az arányszám az 1 értékhez, annál nagyobb lexikai változatosságról beszélünk. Az anyai beszéd type-token aránya/változatossága csökken a beszéd komplexitásának csökkenésével.
4. **type-token arány igékre:** Az igei szótípusok számát osztottuk el azok előfordulási számával az igei szókinccsméret méréséhez. Az anyai beszédben az igék lexikai változatossága csökken a beszéd komplexitásának csökkenésével.
5. **type-token arány főnevekre:** A főnévi szótípusok számát osztottuk el azok előfordulási számával az főnévi szókinccsméret méréséhez. Az anyai beszédben a főnevek változatossága csökken a beszéd komplexitásának csökkenésével.

6. **főnév-ige arány:** Azt vizsgáltuk, hogy egy igére átlagosan hány főnév jut, azaz a főnevek számát elosztottuk az igék számával, ami a két fő szófajhoz tartozó szövegszavak egymáshoz viszonyított arányát fejezi ki.

- szintaktikai komplexitás

7. **átlagos megnyilatkozáshossz (mean length of utterance – MLU):** Az átlagos megnyilatkozáshosszal mérhető a szöveg globális szintaktikai komplexitása. A megnyilatkozások hosszát a szavak számában mértük. Egy megnyilatkozásnak tekintettünk minden egyszerű mondatot és alárendelő összetett mondatot. Külön megnyilatkozásnak kezeltük a mellérendelő összetett mondatok tagmondatait.

8. **Egy megnyilatkozásra eső tagmondatok aránya:** Az alárendelő összetett mondatok sorozatát egy megnyilatkozásnak kezeltük, míg a mellérendelő összetett mondatok (nem hiányos) tagmondatait külön megnyilatkozásnak kezeltük. A szintaktikai komplexitás mérhető az egy megnyilatkozásra eső alárendelő tagmondatok számával, ezért a tagmondatok számát elosztottuk a megnyilatkozások számával, így láthatóvá vált, hogy a megnyilatkozások átlagosan hány tagmondatból állnak. Az alárendelő összetett mondatokat komplexebbnek tekintjük, mint az egyszerű vagy a mellérendelő összetett mondatokat.

Az adatelemzéshez ismételt méréses varianciaanalízist alkalmaztunk, 2 faktoros dizájnnal (a beszéd címettségje: ADS, IDS; gyermek életkora: újszülött kor, 6 hó, 18 hó). Az így kapott eredményeken szfericitás-korrekciót hajtottunk végre, ha szükséges volt, majd Tukey-féle eljárással végeztünk post hoc analízist az egyes szintek összehasonlítására. Ez utóbbi esetén az azonos életkorban felvett IDS és ADS eredményeket hasonlítottuk össze egymással, illetve a különböző életkori IDS eredményeket hasonlítottuk össze egymással, hogy láthatóvá váljon az életkori faktor okozta változás iránya. A kiugró értékeket minden esetben

ellenőriztük. A statisztikai elemzéshez a *jamovi* programot használtuk (The jamovi project, 2019).

4. Eredmények

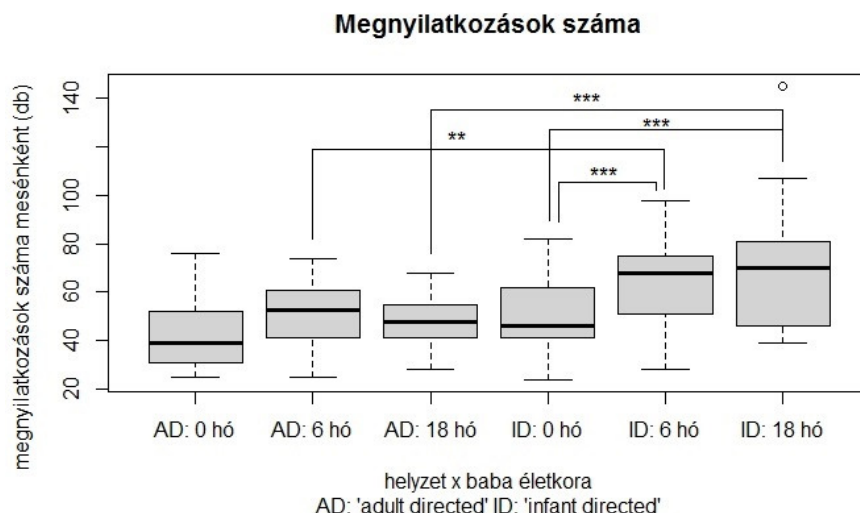
4.1. Terjedelem

4.1.1. Megnyilatkozásszám

A mese címzettjének ($F[1, 21] = 29.3$; $p < 0.001$), a gyermek életkorának ($F[2, 42] = 10.4$; $p < 0.001$) és e két faktor interakciójának ($F[2, 42] = 10.2$; $p < 0.001$) egyaránt szignifikáns hatása volt a megnyilatkozások mennyiségére. A Tukey-féle post hoc tesztek szerint az IDS szövegek szignifikánsan több megnyilatkozást tartalmaznak, mint az ADS szövegek ($p < 0.001$). Az interakció analízise megmutatja, hogy az IDS szövegek még nem térnek el az ADS szövegektől, de a 6 és 18 hós IDS szövegek szignifikánsan hosszabbak mind az azonos időpontban felvett ADS felvételeknél ($p = 0.002$ és $p < 0.001$), mind pedig a 0 hós IDS felvételeknél ($p < 0.001$ és $p < 0.001$), egymástól azonban nem térnek el (II. táblázat; II. ábra). Az ADS szövegek terjedelmi varianciája nem mutatott szignifikáns eltéréseket a gyermek életkora függvényében.

1. táblázat. A megnyilatkozások számának leíró statisztikai adatai

Megnyilatkozások száma	AD 0 hó	AD 6 hó	AD 18 hó	ID 0 hó	ID 6 hó	ID 18 hó
Átlag	42,8	50,8	47,7	48,5	63,3	68,5
Medián	39	52,5	47,5	46	68	70
Minimum	25	25	28	24	28	39
Maximum	76	74	68	82	98	145
Shapiro-Wilk p	0,042	0,892	0,845	0,593	0,577	0,016



1. ábra. A megnyilatkozások számának eloszlása felnőttek (AD) és gyermeknek (ID) szóló történetmesélés esetén, a gyermek 0, 6 és 18 hónapos korában

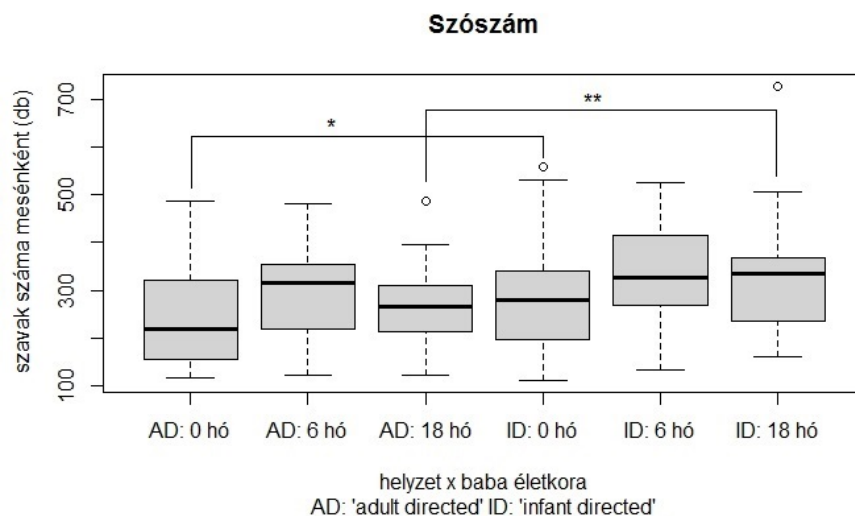
4.1.2. Szószám

A szavak számára hatással volt a mese címzettje ($F[1, 21] = 16.78; p < 0.001$) és a gyermek életkora ($F[2, 42] = 3.50, p = 0.049$). A post hoc eredmények szerint az IDS szöveget szignifikánsan nagyobb szószám jellemzi, mint az ADS-t ($p = 0.001$). Az újszülött kori, 6 hónapos kori és 18 hós IDS mesék egymástól nem térnek el szószám tekintetében. Az újszülött kori és 18 hónapos IDS-ek azonban nagyobb szószámot mutatnak ADS párjukhoz képest ($p = 0.015$ és $p = 0.002$). Az ADS szövegek varianciája nem mutat szignifikáns különbségeket a csecsemők életkora függvényében (2 táblázat; 2 ábra).

Az anyák szignifikánsan több szóval mesélték el ugyanazt a történetet a gyermeküknek, mint egy felnőttnek. 6 hós korban hasonlít leginkább a felnőttnek való meséhez részletességét tekintve, míg újszülött csecsemő és másfél éves gyermek esetén a mese részletesebben kifejtett, jobban magyarázott.

2. táblázat. A szavak számának leíró statisztikai adatai

Szavak száma	AD 0 hó	AD 6 hó	AD 18 hó	ID 0 hó	ID 6 hó	ID 18 hó
Átlag	241	299	269	293	328	331
Medián	220	315	265	280	325	334
Minimum	118	121	123	112	134	160
Maximum	486	481	487	558	524	726
Shapiro-Wilk p	0,078	0,958	0,562	0,191	0,864	0,012



2. ábra. A szavak számának eloszlása felnőttnek (AD) és gyermeknek (ID) szóló történetmesélés esetén, a gyermek 0, 6 és 18 hónapos korában

4.2. Lexikai komplexitás

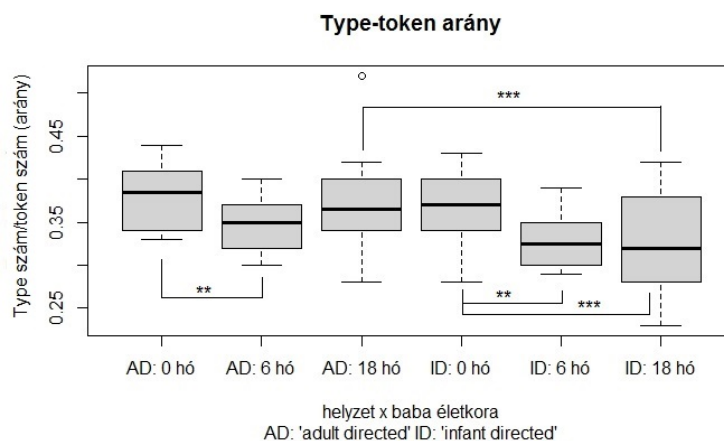
4.2.1. Type-token arány

A beszéd címzettjének ($F[1, 21] = 28.77$; $p < 0.001$), a gyermek életkorának ($F[2, 42] = 10.83$; $p < 0.001$) és e két faktor interakciójának ($F[2, 42] = 4.41$; $p = 0.018$) egyaránt szignifikáns hatása volt a lexikai változatosságra. A Tukey-féle post hoc elemzés kimutatta, hogy az IDS szövegek type-token aránya szignifikánsan alacsonyabb volt, mint az ADS szövegeké ($p < 0.001$). Az újszülött kori és a 6 hónapos kori mese type-token aránya még nem tér el az azonos idő-

pontban felvett ADS-tól ($p = 0.223$ és $p = 0.223$), 18 hós kori szövegek aránya viszont kisebb az azonos időpontban felvett ADS párjukhoz képest ($p < 0.001$). Csak az IDS szövegeket tekintve a 6 és 18 hós kori type-token arányok egyaránt kisebbek, mint az az újszülött kori mérésben ($p = 0.005$ és $p < 0.001$), azonban egymástól nem tértek el (3. táblázat; 3. ábra). Tehát az adatokból az látható, hogy lexikai változatosságot tekintve az anyai dajkanyelv a gyermek 6 és 18 hós korában a legkevésbé komplex. Az ADS szövegek varianciája a 0 és a 6 hónapos kori felvétel esetén mutat csökkenést ($p = 0.005$), a 18 hónapos kori felvételtől egyik sem mutatott eltérést.

3. táblázat. A type-token arány leíró statisztikai adatai

Type-token arány	AD 0 hó	AD 6 hó	AD 18 hó	ID 0 hó	ID 6 hó	ID 18 hó
Átlag	0,384	0,347	0,366	0,367	0,33	0,324
Medián	0,385	0,35	0,365	0,37	0,325	0,32
Minimum	0,33	0,3	0,28	0,28	0,29	0,23
Maximum	0,44	0,4	0,52	0,43	0,39	0,42
Shapiro-Wilk p	0,095	0,175	0,13	0,239	0,206	0,125



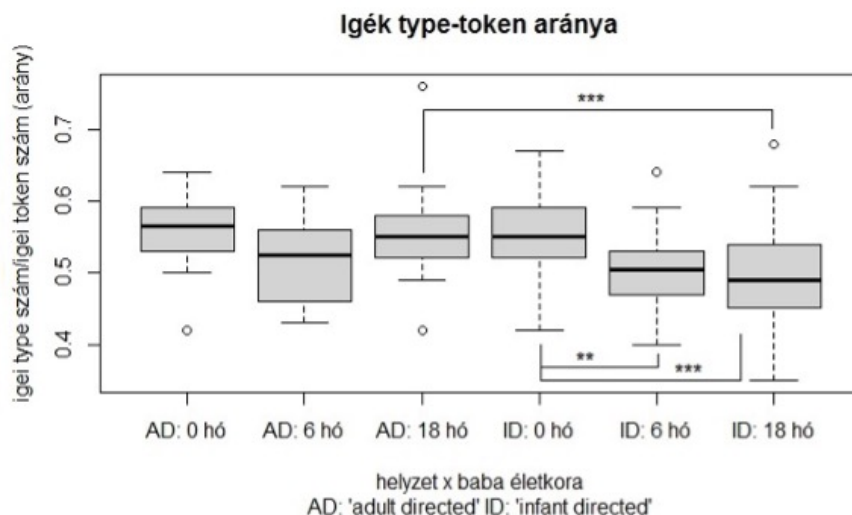
3. ábra. A type-token arány eloszlása felnőttnek (AD) és gyermeknek (ID) szóló történetmesélés esetén, a gyermek 0, 6 és 18 hónapos korában

4.2.2. Type-token arány igék esetén

Az igék változatosságára a beszéd címzettjének ($F[1, 21] = 11.02$; $p = 0.002$), a gyermek életkorának ($F[2, 42] = 7.49$; $p = 0.003$) és e faktorok interakciójának ($F[2, 42] = 3.75$; $p = 0.032$) egyaránt hatása van. A Tukey-féle post hoc analízis megmutatja, hogy ADS-hez képest IDS-ben szignifikánsan alacsonyabb a változatossági arány ($p = 0.003$). Ez az egyes mérési pontokra vetítve a következőt jelenti: 0 hós és 6 hós IDS még nem tér el szignifikánsan ADS párjától, azonban a 18 hós kori IDS szöveg igei változatossága már kisebb, mint mind az ADS párja ($p = 0.001$), mind pedig mint az újszülött kori IDS ($p < 0.001$). A 6 hós IDS type-token arány is kisebb, mint a 0 hós IDS ($p = 0.005$), tehát az életkor előrehaladtával ismét egy komplexitási csökkenést látunk. Az ADS szövegek varianciája nem mutat szignifikáns különbségeket a csecsemők életkora függvényében (4 táblázat; 4 ábra).

