BESZÉDTUDOMÁNY – SPEECH SCIENCE

2024 (4)/1. szám

Szerkesztők/Editors:

Gráczi, Tekla Etelka

Mády, Katalin

HUN-REN Nyelvtudományi Kutatóközpont

HUN-REN Hungarian Research Centre for Linguistics

Budapest

Technikai szerkesztés/Typesetting: Garai, Luca

Borítóterv/Cover design: Iuga, Laura ©

Korrektúra/Proofreading: Jankovics, Julianna & Hegyi, Flóra

A folyóiratszám kiadását a Magyar Kutatási Hálózat (HUN-REN) támogatta.
This volume was supported by the Hungarian Research Network (HUN-REN).

Szerkesztői előszó

Kedves Kollégák!

A korábbi *Beszédkutatás* folyóirat 2020 óta új néven jelenik meg,*Beszédtudomány–Speech Science* címmel. Célunk a beszédtudomány különböző területeiről érkező kutatások ismertetése, emellett a nemzetközi publikációk növelése. A jelen előszóban a kötetben található tanulmányok tematikájának ismertetése előtt egy gyászhírt és néhány változást jelentünk be.

***

Búcsúzunk *Csapó Tamás Gábor*tól (1985–2024). Tamás tragikus hirtelenséggel hunyt el 2024 januárjában. Személyében nemcsak egy kiváló kutatót, hanem egy közvetlen, jóindulatú, szeretetteljes, segítőkész embert veszítettünk el. Mindazt az emberi jót, őszinteséget és szakmai kitartást, hozzáállást, amit kaptunk és tanultunk tőle, megőrizzük emlékével.

***

A folyóiratban három változás következett be.

2024-től a folyóirat szerkesztősége átalakult. Hálásan köszönjük Gyarmathy Dorottyának, Horváth Viktóriának és Krepsz Valériának az eddigi közös munkát, a rengeteg beletett energiát, ötletet. A folyóirat újjáalakításában, gondozásában, jövőjének megalapozásában végzett munkájuk felbecsülhetetlen.

A folyóiratszámokban a továbbiakban nem csak hagyományos értelemben vett tanulmányokat várunk közlésre, hanem rövid tanulmányokat (előtanulmány, esettanulmány, negatív vagy ellentmondásos eredmények) és módszertani írásokat (módszertani leírások és tutoriálok) is. A szerzői útmutató az alábbi linken érhető el mindhárom típusú kézirat benyújtásához: `https://ojs.mtak.hu/index.php/besztud/information/authors`

További változás, hogy a folyóirat évi két számmal fog megjelenni, amelyekre a kéziratok benyújtási határidejét rögzítjük, azaz a felhívás dátuma nem változik

évente. A 2025-ös számtól az alábbiak szerint várjuk a benyújtást: A következő évi két számba szánt kéziratok beküldésének határideje 2025. január 15. és május 15. A kéziratok továbbra is átesnek a szokásos lektorálási folyamaton, így ha az átdolgozás miatt csúszás áll fenn egy elfogadott kéziratban, a soron következő számban tesszük közzé.

A kéziratokat továbbiakban is a beszédtudomány bármely területéről várjuk, például artikuláció, akusztikum és percepció; beszédtechnológia, beszédfelismerés, beszédszintézis, 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 kérdések. A beküldés és a lektorálási folyamat a `https://ojs.mtak.hu/index.php/besztud/index` oldalon keresztül történik. Minden kéziratot két független, a szerzőkétől eltérő kutatócsoporthoz tartozó bíráló véleményez kettős-vak eljárással. A részletek a fentebb írt szerzői útmutatóban megtalálhatóak. Kérjük, ügyeljenek arra, hogy a kézirat első benyújtásakor minden szerzőjét regisztrálják és rendeljék hozzá a benyújtáshoz.