4. táblázat. Az igei type-token arány leíró statisztikai adatai

Igei type-token arány	AD 0 hó	AD 6 hó	AD 18 hó	ID 0 hó	ID 6 hó	ID 18 hó
Átlag	0,561	0,515	0,533	0,55	0,498	0,494
Medián	0,565	0,525	0,55	0,55	0,505	0,49
Minimum	0,42	0,43	0,42	0,42	0,4	0,35
Maximum	0,64	0,62	0,76	0,67	0,64	0,68
Shapiro-Wilk p	0,32	0,13	0,021	0,673	0,62	0,934



4. ábra. Az igék type-token arányának eloszlása felnőttek (AD) és gyermeknek (ID) szóló történetmesélés esetén, a gyermek 0, 6 és 18 hónapos korában

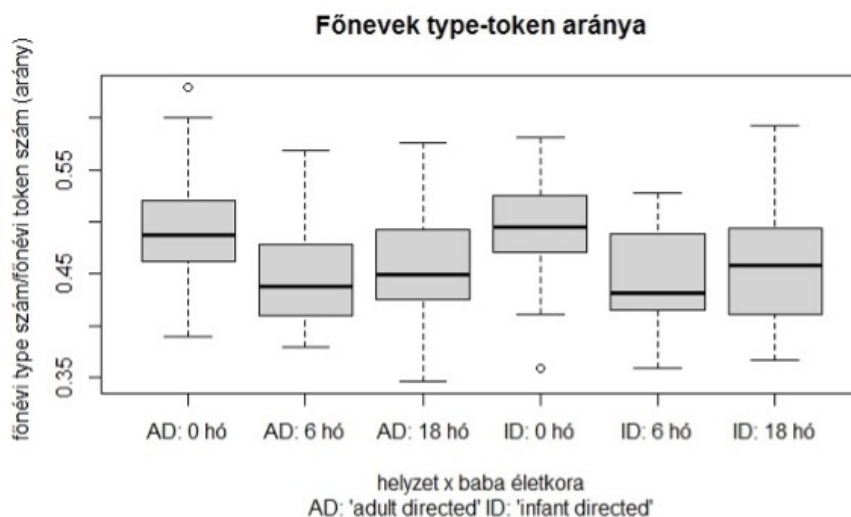
4.2.3. Type-token arány főnevek esetén

A főnevek változatosságát tekintve csak a kor mutat főhatást ($F[1, 21] = 7.12$; $p = 0.002$). A post hoc teszt megmutatta, hogy ez a hatás egy helyzettől független, 0 és 6 hós között látható szignifikáns csökkenés (tehát egyszerűbbé válik), függetlenül attól, hogy az édesanya gyermekének vagy egy felnőttnek meséli a mesét ($p = 0.002$ és $p = 0.043$). Mivel a post hoc teszt nem mutatott azonos életkorban felvett ADS és IDS között különbséget, így ezt az eltérést a vizsgálaton kívüli tényezők okozhatták. Az ADS szövegek varianciája nem mutat szignifikáns különbségeket a csecsemők életkora függvényében (5 táblázat; 5 ábra).

Az eredmények azt mutatják tehát, hogy a történetmondás során használt főnevek változatossági mértéke állandó, az anyák dajkanyelvi környezetben nem módosítják jelentős mértékben a főnévi lexikai választásaikat a gyermek életkorához alkalmazkodva.

5. táblázat. A főnévi type-token arány leíró statisztikai adatai

Főnévi type-token arány	AD 0 hó	AD 6 hó	AD 18 hó	ID 0 hó	ID 6 hó	ID 18 hó
Átlag	0,494	0,446	0,46	0,495	0,444	0,462
Medián	0,485	0,44	0,45	0,495	0,43	0,46
Minimum	0,39	0,38	0,35	0,36	0,36	0,37
Maximum	0,63	0,57	0,58	0,58	0,53	0,59
Shapiro-Wilk p	0,846	0,068	0,423	0,377	0,41	0,583



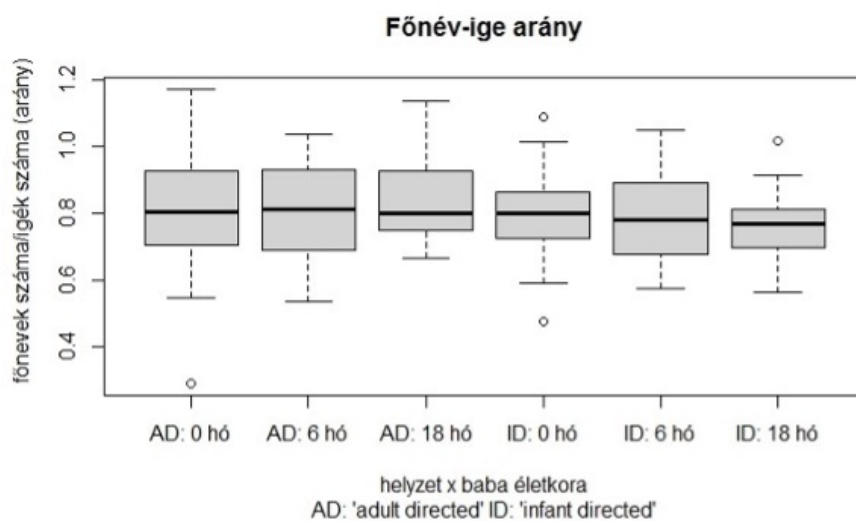
5. ábra. A főnevek type-token arányának eloszlása felnőttnek (AD) és gyermeknek (ID) szóló történetmesélés esetén, a gyermek 0, 6 és 18 hónapos korában

4.2.4. Főnevek és igék aránya a szövegben

Sem a helyzetnek, sem a gyermek életkorának, sem ezek interakciójának nincs hatása a főnevek és igék arányára. Az anyai beszéd főnév-ige aránya megegyezik felnőtthez és gyermekhez szóló beszéd esetén, a gyermek 0, 6 és 18 hós korában egyaránt (6. táblázat; 6. ábra). A szövegekben megjelenő igei túlsúly általánosan jellemző, ahogy azt a pro-drop nyelveknél elvárhatjuk (Genovese et al., 2019).

6. táblázat. A főnévi-ige arány leíró statisztikai adatai

Főnév-ige arány	AD 0 hó	AD 6 hó	AD 18 hó	ID 0 hó	ID 6 hó	ID 18 hó
Átlag	0,804	0,825	0,839	0,789	0,79	0,767
Medián	0,8	0,81	0,795	0,8	0,78	0,765
Minimum	0,29	0,54	0,67	0,48	0,57	0,56
Maximum	1,17	1,04	1,14	1,09	1,05	1,02
Shapiro-Wilk p	0,357	0,57	0,094	0,864	0,72	0,733



6. ábra. A főnevek és igék arányának eloszlása felnőttek (AD) és gyermeknek (ID) szóló történetmesélés esetén, a gyermek 0, 6 és 18 hónapos korában

Tehát az anyai dajkanyelvet szignifikáns lexikai egyszerűsítés jellemzi, amely elsődlegesen az igei szóhasználatban realizálódik.

4.3. Szintaktikai komplexitás

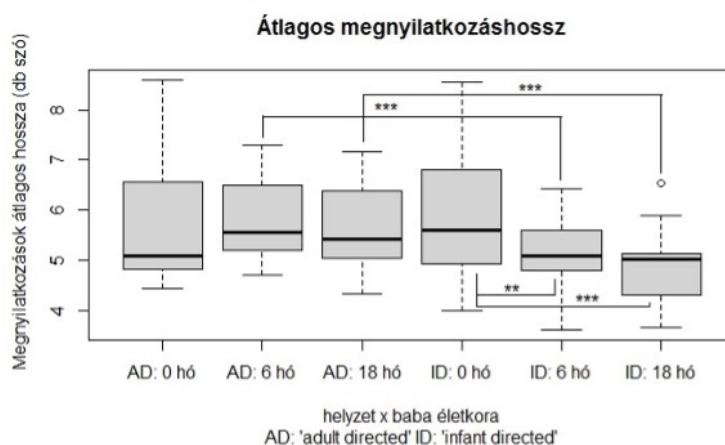
4.3.1. Átlagos megnyilatkozáshossz (MLU)

A mese címzettjének ($F[1, 21] = 13$; $p = 0.002$), a gyermek életkorának ($F(2, 42) = 6.31$; $p = 0.010$) és e két faktor interakciójának ($F[2, 42] = 14.52$; $p < 0.001$) egyaránt erős szignifikáns hatása volt a megnyilatkozások átlagos

hosszára. Az egyes faktoron belüli szintek analízise során láthatóvá vált, hogy az IDS szövegek átlagos megnyilatkozáshosszai szignifikánsan rövidebbek, mint az ADS-ben. Az interakciós szintek vizsgálatakor kiderül, hogy a 0 hós IDS felvételekben az MLU még nem tér el az ADS-tól, azonban a 6 és 18 hós felvételekben már jóval rövidebbek mind ADS-hez ($p = 0.001$ és $p < 0.001$), mind a 0 hós IDS átlagos megnyilatkozáshosszhoz képest ($p = 0.002$ és $p < 0.001$), egymástól viszont nem különböznek lényegesen. Az ADS szövegek varianciája nem tért el szignifikánsan egy életkori szinten sem (7. táblázat; 7. ábra).

7. táblázat. Az átlagos megnyilatkozáshossz leíró statisztikai adatai

Átlagos megnyilatkozáshossz	AD 0 hó	AD 6 hó	AD 18 hó	ID 0 hó	ID 6 hó	ID 18 hó
Átlag	5,59	5,82	5,58	5,88	5,18	4,88
Medián	5,08	5,57	5,42	5,61	5,08	5,03
Minimum	4,45	4,71	4,32	4	3,62	3,67
Maximum	8,59	7,3	7,16	8,55	6,43	6,54
Shapiro-Wilk p	0,003	0,077	0,282	0,336	0,727	0,574



7. ábra. Az átlagos megnyilatkozáshossz (MLU) eloszlása felnőttnek (AD) és gyermeknek (ID) szóló történetmesélés esetén, a gyermek 0, 6 és 18 hónapos korában

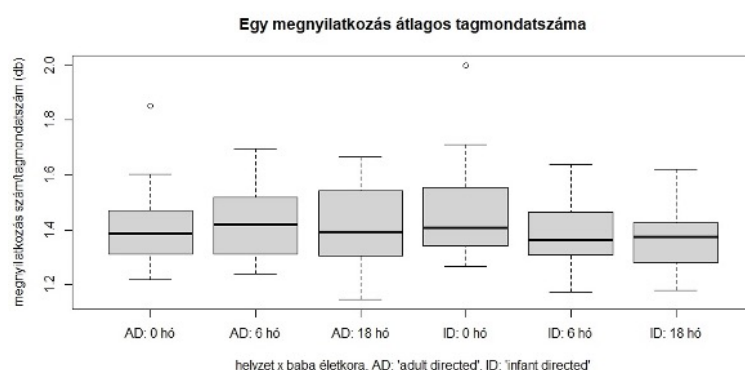
A babákhoz irányított mesében az anyai dajkanyelv egyszerűbb az átlagos megnyilatkozáshosszt tekintve, azaz a 6 és 18 hónapos babákhoz az anyák szignifikánsan rövidebb egységekben beszélnek. Újszülött babákhoz irányított beszédben ez az egyszerűsödés még nem látható.

4.3.2. Tagmondat-megnyilatkozás arány

Az elmondott mesékben átlagosan 1,28 tagmondat jutott egy megnyilatkozásra, ám ebben az esetben sem a címzettnek, sem a gyermek életkorának nem volt hatása erre az arányra (8. táblázat; 8. ábra). Az anyai beszédben az egy megnyilatkozásra eső tagmondatok aránya megegyezik felnőtthez és gyermekhez szóló beszéd esetén, a gyermek 0, 6 és 18 hós korában egyaránt.

8. táblázat. A tagmondat-megnyilatkozás arány leíró statisztikai adatai

Tagmondat-megnyilatkozás arány	AD 0 hó	AD 6 hó	AD 18 hó	ID 0 hó	ID 6 hó	ID 18 hó
Átlag	1,41	1,43	1,42	1,46	1,37	1,37
Medián	1,38	1,42	1,39	1,41	1,36	1,38
Minimum	1,22	1,24	1,23	1,2	1,17	0,96
Maximum	1,85	1,7	1,67	2	1,64	1,62
Shapiro-Wilk p	0,024	0,131	0,167	0,005	0,861	0,2



8. ábra. A tagmondatok és megnyilatkozások arányának eloszlása felnőttnek (AD) és gyermeknek (ID) szóló történetmesélés esetén, a gyermek 0, 6 és 18 hónapos korában

4.4. Eredmények összegzése

Az eredmények azt mutatják, hogy számos terjedelmi és komplexitási mutatót tekintve, a dajkanyelv a csecsemő születését követően még nem tér el jelentősen a felnőttekhez szóló beszédétől, azonban [Genovese et al. \(2019\)](#) eredményeivel összhangban, később jelenik meg az egyszerűsödés. Adatainkban azonban mind terjedelmi, mind a lexikai és szintaktikai komplexitás szempontjából a „padló” a gyermekek **18 hónapos korában látható**.

Terjedelem: Az anyák a gyermek 18 hónapos korában a történetet hosszabb, több megnyilatkozásból, több szóból álló leírásával mutatják be gyermeküknek.

Lexikai komplexitás: A type-token arány, azaz az alkalmazott szókincsük mérete is 18 hónapos korban a legszűkebb, és ez elsősorban az igei választásaikra van hatással, míg a főnévi variancia változatlan. Általános igei túlsúly jellemzi mind ADS, mind IDS szövegeket, ahogy azt a pro-drop nyelvek esetén elvártuk ([Genovese et al. \(2019\)](#)). A főnevek konceptuális egyszerűségének hatásával ([Adi-Bensaid et al. \(2015\)](#)), illetve a történetmondás alapjául szolgáló képanyag sajátosságaival függhet össze, hogy a főnevek lexikai változatossága állandónak bizonyult IDS és ADS helyzetben, míg a konceptuálisan komplexebb igék esetén az anyák kisebb igei szókincset alkalmaznak IDS-ben. A megnyilatkozásszám növekedésével a szószám is szignifikánsan növekedett. Ez a type-token arányok eredményeivel párhuzamban azt jelenti, hogy az anyák szignifikánsan hosszabban, nagyobb szómennyiséggel mesélnek a gyermeküknek, azonban ehhez jóval kisebb szókincset alkalmaznak. Tehát a beszédüket sokkal több ismétlés jellemzi, többször előhívják ugyanazt a szót a történet ismertetése céljából, mint egy felnőtthez irányított történetmesélésben.

Szintaktikai komplexitás: Az átlagos megnyilatkozáshossz is szignifikánsan rövidebb a gyermekeknek mesélt történetek esetén, ahol szintén 18 hónapos korban láthatjuk a legrövidebb megnyilatkozásokat. ADS és IDS aszerint nem tért el, hogy egy megnyilatkozás átlagosan hány tagmondatot tartalmaz – tehát nem jellemezte az ADS-t több összetett alárendelt mondat. Ez azt jelenti, hogy az anyák terjedelmileg bővebben, azonban kisebb és egyszerűbb egységekre tagolva közvetítették a mesék tartalmát gyermekük felé, mint a felnőttek felé. Ebben

az összefüggésben is a 18 hónapos kori felvételek mutatják a legmarkánsabb különbséget, azaz itt látható az anyai beszéd legnagyobb tagoltsága.

Mind a nyolc vizsgált komplexitási tényező esetén látható volt, hogy az ADS felvételek terjedelem és komplexitás szempontjából egységesek voltak, a gyermek életkorának nem volt szignifikáns hatása. Ezért kizárhatjuk azt, hogy az IDS-ekben megjelenő eltéréseket egyéb tényezők okozták. Egyedül a lexikai változatosságnál (type-token arány) volt látható különbség a felnőttnek címzett mesében a gyermek 0 és 6 hónapos kora között, amelyet valamilyen vizsgálaton kívüli tényező okozhatott. Ez a külső hatás azonban időben nem esik egybe a 6 és 18 hós életkorokban felvett IDS-ek közötti eltérést, illetve a 18 hónapos kori ADS és IDS közötti eltérést okozó hatásokkal, így a következtetéseinkre nincs befolyással.

Összefoglalva tehát az anyai nyelv lexikai és morfoszintaktikai tulajdonságait egyfajta ösztönös **stratégiai redundancia** jellemzi, amelynek adataink szerinti csúcspontja 18 hónapos korban látható.

5. Diskusszió

A lineáris növekedési hipotézissel (Phillips 1973; Snow 1972; Henning et al. 2005; Bornstein et al. 1992) szemben a dajkanyelv egyértelműen csökkenő komplexitást, növekvő redundanciát és tagoltságot mutat 0–18 hó között, ami sokkal inkább támogatja a nem lineáris növekedési hipotézist. Genovese és munkatársai (2019) eredményeihez hasonlóan, adatainkból egy komplexitási u-görbe rajzolódik ki, mely szerint a gyermek fejlődésének korai szakaszában valahol van egy komplexitási padló, ahol az anyai beszéd szintaktikai és lexikai szempontból a legegyszerűbb. Az olasz kutatás a preverbális korban feltételezett komplexitási padlót, ezzel szemben a mi adatainkban 18 hónapos korban láthattuk a legegyszerűbb szakaszt, mind lexikai (lexikai változatosság, ismétlések aránya), mind szintaktikai szempontból konzisztensen (átlagos megnyilatkozáshossz, megnyilatkozásszám). Az eredményeinkből így nem látszik a Genovese et al. (2019) által feltételezett u-görbe, csupán egy ereszkedő tendencia. Ez azonban nem

feltétlenül ellentmondás. Adataink nem zárják ki azt, hogy a 6 hónapos kori és a 18 hónapos kori mérés között lehetséges egy még kevésbé komplex szakasz. Kiegészítve tehát a korábbi kutatásokat arra találtunk adatokat, hogy az anyai dajkanyelvi szöveg lexikailag és szintaktikailag 18 hónapos korban még nem éri el az ADS szintet, de még a 6 hónapos kori komplexitási szintre sem tért vissza. Az átlagos megnyilatkozáshossz rövidülése azonban összhangban van a finomhangolási hipotézissel, miszerint a gyermek verbális visszacsatolási képességeihez alkalmazkodva a dajkanyelvi beszéd 2 éves kor körül is eltér ADS-től (Bohannon et al., 1982; Odijk & Gillis, 2020).

Ettől függetlenül lehetséges, hogy a 18 hónapos kori mérési pont már a komplexitási görbe emelkedő szakaszán van. A teljes képhez elengedhetetlen fontosságú a kutatás folytatása nagyobb életkori mérési pontokkal összehasonlítva.

Az adatok egyértelműen illeszkednek a szakirodalmi trendbe, miszerint az anyák jóval szűkebb szókincset alkalmaznak a gyermekük irányába (Broen, 1972; Phillips, 1973; Snow, 1972). Fontos azonban, hogy arra is találtunk adatokat, hogy ez a szűkebb szókincs a magyarban elsősorban az igei szóosztályt érinti. Ez tulajdonképpen illeszkedik ahhoz a feltételezéshez, hogy a főnevek perceptuálisan könnyebben feldolgozhatók a gyermekek számára, ahogy azt (Adi-Bensaïd et al., 2015) héber nyelvi adatai alapján várhattuk, miszerint az ADS-ben tapasztalható igei fölény IDS-ben eltűnik. A fent bemutatott magyar nyelvi adatainkban ez a különbség ugyan nem jelent meg, a főnév-ige arány állandó volt mind ADS-ben, mind IDS-ben. Azonban az egyszerűsítési folyamat az alkalmazott igei szókincsméretben megnyilvánult, azaz az IDS-ben alacsonyabb volt az igei type-token arány, mint ADS-ben. Az anyák beszédükben a cselekvések, történések kifejezését, leírását egyszerűsítik, így alkalmazkodva a gyermekük fejlődési szintjéhez, míg a főnévi szóhasználatukat nem érinti az egyszerűsödési szándék: az minden életkorban és beszédhelyzetben állandó.

Az eredmények azonban mind a szószámot, mind a megnyilatkozások számát tekintve ellentmondanak a korábbi állításoknak, miszerint a gyerekhez irányított beszéd mennyiségileg kevesebb, mint a felnőttekhez irányított beszéd (Snow, 1972; Phillips, 1973; Genovese et al., 2019). A legtöbb kutatás szabad játék-

helyzeti diskurzus szituációt vizsgált. A mi eredményeink alapján az látható, hogy a beszéd mennyiségi csökkenésének jelensége nem általánosítható, mert van olyan diskurzus szituáció, amelyben ennek az ellenkezője igaz. A mesék hosszát tekintve konzisztensen azt az eredményt kaptuk, hogy a gyermekeknek mondott mesék sokkal részletesebbek és hosszabbak a gyermek 6 és 18 hónapos korában. Egy ok lehet az, hogy érvényesül az édesanya tanító stratégiája, amely elvezethet a 9 hónapos kor körül a közös figyelmi helyzet kialakulásához, amely többek között a nyelvtanulásnak is kerete (Tomasello & Farrar, 1986). Ha a 6 és 18 hós gyermekek irányába való beszédbővséget párhuzamba állítjuk ugyanezen életkorokban felvett IDS-ek szignifikánsan rövidebb megnyilatkozáshosszaival, az válik egyértelművé, hogy az anyák jóval hosszabban, de jóval tagoltabban beszélnek gyermekükhöz. Ezt a következő példák szemléltetik (a példák egy ugyanazon beszélőtől származnak, 18 hónapos kori ADS és IDS felvétel leírataiból, ugyanazt az eseményszakaszt elmesélve):

a) **ADS:**

*MOT: és Szut ott tartott a kezében egy nagy diót.

*MOT: és ült egy fészekben.

*MOT: és a dióban volt több kis dió.

*MOT: azt meg osztotta a barátaival.

*MOT: akik megkönnyebbülten a sok ijedelem után jóízűen
falatozták.

b) **IDS:**

*MOT: itt ül fönt a fán.

*MOT: fogja a nagy diót a kezébe.

*MOT: látod?

*MOT: és itt ülnek most együtt.

*MOT: megtalálták Szutot végre sok izgalom után.

*MOT: és jóízűen falatoznak együtt.

*MOT: na mit szólsz ehhez a négy törpe meséhez?

6. Összegzés

Az anyai dajkanyelvet erős redundanciára való törekvés jellemzi lexikai és morfoszintaktikai szinten. Az átlagos megnyilatkozáshossz rövidül, a lexikai változatosság csökken, miközben ezekkel párhuzamban a megnyilatkozások, a tagmondatok és a szavak száma növekszik, tehát a mese hossza növekszik. Ezért a kutatás alapját adó rögtönzött mesemondási szituációban **nagyobb tagoltságra való törekvést** látunk, azaz az édesanya többet magyaráz, a mese eseményeit többször körülírja. Ezt **jobban elkülönülő és könnyebben feldolgozható egységekben** közvetíti a gyermek felé. A lexikai változatosság mérésére szolgáló a type-token arány is az anya a gyermek feldolgozási kapacitásához való alkalmazkodását mutatja. Hosszabban és részletesebben mesél, de a tagoltság mellett kisebb szókincset alkalmaz, **többször használja ugyanazokat a megnevezéseket, többször ismételi**, mint a felnőtthez irányított történetmesélésben. Szófaji szempontból megvizsgálva a lexikai változatosságot azt találjuk, hogy a főnevek változatossága egységes minden életkorban a beszéd címzettjétől függetlenül, a különbséget az igehasználat type-token arányának csökkenése adja, vagyis az **elbeszélés predikátumaiban látható a csökkentett szókincsméret**. Ezekre a dajkanyelvi tényezőkre a gyermek életkorának is hatása van. Adataink szerint az anyai beszéd a gyermek 18 hónapos korában a legegyszerűbb IDS helyzetben, mert szignifikánsan ekkor mesél a legrészletesebben, a legtöbb (igei) ismétléssel, de a legrövidebb egységekben. Genovese et al. (2019) munkájában a legegyszerűbb életkornak a 9 hónapos kor tekinthető, ám ez nem zárja ki az eredményeink illeszkedését. Jelen kutatás nem tért ki a 9 hónapos kori mérésre ugyan, de ugyanúgy csökkenést mutat a gyermek első életévében, tehát mindenképpen kizárja a lineáris hipotézist. Habár a mi adatainkban az u-görbe nem kimutatható, nem is zárják ki azt, hogy a tendencia megegyezik. A 18 hónapos kori mérésakor tapasztalt komplexitási jellemzők lehetséges, hogy

már egy komplexitási „felszálló ág” részét képezik, ami azonban még nem érte el a 6 hónapos kori komplexitási szintet.

Az eredmények tekintetében további kutatási kérdések merültek fel, amelyek alapján vizsgálatunkat folytatni tervezzük. Szükséges a szövegek további kutatása a gyermekek későbbi életkorában is, hogy megtudjuk, mely életkorban kezd a magyar dajkanyelvi beszédmód szintaktikai és lexikai szempontból az ADS-hez közelíteni, illetve milyen hatással van az anyai narratív szövegalkotásra, ha a gyermek már a verbális nyelvfejlődési szakaszba lép. A résztvevők viszonylag alacsony létszáma miatt jelen eredményeinkben megjelenő tendenciákat további résztvevők bevonásával tervezzük ellenőrizni. Kutatásunk következő szakaszában az egyének közötti variabilitást is vizsgálni fogjuk, illetve vizsgálni fogjuk a fonológiai-prozódiai, illetve a lexikai-szintaktikai szinten megjelenő IDS-jellemzők összefüggéseit az anyai viselkedés és személyiség egyéb jellemzőivel.

Köszönetnyilvánítás

A tanulmány az NKFIH 115385 projekt (A korai nyelvfejlődés neuro-kognitív előrejelzői, vezető kutató: Winkler István) és az NKFIH 124477 projekt (Szociális, kognitív és genetikai tényezők szerepe a nyelvi fejlődésben, vezető kutató: Kas Bence) támogatásával készült. A szerzők köszönetet mondanak a kutatásban résztvevő gyerekeknek és szüleiknek, illetve az adatgyűjtést végző munkatársaknak: Balázs Andreának, Haraszi Zsófiának, Horváth Blankának, Őzéné Kende Líviának, Roschéne Farkas Juditnak, Radányi Annának, Szekeres Rebekának, Szerafin Ágnes Katának és Uhrin Beatrixnak. Emellett külön köszönetet szeretnénk mondani Winkler Istvánnak, a TTK Kognitív Idegtudományi és Pszichológiai Intézet, a Hang- és Beszédszlelési kutatócsoport csoportvezetőjének és a NKFIH 115385 számú, *A korai nyelvfejlődés neuro-kognitív előrejelzői* című projekt vezető kutatójának.