***

A jelen kötet tanulmányai ismét felölelik a beszédtudomány számos területét. Szegmentális fonetikai elemzés olvasható a hiátustöltésről, a magyar prozódiára gyakorolt idegen nyelvi hatás feladatvégzés közbeni változásáról, diskurzusjelölők elemzéséről, a dysarthria különböző típusainak akusztikai elkülöníthetőségéről, enyhe értelmifogyatékosok alapfrekvencia-jellemzőiről találhatóak tanulmányok a 2024/1-es számban. Mindezek mellett a módszertani írások között a beszédhallgatás során megjelenő agyi jel dekódolásáról és a beszédadatbázisok leiratozási munkálatairól szóló közleményeket találhatnak.

Üdvözlettel a szerkesztők,

Gráczi Tekla Etelka és *Mády Katalin*

Editorial foreword

Dear Colleagues,

The former journal *Beszédkutatás* (Speech Research) was renamed into *Beszédtudomány – Speech Science* in 2020. One of our goals is to increase the number of international publications, as is signalled by the journal's bilingual title. Before introducing the papers published in the present issue, we have to publish an obituary and share some news on the journal with you.

***

Tamás Gábor Csapó (1985–2024) passed away suddenly and tragically in January 2024. In Tamás, we lost not only an outstanding researcher but also a warm, kind-hearted, loving, and helpful person. We will cherish his memory, holding onto all the goodness, sincerity, professional dedication, and attitude that we received and learned from him. May Tamás rest in peace.

***

We wish to inform you about three changes regarding the journal.

The editors' team has been reorganised in 2024. We are grateful to Dorottya Gyarmathy, Viktória Horváth, and Valéria Krepsz for the joint work, the enormous amount of their energy and thoughts they invested into this journal. Their work on the reformation, the editorial work, and the foundation of the future of the journal is inevitable.

From now on, we not only accept traditional research papers, but we also welcome short papers (pilot and case studies, negative or contradictory results), and methods papers (descriptions of methodology and tutorials). Whichever type of submission you are considering, please follow the authors' guide at: `https://ojs.mtak.hu/index.php/besztud/information/authors`

Another change is that the journal will be published in two issues per year. The deadline for manuscript submission is fixed from the next issues henceforward. For the two issues of 2025, the submission deadlines are the 15^{th} of

January, and for the second issue, the 15$^{\text{th}}$ of May. All manuscripts undergo a double-blind peer-reviewing process. Therefore, if the revision is delayed, the manuscript will be published in the next issue in case of acceptance.

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; bi- and multilingualism; prosody, syntax; pragmatic aspects; clinical research, speech and language disorders; development of corpora and databases; disharmonic speech phenomena and other research questions connected to speech characteristics, processing, and production. The language of the submissions is English or Hungarian. The submission and the review process are managed via `https://ojs.mtak.hu/index.php/besztud/index`. All manuscripts are reviewed by two independent researchers who are not part of the same research group as the authors. Instructions for submissions can be found under the url above. Authors must not forget to register all coauthors and to add them to the submission.

***

The present issue includes papers on various fields of speech sciences. A segmental phonetic study on hiatus-filling, one on the changing effect of a foreign language on L1 prosody in a bilingual experiment can be read. Another study introduces new results on discourse markers, and one observes the possibility of acoustic differentiation between various types of dysarthria. In the first methodological section of the journal, you can read about decoding brain signal during listening to speech, and on speech database annotation works.

Sincerely, the editors,

Tekla Etelka Gráczi and *Katalin Mády*

Tartalomjegyzék/Table of contents

Az *iá* és az *ijá* vokalikus hangsorok megvalósulása magyar álszavakban a nyelvállás akusztikai vetületének szempontjából

Juhász Kornélia[1,2,3], Deme Andrea[1,3]

[1] *Eötvös Loránd Tudományegyetem*
[2] *HUN–REN Nyelvtudományi Kutatóközpont*
[3] *MTA–HUN-REN NYTK Lendület Neurofonetikai Kutatócsoport*