Hivatkozások

- Adi-Bensaid, L., Ben-David, A., & Tubul-Lavy, G. (2015). Content words in Hebrew child-directed speech. *Infant Behavior and Development*, 40, 231–41.
- Bohannon, J., Lotz Stine, E., & Ritzenberg, D. (1982). The "fine-tuning" hypothesis of adult speech to children: Effects of experience and feedback. *Bulletin of the Psychonomic Society*, 19, 201–204.
- Bornstein, M. H., Tal, J., Rahn, C., Galperín, C. Z., Pêcheux, M. G., Lamo-ur, M., & Tamis-LeMonda, C. S. (1992). Functional analysis of the contents of maternal speech to infants of 5 and 13 months in four cultures: Argentina, France, Japan, and the United States. *Developmental Psychology*, 28, 593–603.
- Broen, P. A. (1972). The verbal environment of the language-learning child. *ASHA Monographs*, No. 17.
- Chládková, K., Černáa, M., Paillereaua, N., Skarnitzla, R., & Oceláková, Z. (2019). Prenatal Infant-directed Speech: Vowels and Voice Quality. URL: <https://fonetika.ff.cuni.cz/wp-content/uploads/sites/104/2019/08/Ch1-et19-prenatal-IDS.pdf>.
- Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, 25, 141–168.
- Csibra, G., & Gergely, Gy. (2009). Natural pedagogy. *Trends in Cognitive Science*, 13, 148–153.
- Deme, A., Kohári, A., Mády, K., Reichel, U., & Szalontai, Á. (2019). A magánhangzós hosszúsági fonológiai kontraszt a dajkanyelvben a csecsemő életkorának függvényében. *Beszédkutatás*, 27, 221–242.
- Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, 3, 637–656.

- Genovese, G., Spinelli, M., Romero, L. L., Aureli, T., Castelletti, G., & Fasolo, M. (2019). Infant-directed speech as a simplified but not simple register: A longitudinal study of lexical and syntactic features. *Journal of Child Language*, 47, 22–44.
- Gergely, A., Faragó, T., Galambos, Á., & Topál, J. (2017). Differential effects of speech situations on mothers' and fathers' infant-directed and dog-directed speech: An acoustic analysis. *Scientific Reports*, 7, 13739.
- Gergely, Gy., Egyed, K., & Király, I. (2007). A természetes pedagógiáról: On natural pedagogy. *Magyar Pszichológiai Szemle*, 62, 107–125.
- Henning, A., Striano, T., & Lieven, E. V. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development*, 28, 519–536.
- Indig, B., Sass, B., Simon, E., Mittelholcz, I., Kundráth, P., & Vadász, N. (2019). emtsv – Egy formátum mind felett. In G. Berend, G. Gosztolya, & V. Vincze (Eds.), *MSZNY 2019, XV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2019)* (p. 235–247). Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport.
- Kohári, A., Deme, A., Reichel, U., Szalontai, Á., & Mády, K. (2019). A dajkanyelv temporális jellemzői 4 és 8 hónapos csecsemőkhöz szóló beszédben. *Beszédkutatás*, 27, 243–258.
- Kohári, A., Deme, A., Reichel, U., Szalontai, Á., & Mády, K. (2020). A beszédritmus időbeli dimenziójának jellegzetességei a dajkanyelvben. In *Nyelvfejlődés csecsemőkortól kamaszkorig* (p. 195–208). Budapest: ELTE Eötvös Kiadó.
- MacWhinney, B. (2017). Tools for analyzing talk part 2: The CLAN program. URL: <http://talkbank.org/manuals/CLAN.pdf> retrieved from.
- Mády, K., Reichel, U., Szalontai, Á., Kohári, A., & Deme, A. (2018). Prosodic characteristics of infant-directed speech as a function of maternal parity. In

- K. Klessa (Ed.), *Proceedings 9th International Conference on Speech Prosody 2018, Poznań, Poland* (p. 294–298).
- Mittelholcz, I. (2017). emToken: Unicode-képes tokenizáló magyar nyelvre. In V. Vincze (Ed.), *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017). Szegedi Tudományegyetem Informatikai Tanszékcsoport* (p. 70–78).
- Novák, A. (2003). Milyen a jó Humor? In *I. Magyar Számítógépes Nyelvészeti Konferencia* (p. 138–144).
- Novák, A. (2014). A new form of Humor – Mapping constraint-based computational morphologies to a finite-state representation. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14). European Language Resources Association (ELRA)*. Reykjavik, Iceland.
- Novák, A., Siklósi, B., & Oravecz, Cs. (2016). A new integrated open-source morphological analyzer for Hungarian. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA)*. Paris, France.
- Odijk, L., & Gillis, S. (2020). Fine lexical tuning in infant directed speech to typically developing children. *Journal of Child Language*, 48, 1–14.
- Orosz, Gy., & Novák, A. (2012). PurePos 2.0 – an open source morphological disambiguator. In B. In. Sharp, & M. Zock (Eds.), *Proceedings of the 9th International Workshop on Natural Language Processing and Cognitive Science. Wroclaw* (p. 53–63).
- Orosz, Gy., & Novák, A. (2013). PurePos 2.0: a hybrid tool for morphological disambiguation. In *Proceedings of the International Conferen-*

- ce *Recent Advances in Natural Language Processing RANLP 2013.INCOMA Ltd* (p. 539–545). Shoumen, BULGARIA, Hissar, Bulgaria. URL: <https://www.aclweb.org/anthology/R13-1071>.
- Phillips, J. R. (1973). Syntax and vocabulary of mothers’ speech to young children: age and sex comparisons. *Child Development*, 44, 182–185.
- Réger, Z. (2002). *Utak a nyelvhez*. Budapest: Soros Alapítvány, MTA Nyelv-tudományi Intézet.
- Simon, E., Indig, B., Kalivoda, Á., Mittelholcz, I., Sass, B., & Vadász, N. (2020). Újabb fejlemények az e-magyar háza táján. In G. Berend, G. Gosztolya, & V. Vincze (Eds.), *MSZNY 2020, XVI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2020* (p. 29–42). Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport.
- Snow, C. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, 4, 1–22.
- Snow, C. E. (1972). Mothers’ speech to children learning language. *Child Development*, 43, 549–565.
- Snow, C. E., & Ferguson, C. A. (1977). *Talking to Children*. Cambridge: Cambridge University Press.
- Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: infant age and context related changes. *Journal of Child Language*, 10, 1–15.
- Szalontai, Á., Mády, K., Deme, A., & Kohári, A. (2018). Prosodic boundaries in Hungarian infant-directed speech. In *Proceedings Experimental and Theoretical Advances in Prosody (ETAP) 4*. Paper: 53.
- The jamovi project (2019). jamovi (version 0.9) [computer software]. URL: <https://www.jamovi.org>.

- Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7, 53–71.
- Tomasello, M. (2019). *Becoming human: A Theory of Ontogeny*. Cambridge, MA: Belknap Press of Harvard University Press.
- Tomasello, M., & Farrar, M. (1986). Joint attention and early language. *Child Development*, 57, 1454–1463.
- Várad, T., Simon, E., Sass, B., Gerőcs, M., Mittelholcz, I., Novák, A., Indig, B., Prőszék, G., Farkas, R., & Vincze, V. (2017). Az e-magyar digitális nyelvfeldolgozó rendszer. In *MSZNY 2017, XIII. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, Szegedi Tudományegyetem Informatikai Tanszékcsoport* (p. 49–60).
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science*, 24, 2143–2152.
- Zsibrita, J., Vincze, V., & Farkas, R. (2013). magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In *Proceedings of RANLP 2013* (p. 763–771).

Artikulációs fonetikai jellemzők verifikálása kvantitatív adatokkal

Trencsényi Réka¹, Czap László²

¹*Debreceni Egyetem, Villamosmérnöki Tanszék*

²*Miskolci Egyetem, Automatizálási és Infokommunikációs Intézet*

Abstract

This paper aims to verify the phonetic features of articulation by quantitative data, whereby it becomes possible to determine the base data set of visemes – the visual counterparts of phonemes – with quantitative data in order to provide accurate input for visual speech synthesis (a talking head that supports the training of speech production of deaf and hard of hearing children). Measurement-based features extend the existing data and refine our previously used dynamic model of articulation. This endeavour requires the definition of two major types of data simultaneously: 1.) Information connected to the shape of the mouth, which can be examined relatively simply in an ordinary camera image. 2.) Parameters describing the position of the tongue, gaining of which requires the use of medical-level imaging devices and the processing of their signals. The place of articulation of sounds can be described by the shape and position of the tongue. In the case of vowels, we estimated the tongue position with the centroid of the tongue, while in the case of consonants, we define the place of articulation with the measured distance of the tongue from the palate. In our examinations, we use dynamic MRI images and determine the relevant tongue contours by running automatic algorithms. On the track of our analysis, such a data set is created that statically defines the articulatory key frames (fixing the tongue position belonging purely to the given speech sound, without the properties of sound transitions) playing an important role in visual speech synthesis.

Keywords: articulation, phonetic characteristics, quantitative phonetics, articulatory chart, dynamic MRI recording, tongue and palate contour tracking

1. Bevezető

A korábbi tanulmányok azt mutatják, hogy az emberi beszéd fiziológiai folyamataihoz kapcsolódó vizuális információk nagyban hozzájárulnak a beszédképzés komplex mechanizmusának megértéséhez, és ezáltal a beszéd-szintézis

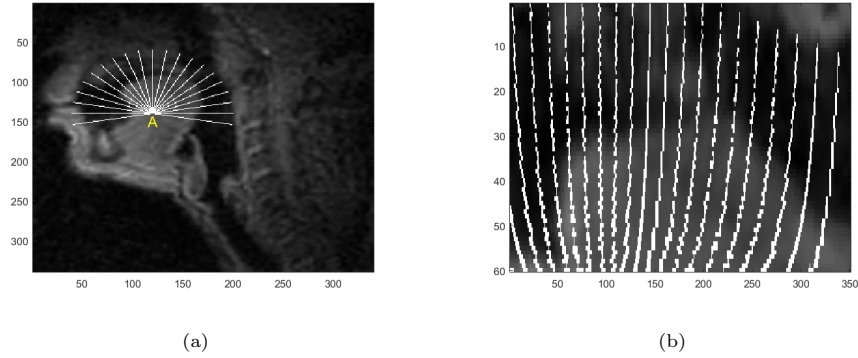
Email addresses: trencsenyi.reka@science.unideb.hu (Trencsényi Réka),
czap@uni-miskolc.hu (Czap László)

módszereinek hatékony fejlesztéséhez (Barnaud et al., 2019). A jelenleg elérhető radiológiai és monitorozó eljárások, mint például a mágneses rezonanciás képalkotás (MRI) (Douros et al., 2020), a komputer tomográfia (CT) (Baum et al., 1990), az ultrahang (UH), az elektropalatográfia (EPG) (Recasens, 1991) vagy az elektromágneses artikulográfia (EMA) (Serrurier et al., 2012) nélkülözhetetlenek az artikuláció dinamikus tulajdonságainak megismerésében. Ez azzal indokolható, hogy a képalkotó technikák segítségével kapott morfológiai és geometriai adatok felhasználhatók az adott beszédjelhez tartozó artikulációs mozgások leképezéséhez, ami alapvető szerepet játszhat például az artikulációt imitáló beszélő fej paraméterezésében. Kutatómunkánk során kvantitatív adatokat származtattunk MRI-képek sorozataiból, ami megfelelő paramétereket biztosíthat az animációs algoritmusunkhoz. Az alkalmazás célja a nyelvmozgások vizuális megjelenítése egy átlátszó arccal ellátott beszélő fej segítségével, ahol az animáció alapelemeit a vizémák képezik. A rendszer jól alkalmazható a beszédterápiában (Czap et al., 2019; Daassi-Gnaba & Krahe, 2009; Zhao et al., 2010), a nem anyanyelvi nyelvoktatás megújításában (Peng et al., 2020; Segaran et al., 2014; Wang et al., 2014) vagy az artikulációs-akusztikai konverziót megvalósító beszéd szintetizátorok konstrukciójában (Csapó et al., 2019; Fagel & Clemens, 2004; Mattheyses & Verhelst, 2015).

Jelen tanulmány célja a nyelvmozgások mélyebb feltérképezését és jellemzését elősegítő kvantitatív adatok meghatározása. A siket és nagyothalló emberek kommunikációjának elengedhetetlen mozzanata a szájról olvasás, de ennek során nem figyelhető meg a nyelv helyzete, alakja és mozgása. A teljes akusztikai percepció hiányában csak a beszéd vizuális modalitására támaszkodhatnak, hogy képesek legyenek a speciális beszédjeleik kialakítására. A kapott kvantitatív adatok stabilan támogatják a transzparens beszélő fej tökéletesítését.

2. Módszerek

A statikus és dinamikus analízis során az MRI-felvételek feldolgozását MATLAB környezetben megírt programjaink felhasználásával végeztük el, melynek



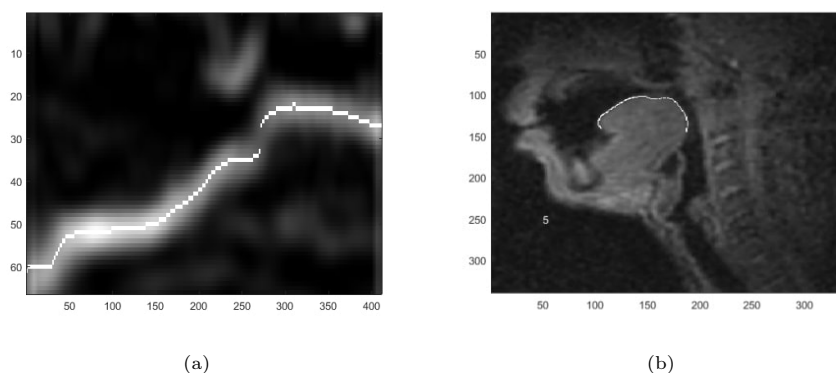
1. ábra. Az előfeldolgozás első lépése: (a) az MRI-kép radiális újra-mintavételezése, (b) az újra-mintavételezett kép descartes-i oszlopmátrixa. A könnyebb áttekinthetőség végett a radiális egyenesek 10° -onként vannak megrajzolva, de az újra-mintavételezés valójában 1° -onként történik.

keretében, dinamikus programozás révén (Czap, 2021), segédgörbét illesztettünk a nyelv felszínére.

Az 1a ábrán látható nyers MRI-képek felbontása 320×320 pixel. Az előfeldolgozás első lépéseként a képet radiálisan újra-mintavételezzük a midszagittális metszet mentén radiális vonalakat képezve egy vizuálisan kijelölt körközéppontból (az 1a ábra A pontja) kiindulva. Erre azért van szükség, hogy elkerüljük a nyelvkontúr visszahajlásából adódó, a kép ugyanazon pixeloszlopában megjelenő, egynél több kontúrponthoz, amit az éldetektáló algoritmus nem tudna kezelni. Az átláthatóság kedvéért az 1a ábra 10° -onként ábrázolja a radiális egyeneseket, de az újra-mintavételezés valójában 1° -onként történik. Az így módon kapott radiális metszeteket descartes-i oszlopokba rendezve egy mátrixstruktúrához jutunk. Ennek megfelelően az újra-mintavételezett képet egy descartes-i koordináta-rendszerben helyezhetjük el, vagyis a radiális geometriát négyzetes geometriává alakíthatjuk át, amit az 1b ábra illusztrál. Az 1b ábrán az 1a ábra transzformált radiális vonalai is megfigyelhetők, az illesztett nyelvkontúrt pedig a 2a ábra jeleníti meg.

A második lépésben, dinamikus programozással végrehajtott élkiemelést követően, megkeressük a kép mátrixában a legnagyobb kumulatív világossággal

rendelkező görbét, amit a 2a ábra példáz. A feldolgozás a kép bal szélső oszlopától a jobb szélső oszlopának irányába halad. Az ily módon azonosított kontúrt a 2a ábra fehér pontjai jelölik ki. Az ezt követő feldolgozást megelőzően az egyenetlen nyelvkontúrt szűréssel simítjuk, így az artikuláció további elemzésének alapjául a simított nyelvkontúr szolgál, megteremtve ezzel a különféle kvantitatív vizsgálatok lehetőségét. Az eredeti keretre visszavetített nyelvkontúr a 2b ábrán figyelhető meg.



2. ábra. Az előfeldolgozás második lépése: (a) a kép descartes-i oszlopmátrixának kiemelt éle, amely fehér pontok sorozataként definiálja a nyelvkontúrt, (b) az eredeti keretre visszavetített nyelvkontúr.

2.1. A nyelv pozíciójának vizsgálata

A nyelvkontúr automatizált kijelölése lehetővé teszi a geometriai jellemzők számítását. Magyar nyelvű felvételek hiányában a Dél-Kaliforniai Egyetem vizuális MRI-adatbázisát (website3) használtuk az egyes beszédhangokhoz kapcsolódó nyelvpozíciók meghatározására egy amerikai angol anyanyelvű, férfi adatközlő által kiejtett magánhangzók és VCV hangkapcsolatok (V: magánhangzó, C: mássalhangzó) tanulmányozásával. Az artikuláció helyének felderítése révén eljutottunk az egyes beszédhangok statikus vizémaadataihoz.

2.1.1. A magánhangzókhoz tartozó nyelvpozíciók meghatározásának kvantitatív módszere

A módszer kiindulópontja a nyelvtest súlypontjának megállapítása az MRI-keretek által adott keresztmetszeti képeken, ami által kvantitatív adatok nyerhetők az aktuális beszédhang nyelvkarakterisztikájának horizontális és vertikális pozícióiról. A nyelv

$$Cxy = [\bar{x}, \bar{y}] \quad (1)$$

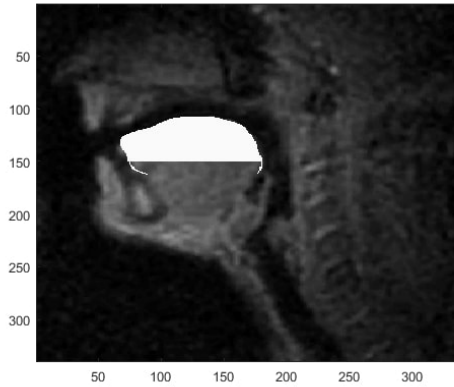
formula által adott súlypontjának x és y koordinátáit a [3] ábrán az e hang esetében prezentált feltöltött nyelvtest fehér pontjainak horizontális és vertikális koordinátaival számított elsőrendű momentumként (Hu [1962] Mukundan & Ramakrishnan [1998]) származtatjuk a

$$\bar{x} = \frac{1}{n} \sum_x \sum_y x \cdot f(x, y), \quad (2)$$

$$\bar{y} = \frac{1}{n} \sum_x \sum_y y \cdot f(x, y) \quad (3)$$

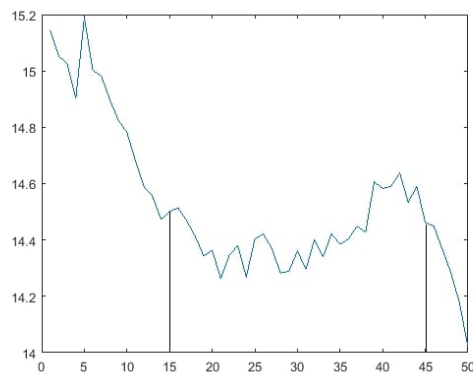
felírás szerint, ahol az f függvény értéke $f(x, y) = 1$ a fehér területen belül, és $f(x, y) = 0$ azon kívül, n pedig az adott tartományt lefedő fehér pontok számát megadó konstans. A nyelvtest feltöltése a nyelvfelszín legfelső pontjából indul, és a [3] ábrával összhangban a kép alsó része felé irányul.

A súlypont kiszámításához meg kell határoznunk a nyelvtestet feltöltő tartomány pixelsorának optimális számát is. Ugyanis, ha túlságosan kevés pixelsort veszünk figyelembe, akkor a nyelv helyzetére vonatkozóan fals és pontatlan mérési adatokat kaphatunk. Ha pedig túlságosan sok pixelsor kerül be a feltöltésbe, akkor a túlsordulás miatt elveszíthetjük a nyelvpozíció valódi jellemzőit. A feltöltési mélység optimalizációjához megvizsgáltuk az említett többnyelvű video-adatbázis 28 magánhangzójához tartozó súlypontok szórását a legnagyobb különbség maximumának megkeresésével. A [4] ábra tanúsága szerint 45-nél több pixelsoros feltöltés esetén a szórásnégyzet meredeken csökken miközben a nyelvgyök szerepe egyre kevésbé hangsúlyos. 15-nél kevesebb pixelsor kije-



3. ábra. A nyelvtest keresztmetszetének feltöltése az ϵ hang esetében

lölése azonban csak a nyelvtestnek a nyelvfelszínhez közeli, felső tartományát reprezentálná, és nem az artikuláció során mérvadó, teljes tömegét. A néhány soros feltöltéskor nagy szórást kapunk, de a legmagasabb pont nem biztos, hogy helyesen tükrözi a nyelv helyzetét.



4. ábra. Az MRI-adatbázis 28 magánhangzójához tartozó súlypontok szórása a feltöltési mélység függvényében

A fentebbi érvelés értelmében a magánhangzók által adott nyelvvalakok súlypontját a $[15, 45]$ intervallum maximumát definiáló 42 pixelsoros feltöltési mélység rögzítésével tanulmányoztuk. A fizikai dimenzióban ez a kijelölés a nyelv

szagittális metszetének felső 22 milliméteres tartományát fedi le, ami a fej függőleges irányú méretének kb. 8-10%-a.

2.1.2. A mássalhangzókhoz tartozó artikulációs helyek meghatározásának kvantitatív módszere

A mássalhangzók artikulációja lényegi különbséget mutat a magánhangzók képzéséhez viszonyítva. Ez az artikuláció helyével írható le, amit az ajak-nyelv-állkapocs mozgás által létrehozott rés vagy zár lokalizál (Erdogan & Wei 2019). Tehát az artikuláció helye a vokális traktus legkisebb szűkületű helyéhez rendelhető.

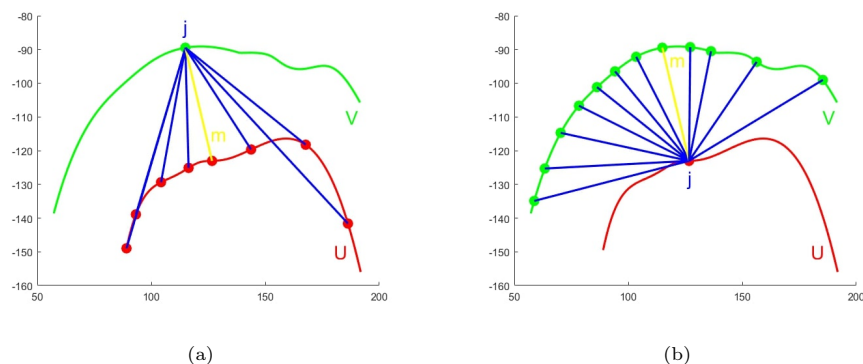
A nyelvkontúr automatikus követését lehetővé tevő algoritmust alapul véve, analóg módon meghatározható a fogmeder és a szájpád kontúrja is. A különbség mindössze abban nyilvánul meg, hogy az MRI-kereteken kiválasztott körközepptől távolodva nem csökkenő, hanem növekvő világosságértéket kell detektálnunk. Emellett az algoritmus paramétereit is a szájpád által adott régió geometriai sajátosságaihoz kell igazítanunk. Azon képek esetében, ahol a szájpád-fogmeder nem rajzolódik ki élesen, és ezáltal bizonytalanná válik a határvonal azonosítása, a kontúrt több kereten végzett átlagolással valószínűsítjük.

A nyelv- és szájpádkontúr ismeretében a két görbe távolsága pontonként kiszámítható például az

$$NND(U, V) = \frac{1}{(n+m)} \left(\sum_{i=1}^n \min_j |u_i - v_j| + \sum_{i=1}^m \min_j |v_i - u_j| \right) \quad (4)$$

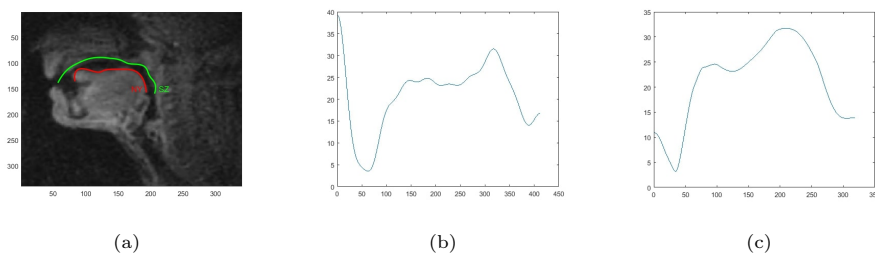
Nearest Neighbour Distance (NND) távolságmérték alkalmazásával, ami kiváltképp alkalmas eltérő számú pontból felépülő görbék távolságának meghatározására (Zharkova & Hewlett 2009). Tegyük fel, hogy a nyelvkontúr az $U = [u_1, u_2, \dots, u_n]$, a szájpádkontúr pedig a $V = [v_1, v_2, \dots, v_m]$ pontok által van kifeszítve. Ekkor a V görbe adott pontjának az U görbétől mért távolsága az U görbe hozzá legközelebb eső pontjától mért távolsággal egyezik meg, amit matematikailag (3) első összegjárulékában szereplő minimum ír le, és grafikusan az 5a ábra szerint értelmezhető. Hasonló gondolatmenetet követve, (3) második

összegjárulékának minimuma a V görbe adott pontjának és az U görbe hozzá legközelebb eső pontjának távolságát adja meg, amit az 5.b ábra szemléltet. Az 5. ábra távolságainak m minimuma sárgával van megjelölve. Gondoljuk meg, hogy a (3) kifejezés két összegjárulékában számított távolságok nem feltétlenül egyeznek meg egy rögzített görbepár esetében.



5. ábra. Az NND távolságmérték grafikus reprezentációja

A 6.a ábra ugyanazon MRI-kereten piros (NY), illetve zöld (SZ) görbékkel megrajzolt nyelv- és szájpadkontúrt demonstrál az r hang esetében. A 6.b ábra a nyelvkontúr pontjainak (vízszintes tengely) szájpadlástól mért távolságát, a 6.c ábra pedig a szájpadkontúr pontjainak (függőleges tengely) nyelvfelszíntől mért távolságát mutatja be. Ebben a megközelítésben, a nyelv legkisebb távolsághoz tartozó pontját tekintjük az artikuláció helyének.



6. ábra. (a) Az illesztett nyelv- (NY) és szájpadkontúr (SZ) az r hang esetében, (b) a nyelvkontúr pontjainak szájpadlástól mért távolsága, (c) a szájpadkontúr pontjainak nyelvfelszíntől mért távolsága.