Abstract

In this acoustic analysis we compare the realization of *iá* /ia:/ and *ijá* /ija:/ in Hungarian pseudowords. We expect that the orthographical representation induces contrast between these forms in the phonetic realisation, more particularly, between the [j] that is not present in the orthographical representation of the pseudoword (e.g., in *iá* /ia:/) and the [j] that is present in orthography (e.g., in *ijá* /ija:/). We suggest that the investigation of these realisations may serve as a basis for future analyses where i) epenthetic [j] appearing in hiatus, and ii) [j] present in the assumed phonological representation of a word are compared, since *j* is never marked in hiatus by orthography. We propose that through orthographic facilitation, the setting of the present study forces speakers to maximally exaggerate any possible phonetic contrasts between marked and non-marked [j]-realisations (in otherwise identical phonetic contexts), and this is analogous to the phonemic and non-phonemic *j*. Therefore, the present study can clarify if any difference may be expected in the comparison of phonemic and non-phonemic [j]. We analyse the acoustic traits of tongue height differences between the two [j] realizations in /ia:/ and /ija:/ sequences. The phonemic /j/ is claimed to be an approximant, and a liquid, and thus is characterized by more constricted vocal tract than, e.g., the high vowel /i/. The epenthetic [j] in hiatus resolution is, however, considered to be a glide which – from a phonetic viewpoint – is the result of the acoustic transition between the articulatory/acoustic targets of /i/ and /a:/. On this basis, we expect that the epenthetic [j] in the sequence *iá* /ia:/ is articulated with a less constricted (more vowel like) vocal tract than that observable in the realisation of the phonemic /j/ in the sequence *ijá* /ija:/. To test this, we analyse the acoustic traits of tongue height differences between the two [j] realizations in *iá* /ia:/ and *ijá* /ija:/ sequences, that is, we measure and analyse F_1. We expect that /j/ in *ijá* /ija:/ features a narrower constriction in the oral cavity reflected in lower F_1, than *iá* /ia:/. We recorded [j] realizations in /ia:/, and /ija:/ shaped vocalic sequences in nonsense words in two sibilant contexts produced in isolation by 14 Hungarian female speakers. We extracted F_1 frequencies automatically at every 5^{th} ms throughout the whole quasiperiodic signal phase in Praat. The resulting F_1 curves were submitted to GAMMs, where we analysed the effect of the normalized timepoint predictor on the dependent

Email addresses: juhasz.kornelia@nytud.hun-ren.hu (Juhász Kornélia), deme.andrea@btk.elte.hu (Deme Andrea)

Beszédtudomány – Speech Science 2024/1. 9–41. DOI: 10.15775/Besztud.2024.1.9-41

variable, F_1, and added vocalic sequence as a parametric term to each model, as well as random smooth by each trajectory. Our results showed that regardless of the sibilant context, there was a significant difference between *iá* /ia:/ and *ijá* /ija:/ in the transitional phase connecting the two targets (/i/ and /a:/), since /ija:/ showed lower F_1 than /ia:/, which reflects a narrower constriction in the oral cavity in *ijá* /ija:/. Therefore, we concluded that speakers may differentiate [j] variants that are marked or not marked in orthography, and it is possible that they apply this differentiation when producing phonemic and epenthetic [j] that surfaces in the case of hiatus resolution.

---

1. Bevezetés

Ebben az akusztikai fonetikai vizsgálatban azt elemezzük, hogy eltérhet-e az *iá* /ia:/ és az *ijá* /ija:/ megvalósítása abban az esetben, amikor az ortográfiai megjelenítés és a kísérleti elrendezés egyaránt a két vokalikus hangsor közötti kontraszt megvalósítását facilitálja, tehát azt, hogy a beszélők próbálják tudatosan megjeleníteni a [j] hangot a hangsorban (*ijá* /ija:/), vagy nem (*iá* /ia:/). A vizsgált álszavak ortográfiai megjelenítése egy olyan – a magyar nyelvben nem kontrasztív – szembenállás produkciójára igyekszik rákényszeríteni a beszélőket, amely analóg egy a spontán beszédben megjelenő, jelentéssel rendelkező szavakat érintő esettel: a szakirodalom szerint két különböző [j]-realizáció megjelenése feltételezhető a fonemikusan is [j]-t tartalmazó hangsorokban (pl. *nyá**j**ig*) és a hiátustöltő [j]-t tartalmazó hangsorokban (pl. *fáig*). Az analógia alapját az képezi, hogy a hiátustöltőként megjelenő [j]-t a helyesírás sosem jelöli – bár az is valószínűsíthető, hogy az írásban nem jelölt [j] nem okvetlenül elemezhető mindig hiátustöltőként (vö. pl. Siptár, 2011). A jelen kísérlet célja az, hogy minden eszközzel (az ortográfiai megjelenítéssel, illetve egymás után ejtendő, lényegében minimális párként jelentkező hangsorokban) a lehető leginkább facilitáljuk a [j]-realizációk potenciális akusztikai elkülönítését. Arra a kérdésre keressük a választ, hogy egyáltalán képesek lehetnek-e a beszélők álszavak esetében elkülöníteni ezt a valószínűsíthetően kétféle [j]-realizációt tartalmazó hangsort akkor, ha minden körülmény erre az elkülönítésre vezeti őket, illetve amikor erre a (fonetikai) körülmények a legideálisabbak. Az eredmények megalapozzák a valódi, potenciálisan fonemikus, illetve hiátustöltő [j]-t tartalmazó szavak vizsgálatát, hiszen ott eltérés csak akkor várható, ha a je-