3. Eredmények

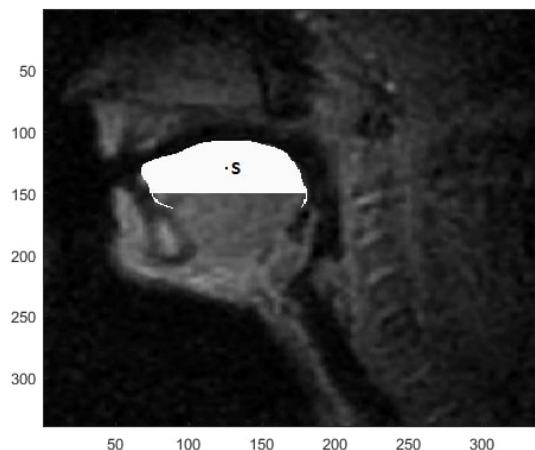
Méréseink során, a nyelv beszéd közbeni mozgását dinamikusan követő nyelvkontúrjaink segítségével olyan geometriai adatokra tettünk szert, melyek alapvető fontosságúak lehetnek az artikuláció kvantitatív jellemzésében. A magánhangzókra és mássalhangzókra kapott eredményeinket elkülönítve tárgyaljuk, párhuzamot vonva a hagyományos leírás fonetikai paramétereivel.

3.1. Magánhangzók

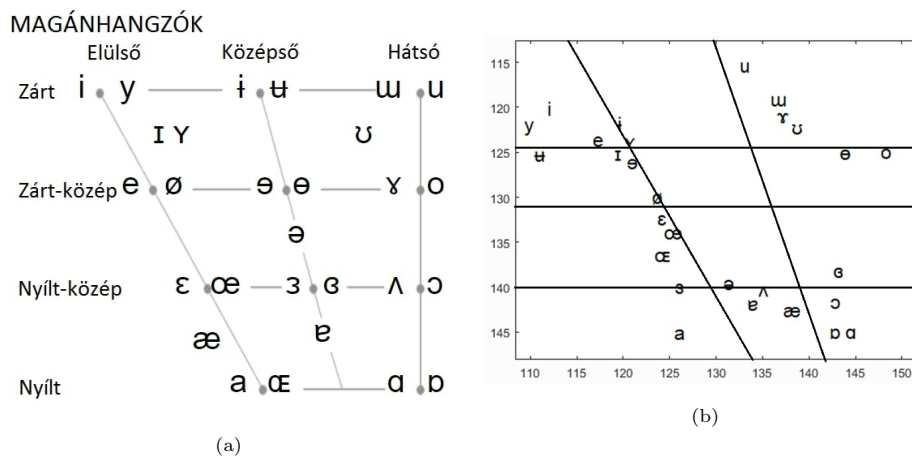
A magánhangzók esetében a vokális traktus longitudinális tengelye mentén képzett keresztmetszeti adatokat az állkapocs nyitottsága és a nyelv helyzete befolyásolja. A vokális traktus keskenyebb és szélesebb metszetei, valamint az ajkak alakja együttesen meghatározzák a gégeből kiinduló gerjesztőjel spektrális tulajdonságait [Ivanova & Hasko \(2019\)](#). A [8](#)a ábra a magánhangzók artikulációjához kapcsolódó nyelvállásokat foglalja rendszerbe a fonetikai paraméterek hagyományos szakirodalmi reprezentációjának megfelelően. Az ábrát az International Phonetic Alphabet (IPA) weboldaláról emeltük át ([website1](#)).

A magánhangzók bemondását megvalósító MRI-felvételeken az adott hang kulcskeretére (a hangátmenetek nélküli, tiszta hangfázis középső kerete) automatikusan illesztett nyelvkontúr által határolt nyelvtartományt a nyelvfelszín legmagasabb pontjától mért 42 pixelsoros mélységgel, majd a 2.1.1. alfejezetben ismertetett módon meghatároztuk a súlypontokat. A súlypont pozícióját (S) az ϵ hang példáján keresztül a [7](#) ábra mutatja be.

A [8](#)b ábra a feltöltött nyelvtest súlypontjait ábrázolja a tanulmányozott 28 magánhangzó esetében az MRI-keretek 320×320 pixeldimenziója által kifesztett koordináta-rendszerben, amely például az [1](#)a vagy [2](#)b ábrákon szemléltetett szájjüreg releváns régióját képezi le. Ennek értelmében a [8](#)b ábra vizuálisan visszaadja a [8](#)a ábra fonetikai elrendezését.



7. ábra. A súlypont pozíciója (S) az ϵ hang esetében



8. ábra. (a) A magánhangzók artikulációs térképe (IPA), (b) az MRI-képeken mért nyelv súlypontok (a szájüreg hátsó része az ábra jobb oldalán, elülső része pedig a bal oldalán található).

A nyelvpozíciók hagyományos artikulációs térképét, – mely vertikálisan négy, horizontálisan pedig három részre osztható – összehasonlítottuk a mért súlyponti adatainkkal. A 8.a és 8.b ábrák összevetése szubjektív úton történt a 8.b ábrára rávetített segédvonalak kijelölésével, melyek létrehozzák a 8.a artikulációs tér-

kép vertikális és horizontális zónáit. Következtetéseinket az [8a](#) táblázat összegzi, ahol az IPA-térképnek megfelelő helyes tartományokba eső nyelvsúlypontokat üres cellák jelzik, a horizontális vagy vertikális irányban eltérést mutató nyelvsúlypontokat pedig a szürke mezők emelik ki. A mezőkben látható 1-es számérték arra utal, hogy az IPA-térkép horizontálisan három-, vertikálisan négylépcsős skáláján maximálisan egy egységnyi eltolódást tapasztalunk a jelzett irány mentén. A [8b](#) ábra vonalait érintő hangokat tekintve a besorolás nem mindig egyértelmű, ezért, ezen hangok pozícióit határesetként kezelve, nem jelöltünk eltolódást az adott irány mentén (pl. e, ɐ, ɪ, ʒ, ʌ). A beszédhangok pozícióinak bizonytalansága, illetve a tapasztalt horizontális és vertikális eltolódások nem rögzíthetők egyértelmű hibaként, hiszen azontúl, hogy az egyes beszélők hangképzése között is mutatkozhatnak különbségek, az IPA-térképek hozzávetőleges táblázatok, melyek az egyes nyelvek beszédhangjait az IPA-táblázat legközelebb eső hangjelével írják le.

A nyelvpozíciók kvantitatív meghatározására vonatkozóan más módszerekkel is találkozhatunk a szakirodalomban, melyek közül érdemes megemlíteni egy kísérleti tanulmányt ([Wang et al., 2013](#)). A jelzett cikk szerzői elektromágneses artikulográfiával (EMA) nyert adatokra támaszkodva állapították meg a nyelv pozícióját 8 angol magánhangzó és 11 mássalhangzó esetében 10 amerikai angol anyanyelvű, női adatközlő bevonásával. A nyelvvalakok meghatározása Prokrusztész-analízis segítségével történt, a kapott adathalmaz osztályozását pedig gépi tanulóalgoritmusok bevetésével végezték el. Az egyes beszédhangokhoz tartozó nyelvpozíciók relatív helyzetét többdimenziós skálázással egy kétdimenziós geometriai térbe leképezve előállítottak egy grafikus reprezentációt, amely összevethető a konvencionális IPA-térképpel. Összehasonlítva az EMA-alapú eredményeket a jelen publikációban tárgyalt MRI-alapú eredményeinkkel, megállapítható, hogy a több beszélő által adott statisztikai EMA-adathalmazzal jó közelítéssel összhangban vannak az MRI-beszélő releváns hangjaira kapott eredményeink, tehát a két technika és a különböző módszerek egy irányba mutatnak.

1. táblázat. A nyelvsúlypontok lokalizációjának pontossága az IPA-térképhez viszonyítva

horizontális	vertikális	horizontális	vertikális
i		o	
y		ə	1
ɨ		ɛ	
ʉ	1	œ	
ɯ		ɜ	1
u		ɔ	1
ɪ		ʌ	1
ʏ		ɔ	1
ʊ		æ	1
e		ɐ	
ø		a	
ɘ		œ	1
ɵ	1	ɑ	
ɤ		ɒ	

3.2. Mássalhangzók

A [2] táblázat a mássalhangzókat a típusuk és képzési helyük szerint osztályozza ([website2]). Munkánk során a bilabiális és labiodentális hangok artikulációs sajátosságait és nyelvéllásait nem vizsgáltuk. Ennek az az oka, hogy ezen hangok képzésekor a nyelv helyzete határozatlan, mivel az aktuális szomszédos hangok nyelvpozícióihoz igazodik. Ebből adódóan a nyelvvalak és a nyelvéllás nem kezelhető az artikulációs effektusok hangkörnyezettől független, hiteles indikátoraként. Ezekben az esetekben az artikuláció helyét kizárólag a fogak és az ajkak határozzák meg.

2. táblázat. A mássalhangzók típusai és képzési helyei

MÁSSALHANGZÓK © 2005 IPA

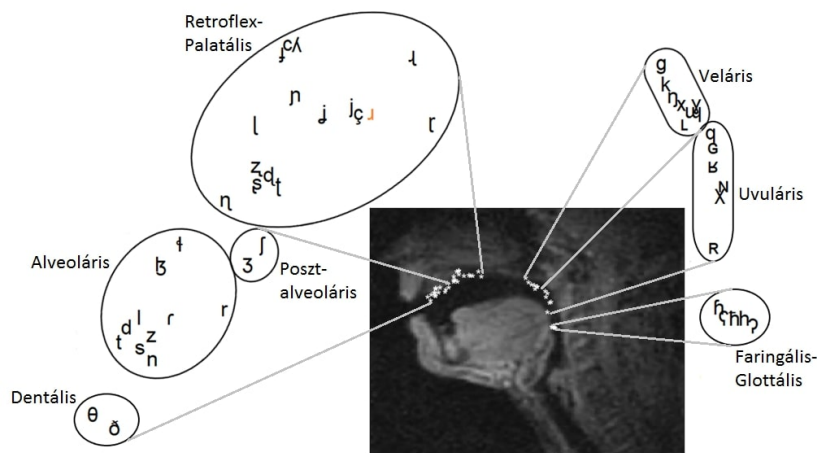
	Bilabiális	Labiodentális	Dentális	Alveoláris	Posztalveoláris	Retroflex	Palatális	Veláris	Uvuláris	Faringális	Glottális
Ploziva	p b		t d			ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nazális	m	ɱ	n			ɳ	ɲ	ŋ	ɴ		
Pergő		β	r						ʀ		
Érintő			ɹ			ɻ					
Frikatíva	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Laterális frikatíva			ɬ ɮ								
Approximáns		ʋ	ɹ			ɻ	j	ɰ			
Laterális approximáns			l			ɭ	ʎ	ʟ			

A 2.1.2. alfejezetben részletezett NND távolságmérték segítségével azonosított képzési helyeket a [9] ábra vizualizálja egy olyan MRI-kereten, ahol jól elkülönül egymástól a nyelv és a szájpád vonala. Megfigyelhető, hogy a [2] táblázatban megadott elméleti és az általunk számított képzési helyek összhangban vannak egymással. Ez alól az egyetlen kivételt az *r* hang képezi (az ábrán pirossal kiemelve), hiszen a felhasznált MRI-felvétel tanúsága szerint az artikuláció helye a szájjüregben hátrébb lokalizálható, mint ahogyan az a [2] táblázat alapján várható. Ez valószínűleg az amerikai angol adatközlő anyanyelvi adottságaiból ered, ami azt eredményezi, hogy az általa kiejtett *r* hang vizuálisan retroflex típusúnak tűnik. Összességében véve tehát elmondható, hogy a kvantitatív úton kapott képzési helyek megegyeznek a fiziológiai meghatározásokkal.

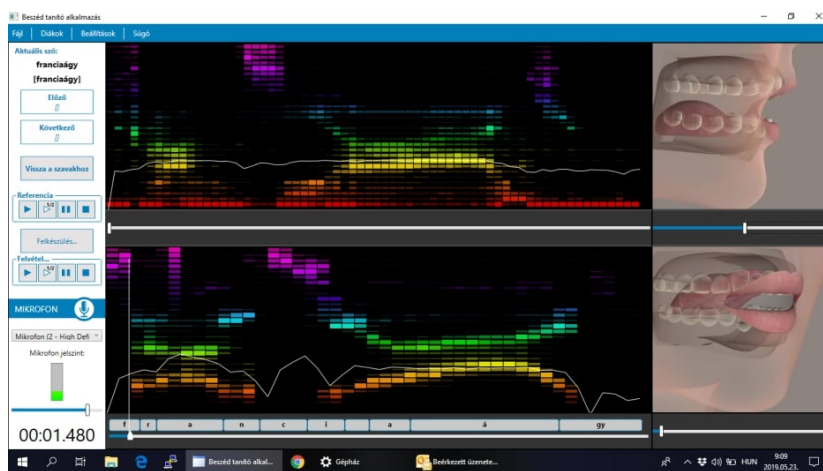
3.3. Az eredmények alkalmazhatósága a Beszédasszisztens rendszerben

A magyar nyelvre kifejlesztett Beszédasszisztens rendszer a siketek és nagyothallók beszédtanulását támogatja audio- és vizuális elemek egyesítésével. A mostani eredményeink a későbbiekben jó támpontot adhatnak és elősegíthetik a Beszédasszisztens rendszer működésének finomítását és továbbfejlesztését is (Czap et al., 2019). A [10] ábra az alkalmazás monitorképét jeleníti meg a hangokhoz tartozó színkódolt absztrakt képi jelek, illetve az átlátszóvá tehető, élethű arcrészlet formájában. A blokk alsó részén a referencia, a felső részén pedig a gyakorló által kiejtett szóhoz rendelt színskála jelenik meg, amely az

adott hang frekvenciakomponenseit és a hangerőt együttesen kódolja. A transzparens fej forgatható, így különböző szögekhez tartozó nézetek állíthatók be. Az animáció során felhasznált paraméterek (pl. a nyelvhegy és a nyelv hát vízszintes és függőleges irányú helyzete, a nyelv visszahúzásának és visszahajlásának mértéke) beállítása reményeink szerint tökéletesíthető a jelen publikációban bemutatott eredményeinkre támaszkodva.