len elemzésben is találunk különbségeket – ilyen jellegű megalapozó vizsgálatot a szerzők tudomása szerint korábban nem végeztek. A jelen tanulmányban a [j]-megvalósulásokat az egymástól akusztikai és artikulációs tekintetben is legmesszebb képzett palatális legfelső nyelvállású /i/ és centrális legalsó nyelvállású /aː/ magánhangzók esetében vizsgáljuk dentialveoláris /s/ és posztalveoláris /ʃ/ mássalhangzók kontextusában. A bemutatott kísérlet újszerűsége abban (is) rejlik, hogy a fókuszában egy dinamikus elemzés áll: míg a megelőző, a fonemikus és hiátustöltő [j] megvalósulásának elemzésére irányuló akusztikai vizsgálatok a két [j]-realizáció lehetséges eltéréseit – a szerzők tudomása szerint – kizárólag statikus szempontból vizsgálták, a jelen vizsgálatban az első formáns frekvenciájának változását elemezzük. Ebből a szájüregben létrejövő szűkület mértékére következtethetünk, tehát közvetetten arra, hogy a kérdéses beszédhangok egyike, a helyesírásban is jelölt [j] (ami a fonemikusan megjelenő [j] esetével analóg) mássalhangzósabb-e a helyesírásban nem jelölt [j]-nél (ami a hiátustöltőként megjelenő [j]-vel áll analógiában). A mássalhangzósabb ejtés ugyanis kisebb átmérőjű szűkülettel jár együtt. Mivel a vizsgálatunk a fonemikusan és hiátustöltőként megjelenő [j] összevetését alapozza meg, illetve ezzel analóg helyzeteket elemez, ezért először tehát azt kell számba vennünk, hogy a fonológiai és fonetikai szakirodalom eddig milyen következtetéseket összegzett az említett két [j]-realizáció fonológiai és fonetikai tulajdonságaival kapcsolatban.

A továbbiakban elsőként a fonemikus [j]-realizációk nyelvi viselkedését és kategorizációját ismertetjük, és ezután kerítünk sort a hiátustöltő [j] megjelenésének jellemzésére, illetve a megjelenése mögött megbúvó fonológiai folyamatok ismertetésére. Ezután bemutatjuk ezen fonológiai jegyekkel és folyamatokkal összefüggésben álló artikulációs-akusztikai tulajdonságokat, valamint a kétfajta [j]-realizációval kapcsolatos korábbi empirikus eredményeket, amik a jelen vizsgálat hipotéziseihez vezettek el.