9. ábra. A mássalhangzók számított képzési helyei egy MRI-képen



10. ábra. A Beszédasszisztens gyakorlás közbeni monitorképe

4. Összefoglaló

Az artikuláció fonetikai jellemzőit vizsgáltuk dinamikus MRI-felvételekből származó kvantitatív adatok segítségével. A hangok képzési helyét a nyelv helyzetének megadásával jellemezhetjük, de a magánhangzók és mássalhangzók osztályát különböző módszerrel tanulmányoztuk. A nyelv pozícióját magánhangzók esetében a nyelv súlypontjának, mássalhangzók esetében pedig a nyelv szájpaddalástól mért távolságának segítségével becsültük. A nyelv súlypontját elsörendű momentumként értelmeztük, a nyelv szájpaddalástól mért távolságát pedig az NND távolságmérték alkalmazásával határoztuk meg. Az általunk számított nyelvpozíciók és artikulációs helyek jó egyezést mutattak a beszédhangok tradicionális fonetikai osztályozását vizualizáló IPA-táblázatokkal. Ez azt bizonyítja, hogy az általunk megvalósított megközelítés stabil kvantitatív háttérrel támogatja a kvalitatív fonetikai osztályozást. A kvantitatív elemzések azonban további vizsgálatokat igényelnek például a beszélők számának kiterjesztésével, illetve a különböző hangkörnyezetek által definiált hangátmenetek részleteinek feltérképezésével.

Hivatkozások

- Barnaud, M. L., Schwartz, J. L., Bessière, P., & Diard, J. (2019). Computer simulations of coupled idiosyncrasies in speech perception and speech production with COSMO, a perceptuo-motor Bayesian model of speech communication. *PLoS ONE*, 14, 1.
- Baum, S. R., Blumstein, S. E., Naeser, M. A., & Palumbo, C. L. (1990). Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language*, 39, 33–56. URL: [https://doi.org/10.1016/0093-934X\(90\)90003-Y](https://doi.org/10.1016/0093-934X(90)90003-Y). doi:10.1016/0093-934X(90)90003-Y.
- 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. *Proc. Interspeech*, (p. 894–898). doi:[10.21437/Interspeech.2019-2046](https://doi.org/10.21437/Interspeech.2019-2046).
- Czap, L. (2021). Impact of Preprocessing Features on the Performance of Ultrasound Tongue Contour Tracking, via Dynamic Programming. *Acta Polytechnica Hungarica*, 18, 19. URL: http://acta.uni-obuda.hu/Czap_109.pdf.
- Czap, L., Pintér, J. M., & Baksa-Varga, E. (2019). Features and results of a speech improvement experiment on hard of hearing children. *Speech Communication*, 106, 7–20. URL: <https://doi.org/10.1016/j.specom.2018.11.003>. doi:[10.1016/j.specom.2018.11.003](https://doi.org/10.1016/j.specom.2018.11.003).
- Daassi-Gnaba, H., & Krahe, J. L. (2009). Universal combined system: speech recognition, emotion recognition and talking head for deaf and hard of hearing people. In *Conférence-AAATE* (p. 503–508).
- Douros, I. K., Kulkarni, A., Dourou, C., Xie, Y., Felblinger, J., Isaieva, K., Vuissoz, P., & Laprie, Y. (2020). Using Silence MR Image to Synthesise Dynamic MRI Vocal Tract Data of CV. *Proc. Interspeech*, (p. 3730–3734). doi:[10.21437/Interspeech.2020-1173](https://doi.org/10.21437/Interspeech.2020-1173).
- Erdogan, N., & Wei, M. (2019). Articulatory Phonetics: English Consonants. In N. Erdogan, & M. Wei (Eds.), *Applied Linguistics for Teachers of Culturally and Linguistically Diverse Learners* (p. 263–284). IGI Global. URL: <http://doi:10.4018/978-1-5225-8467-4.ch011>.
- Fagel, S., & Clemens, C. (2004). An articulation model for audiovisual speech synthesis—Determination, adjustment, evaluation. *Speech Communication*, 44, 141–154. URL: <https://doi.org/10.1016/j.specom.2004.10.006>. doi:[10.1016/j.specom.2004.10.006](https://doi.org/10.1016/j.specom.2004.10.006).
- Hu, M. K. (1962). Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, 8, 179–187. URL: <https://doi.org/10.1109/TIT.1962.1057692>.

- Ivanova, S. A., & Hasko, V. (2019). Articulatory Phonetics: English Vowels. In N. Erdogan, & M. Wei (Eds.), *Applied Linguistics for Teachers of Culturally and Linguistically Diverse Learners* (p. 285–301). IGI Global. URL: <http://doi:10.4018/978-1-5225-8467-4.ch012>.
- Mattheyses, W., & Verhelst, W. (2015). Audiovisual speech synthesis: An overview of the state-of-the-art. *Speech Communication*, 66, 182–217. URL: <https://doi.org/10.1016/j.specom.2014.11.001>.
- Mukundan, R., & Ramakrishnan, K. R. (1998). *Moment functions in image analysis*. Singapore: World Scientific Press.
- Peng, X., Chen, H., Wang, L., Tian, F., & Wang, H. (2020). Talking Head-based L2 Pronunciation Training: Impact on Achievement Emotions, Cognitive Load, and Their Relationships with Learning Performance. *International Journal of Human-Computer Interaction*, 36, 1487–1502. URL: <https://doi.org/10.1080/10447318.2020.1752476>.
- Recasens, D. (1991). On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics*, 19, 267–280. URL: [https://doi.org/10.1016/S0095-4470\(19\)30344-4](https://doi.org/10.1016/S0095-4470(19)30344-4).
- Segaran, K., Ali, A. Z. M., & Hoe, T. W. (2014). Usability and user satisfaction of 3D talking-head mobile assisted language learning (MALL) app for non-native speakers. *Procedia-Social and Behavioral Sciences*, 131, 4–10. URL: <https://doi.org/10.1016/j.sbspro.2014.04.069>.
- Serrurier, A., Badin, P., Barney, A., Boë, L. J., & Savariaux, C. (2012). The tongue in speech and feeding: Comparative articulatory modelling. *Journal of Phonetics*, 40, 745–763. URL: <https://doi.org/10.1016/j.wocn.2012.08.001>.
- Wang, J., Green, J. R., Samal, A., & Yunusova, Y. (2013). Articulatory distinctiveness of vowels and consonants: a data-driven approach. *Journal*

- of Speech, Language and Hearing Research*, 56, 1539–1551. URL: [https://doi.org/10.1044/1092-4388\(2013/12-0030\)](https://doi.org/10.1044/1092-4388(2013/12-0030)).
- Wang, X., Hueber, T., & Badin, P. (2014). On the use of an articulatory talking head for second language pronunciation training: the case of Chinese learners of French. In *10th international seminar on speech production (issp 2014)* (p. 449–452).
- website1 (n.d.). URL: <https://www.internationalphoneticalphabet.org/ipa-charts/vowels> Accessed 22.04.2020.
- website2 (n.d.). URL: <https://www.internationalphoneticalphabet.org/ipa-charts/consonants/> Accessed 22.04.2020.
- website3 (n.d.). URL: sail.usc.edu/span/rtmri_ipa/je_2015.html Accessed 15.11.2009.
- Zhao, J., Lirong, W., Chao, Z., Lijuan, S., & Jia, Y. (2010). Pronunciation of rehabilitation methods based on 3d-talking head. In *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering* (p. 17–20).
- Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/. *Journal of Phonetics*, 37, 248–256. URL: <https://doi.org/10.1016/j.wocn.2008.10.005>

End-to-End Recognition of Spontaneous Speech on the Hungarian BEA Database

Tímea Fekete¹, Péter Mihajlik^{1,2}

¹*Budapest University of Technology and Economics*

²*Hungarian Research Centre for Linguistics*

Abstract

The end-to-end deep neural network based speech recognition approach is increasingly popular due to its fully data driven nature - no language-specific knowledge is needed beyond the transcribed speech data. However, most of the end-to-end speech recognition experiments are performed on read speech and no Hungarian language results are available for the Speech Community. In this paper, we make the first attempt to train and evaluate a Hungarian speech recognition system based on the studio-quality Hungarian BEA (Spoken Language Speech Database) in an end-to-end neural manner. We present the challenge of recognising spontaneous speech: even without any significant background noise, the word error rate on spontaneous speech is an order of magnitude higher than in the case of planned speech - both recorded with the same speakers in the same environment. This emphasises the need for more thorough studies of spontaneous speech and possibly for more data.

Keywords: automatic speech recognition, end-to-end neural model, deep learning, Hungarian

1. Introduction

With the spread of artificial neural networks and deep learning, an incredible progress is taking place in numerous fields of technology. Nowadays' best results in automatic speech recognition (ASR) come from end-to-end models, models that are based entirely on deep neural networks. Despite this fact, no previous publication is known to the authors aiming at Hungarian language end-to-end speech recognition on publicly available databases.

As Mihajlik (2020) argues, end-to-end ASR can be considered as a result of natural evolution of ASR systems: statistics and machine learning have been

Email addresses: fe.timea@gmail.com (Tímea Fekete), mihajlik.peter@nytud.hu (Péter Mihajlik)

always essential "ingredients" in an ASR recipe (Bahl et al. 1983). The key step in building the entire ASR on one neural network was the introduction of CTC (Connectionist Temporal Classification) algorithm (Graves et al. 2006), still popular in research and production level ASR systems. This allowed to train acoustic models directly on characters (graphemes), without the need for any pronunciation dictionary or phonological or phonetic knowledge. The approach, however, is mostly applied for isolating languages - like English - where shorter words fit the algorithm more. The effectiveness of the purely CTC-based end-to-end approach is questionable in the case of agglutinating languages where the average word length can be much more. On the other hand, spontaneous speech tends to use shorter words.

Therefore, we decided to train and evaluate an end-to-end deep neural ASR network on the Hungarian language BEA database where we can investigate the challenges of spontaneous speech in contrast to the much more frequent read speech databases.

2. Experiments

One of the crucial elements in the development of automatic speech recognition is data. A publicly available option for English is LibriSpeech, a corpus derived from audiobooks; created specifically for training and evaluating speech recognition systems (Panayotov et al. 2015). It has become one of the most used databases for such purposes, providing a basis of comparison for different models.

BEA is a Hungarian speech corpus developed by the Research Institute for Linguistics in Hungary for various research purposes. This database, unlike LibriSpeech, contains spontaneous speech for the most part, but there are also some other kinds of speech during which the participants repeated sentences or read a text aloud (Gósy et al. 2012) - for simplicity, we will refer to this 'not spontaneous' part of data collectively as 'planned speech'. Most public databases contain read texts exclusively, which makes BEA stand out.

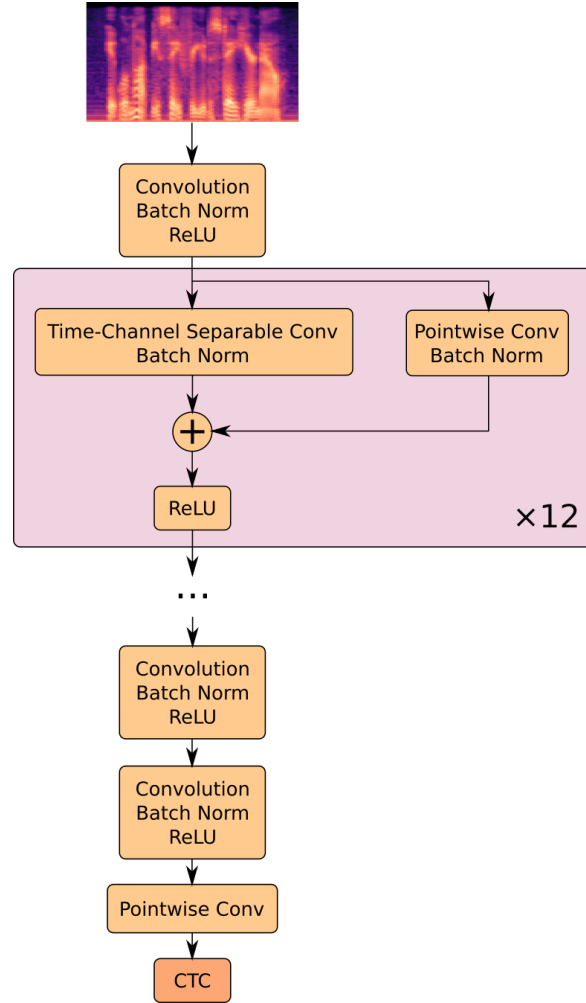


Figure 1: The QuartzNet 12 $\times$ 1 end-to-end deep neural network architecture

We trained an end-to-end ASR model on both LibriSpeech and BEA, enabling a comparison between the two performances. For BEA, the training dataset contained 87.59 hours of speech, the evaluation dataset contained 7.62 hours - distributed similarly to the LibriSpeech dataset in use, which had a 92.34hrs/4.28hrs ratio for the train-clean-100 and test-clean subsets.

We used QuartzNet, a state-of-the-art CTC-based neural acoustic model developed by NVIDIA (Kriman et al., 2019). It is composed of multiple blocks

containing one or more modules with 1D time-separable convolutional layers, batch normalisation and ReLU layers. The blocks connected with residual connections are denoted by B whereas C stands for other (non-residual) blocks. Our model shape is 12×1 , which means that there are 12 internal blocks ($B_1 - B_{12}$), each containing one module with residual connection and this internal structure is completed by leading and trailing C-type blocks ($C_1 - C_4$). The total number of parameters are around 5M. For the details, see Figure 1 and Table 1. Although more complex structures are possible with QuartzNet (which also might bring better results), we had to consider resource-efficiency. There are multiple end-to-end architectures using convolutions, such as wav2letter (Pratap et al., 2019) and Jasper (Li et al., 2019). QuartzNet works with fewer parameters - time-channel separable convolutions use less weights ($\text{kernel size} \times \text{input channels} + \text{input channels} \times \text{output channels}$) than regular 1D channel convolutions ($\text{kernel size} \times \text{input channels} \times \text{output channels}$).

Table 1: The main parameters of the consecutive convolutional blocks of the QuartzNet architecture. There are 36 output labels: the 35 Hungarian characters plus the space character

Block	Kernel size	Output channels
C_1	33	256
$B_1 - B_3$	33	256
$B_4 - B_6$	39	256
$B_7 - B_9$	51	512
$B_{10} - B_{12}$	63	512
C_2	75	512
C_3	1	1024
C_4	1	36 (labels)

To reduce the possibility for overfitting, data augmentation, i.e. adding slightly modified copies of our data was used. In speech recognition, a well-

performing method is SpecAugment, during which three kinds of modifications are applied: time warping (deforming the features through time), frequency channel masking (a block of frequencies removed) and time masking (a block of series removed) (Park et al., 2019).

The minibatch length for training was set to 32 and to 1 for evaluation. A learning rate scheduler was applied with a linear warmup of 5% of training time targeting a maximum value of 0.05, then decreasing dynamically with cosine annealing, keeping a minimum of 0.001. The training lasted 100 epochs. Novograd optimizer (Ginsburg et al., 2020) was used with a weight decay of 10^{-4} and $\beta_1 = 0.95$, $\beta_2 = 0$. During inference, greedy CTC decoding was performed only because we wanted to avoid the application of any external lexicon or language model so that the comparison of planned and spontaneous results are not biased by prior knowledge. All the experiments were performed on an NVIDIA GeForce GTX 1070 GPU.

3. Results

3.1. Quantitative analysis

The most widely used metric for evaluating speech recognition models is WER, standing for Word Error Rate. It shows the rate of errors on a word level using the minimum number of single-word edits as a distance between the reference and hypothesis. The less the value, the better the performance it indicates. Similarly, we can calculate CER (Character Error Rate) on a character level. The later might be even more important for Hungarian, as it's a morphologically rich language.

The training took about three times as many steps for the Hungarian dataset, because it had three times more files than the English one, despite totalling a similar length of time (segments containing only a few words were frequent in the BEA dataset, not so in LibriSpeech). Different results can be expected for the planned and spontaneous parts of the BEA evaluation set, therefore we also evaluated them independently. In Figure 2 we can see how the WER had

changed during the training of the models for both languages (trained with the same parameters).

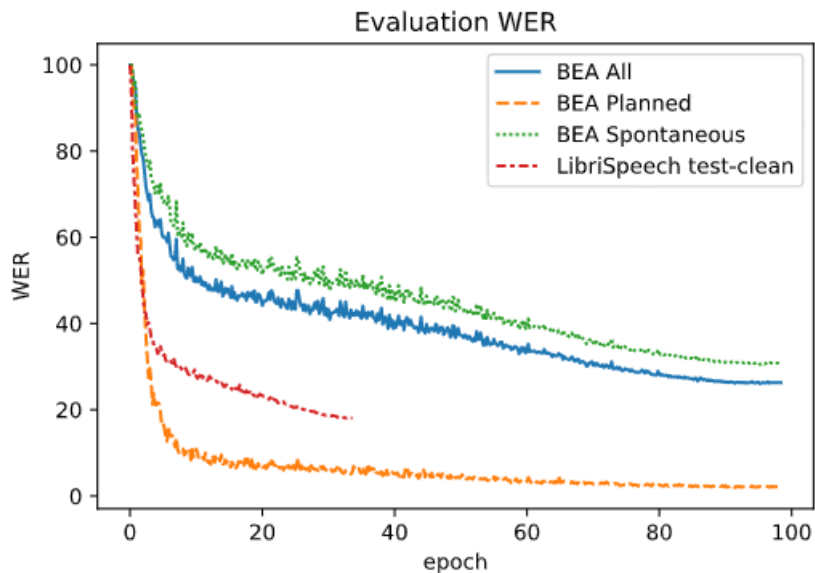


Figure 2: Evaluation WER

Table 2: The model’s performance on different evaluation datasets

	Word count	WER	Char count	CER
BEA Planned	7,679	2.09%	63,701	0.46%
BEA Spontaneous	40,561	30.84%	269,384	8.98%
BEA All	48,240	26.28%	333,085	7.35%
LibriSpeech test-clean	49,957	18.16%	281,530	5.81%

The error rates for the planned subset of BEA got quite close to zero (see Table 2 and Figure 3), due to the nature of the data: the same sentences were repeated by different participants, so it can be expected from the model to recognise these particular sentences almost perfectly. On the other hand, it did not perform so well on the spontaneous subset, even compared to the

model trained on LibriSpeech - spontaneity brings obvious difficulty into speech recognition. Our results aren't necessarily on par with the latest state-of-the-art ones, but note that we trained a less complex QuartzNet architecture on a smaller amount of data (only 100 hours instead of the usual 1000 hours) than most reported models.

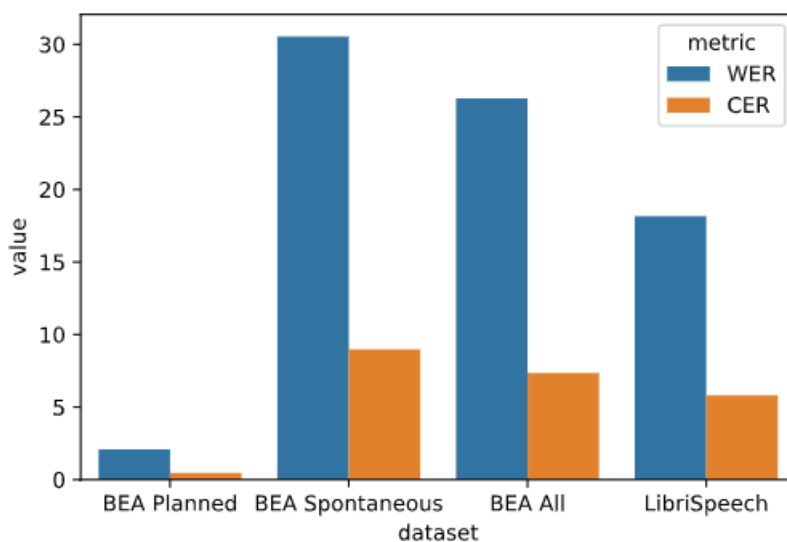


Figure 3: Error rates

3.2. Subjective analysis of spontaneous recognition errors

The magnitude of WER on spontaneous speech looks unacceptably high. However, if we zoom in on some random recognition results on the spontaneous subset, it turns out that many of the errors seem acoustically forgivable - reading aloud some of the hypotheses, one could hardly notice the difference from the reference transcription.

1.	/English/ Reference Hypothesis	/but I really like to cook and bake/ de nagyon szeretek főzni sütni de nagyon szeretek főzni süt néy
----	--------------------------------------	---

Here, similarly to the English term *chutney*, the pronunciation of *süt néy* can be quite close to *sütni*.

2.	/English/	/and the problem and solution must be sought in the family background as well/
	Reference	és a családi háttérben kell keresni a problémát és a megoldást is
	Hypothesis	és a családi áttérben kell keresni a problémát és a megoldást is

The lack of *h* in this case could go unnoticed in spontaneous situations.

3.	/English/	/roots have been grown where they have been planted with a foreign plant/
	Reference	gyökeret növesztettek ahol idegen növény mellé ültették
	Hypothesis	gyökeret növesztettek ahol idegán növény mellé ültették

In this particular case, the cause of recognition error could be due to an accent typical in the Highland region of former Hungary.

4.	/English/	/it is still easier to accept this way but he has beheaded the citizens/
	Reference	mégiscsak egyszerűbb így elfogadni de a polgárokat lefejeztette
	Hypothesis	mégiscsak egyszerűbb úgy el fog adni de a polgárokat lefejeztette

Now we can observe two types of errors. The first one is explainable with a kind of co-articulation of *egyszerű* and *így*. The second one has a semantically false segmentation but there is not necessarily an acoustical difference between *elfogadni* and *el fog adni*. A possible explanation is that the shorter words might have occurred more frequently in the training set.

5.	/English/	/so his hand was very roughly cut apart because he got it in front of his face/
	Reference	hát nagyon durván szét volt vágva a keze mer az arca elé kapta
	Hypothesis	hát nagyon durván szét volt vágal kezdé mer az arc elé kapta

This type of misrecognitions are not easy to explain - we have to admit that the artificial neurons were probably not trained adequately.

6.	/English/	/but if I go down to her let's say she lives very far away just/
	Reference	de hogyha hozzá lemegyek mondjuk ő nagyon messze lakik csak
	Hypothesis	de hogyha hozzále megyek mondjuk ő nagyon messzel akik csak

An illustrative example where the only source of word errors is false segmentation causing a local WER of 70% , whereas the hypothesis is totally correct on the phonetic level.

7.	/English/	/finish the ice cream and even top with a good deal of chocolate so/
	Reference	befejezni a fagyit és még a tetejére is jó sok csokit tehát így
	Hypothesis	befejezni a pajyit és még a teteére is jó sok csokit tehát így

The *pajyit* hypothesis for *fagyit* is odd but not so impossible - the phonetic representation of both *f* and *p* is labial, and the Hungarian speech sound for *gy* is close to the assimilated form of *d+j*. The dropping of *j* from *teteére* would be completely tolerable perceptually if this hypothesis was synthesised.

8.	/English/	/and I actually put myself on these pages to practice English/
	Reference	és tulajdonképpen azért tettem föl magam ezekre az oldalakra hogy gyakoroljam az angolt
	Hypothesis	és csak azért tettem föl magam ezekre zudorkho gyakorolan az angolt

It might look bizarre that a longer word (*tulajdonképpen*) was mistaken for a completely different short one (*csak*). In this case, the speaker was speaking

so fast that only trained ears could understand the other highlighted part as well.

Some errors are due to incorrect word segmentation, non-space characters do match in such cases. This shows that the boundaries of words aren't always clear during spontaneous speech. Other times, a few letters were left out or changed, resulting in nonsense words. Such problems could have been avoided using a language model, which was out of scope from this study. Errors on the planned part of test data hardly ever occurred.

Examples from the model trained on LibriSpeech:

1.	Reference	yes all alone by himself asserted jasper vehemently and winking furiously to the others to stop their laughing he did now truly phronsie
	Hypothesis	yes all alone by himself a serted jasper veemently and winking puriously to the others to stop their laughing he did now truly phronsie
2.	Reference	we sat with the officers some little time after dinner and then went ashore
	Hypothesis	we sat wit the officer some little time after dinner and then wen ashore
3.	Reference	from the norwegian graveyard one looks out over a vast checker board marked off in squares of wheat and corn light and dark dark and light
	Hypothesis	from the norregiond grave yard one looks out over a vast checkerbord marked off end squares of wheat and corn light and dark dark and light

4. Conclusions

We have introduced the first Hungarian language end-to-end speech recognition results on high-quality spontaneous speech. The subset of BEA database used for training and evaluation is available for research purposes. We showed

that ASR error metrics are significantly higher for spontaneous speech than for the planned one even in the case of identical speakers in the same environment. The results suggest that spontaneous speech needs special care in modelling and/or more data in order to cope with the high variability of spontaneously articulated or fast speech. The validity of the findings is limited though by the fact that the planned part of the database is based on the same text for each speaker. As for future work, we are going to re-sample the planned part of the corpora to decrease the mismatch of planned/spontaneous evaluation. Also, we plan to add a language model to the model/decoder since it obviously could reduce the word error rates according to the results of subjective evaluation. All in all, the spontaneous results can serve as baseline for future improvements.

References

- Bahl, L. R., Jelinek, F., & Mercer, R. L. (1983). A maximum likelihood approach to continuous speech recognition. *IEEE transactions on pattern analysis and machine intelligence*, 5(2), 179–190.
- Ginsburg, B., Castonguay, P., Hrinchuk, O., Kuchaiev, O., Lavrukhin, V., Leary, R., Li, J., Nguyen, H., Zhang, Y., & Cohen, J. M. (2020). Stochastic gradient methods with layer-wise adaptive moments for training of deep networks. *arXiv*, 1905.11286.
- Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006). Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning* (pp. 369–376).
- Gósy, M., Gyarmathy, D., Horváth, V., Grácz, T. E., Beke, A., Neuberger, T., & Nikléczi, P. (2012). BEA: Beszéltnyelvi adatbázis. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* (pp. 9–25). Budapest: Akadémiai Kiadó.

- Kriman, S., Beliaev, S., Ginsburg, B., Huang, J., Kuchaiev, O., Lavrukhin, V., Leary, R., Li, J., & Zhang, Y. (2019). QuartzNet: Deep automatic speech recognition with 1D time-channel separable convolutions. *arXiv*, 1910.10261.
- Li, J., Lavrukhin, V., Ginsburg, B., Leary, R., Kuchaiev, O., Cohen, J. M., Nguyen, H., & Gadde, R. T. (2019). Jasper: An end-to-end convolutional neural acoustic model. *arXiv*, 1904.03288.
- Mihajlik, P. (2020). How does an AI recognize speech?—about end-to-end deep neural network based speech recognition. In *Speech Research conference* (pp. 68–70). volume 14.
- Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). LibriSpeech: An asr corpus based on public domain audio books. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5206–5210). doi:[10.1109/ICASSP.2015.7178964](https://doi.org/10.1109/ICASSP.2015.7178964).
- Park, D. S., Chan, W., Zhang, Y., Chiu, C.-C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). SpecAugment: A simple data augmentation method for automatic speech recognition. *Interspeech*, 2019. URL: <http://dx.doi.org/10.21437/Interspeech.2019-2680>.
- Pratap, V., Hannun, A., Xu, Q., Cai, J., Kahn, J., Synnaeve, G., Liptchinsky, V., & Collobert, R. (2019). Wav2letter++: A fast open-source speech recognition system. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019. URL: <http://dx.doi.org/10.1109/ICASSP.2019.8683535>. doi:[10.1109/icassp.2019.8683535](https://doi.org/10.1109/icassp.2019.8683535).