1.1. Háttér

Fonológiai szempontból a mögöttes /j/ fonéma felszíni megvalósulása, illetve viselkedése Siptár (2013), illetve Siptár & Törkenczy (2000) szerint viszonylag

heterogén. Egyfelől a /j/ mássalhangzós tulajdonságokkal rendelkezik, hiszen vannak obstruens allofónjai (szóvégi esetében, pl. *kapj* [kɒpç] vagy *dobj* [dobɟ]), a /j/-kezdetű szavak előtt a határozott névelő a mássalhangzók előtt jellemző *a* alakban jelenik meg, és a /j/-végű szavak a *-val/-vel* toldalékkal ugyancsak a mássalhangzós, a *v*-t hasonító megjelenést mutatják (pl. *vajjal*) (Siptár & Törkenczy, 2000, 85). A /j/ ugyancsak lehet asszimiláció kiváltója, illetve elszenvedője is. Az asszimiláció kiváltójaként az *l*-palatalizációban a /j/ képzési jegyei átterjednek a szomszédos laterális /l/-re (pl. *alja* [ɒjːɒ]) (Siptár & Törkenczy, 2000, 178), míg az asszimiláció elszenvedőjeként a /j/ felszólító módú igealakok esetében posztlexikálisan asszimilálódik a szibilánsvégű igetőhöz (pl. *moss*) (Siptár, 2016). Habár ezek alapján a tulajdonságok alapján úgy tűnik, hogy a /j/ mássalhangzó, viszont egyértelműen nem obstruens, hiszen például nem vesz részt zöngésségi hasonulásban sem annak környezeteként (pl. *ajtó* *[ɒçtoː])[1], sem kiváltójaként (*fáklya* *[faːgjɒ]). Valamint fonetikai szempontból tekintve a képzése is közelítőhangszerű (approximáns), azaz az ejtését nem kíséri a réshangokra jellemző aperiodikus turbulens zörej. Mindenezek alapján Siptár (2013), illetve Siptár & Törkenczy (2000) a /j/ alapvariánsát likvidaként [+msh, +szon] határozza meg, mégpedig azért, mert az /r/-hez és az /l/-hez hasonlóan a /j/ is részt vesz olyan folyamatokban, mint a nazálisokhoz történő képzés hely szerinti asszimiláció, vagy a likvidatörlés. A nazálisokhoz történő asszimiláció esetében a nazálist követő likvida teljesen hasonulhat a megelőző beszédhanghoz spontán beszédben, pl. *olyan jó* [ojɒjoː], *olyan rossz* [ojɒrosː], *olyan lassú* [ojɒlːɒʃːu] (Siptár & Törkenczy, 2000, 209–210; Siptár, 2013). A likvidatörlés esetében pedig – szintén az /r/-hez és az /l/-hez hasonlóan (pl. *arra* [ɒːrɒ], *merre* [mɛːrɛ]; és *balra* % [bɒːrɒ]) – a /j/ is pótlónyúlással törlődik (pl. *mélység* [meːseːg], *éjszaka* [eːsɒkɒ]) (Siptár & Törkenczy, 2000, 212–213; Siptár, 2013). Mindezek alapján a jelen megalapozó jellegű vizsgálatunkban abból indulunk

[1] Az átiratok, szóalakok előtti csillag (*) és százalékjel (%) rendre arra utal, hogy az adott formát a beszélőközösség egésze vagy egy része elutasítja. Ezek a jelölések a jelen szövegben a hivatkozott forrásból idézett, nem revideált átvételek.

ki, hogy a /j/ fonéma, illetve annak a felszínen megjelenő alapvariánsa likvida, tehát egy dominánsan mássalhangzós tulajdonságokkal rendelkező beszédhang.

A hiátustöltőként megjelenő [j]-realizációt illetően először azt kell definiálnunk röviden, hogy mi is az a hiátus és miért jön létre hiátustöltés. A hiátus olyan magánhangzó-kapcsolatot jelöl, ahol a szekvencia két vokalikus eleme két különböző szótag szótagmagi magánhangzójának tekinthető; a hiátus jelentése 'hangűr', és a magánhangzók közül hiányzó mássalhangzóra (ill. a hiányra magára) utal (Siptár, 2002; Markó, 2012; Gósy, 2014). Az ilyen módon létrejövő heteroszillabikus magánhangzó-kapcsolat (V_1V_2), azaz hiátus előfordulását sok nyelv nem tekinti jólformáltnak, ezért számos nyelvspecifikus stratégia létezik annak feloldására. Ilyen stratégia például a) a magánhangzó-törlés, b) a magánhangzó-kapcsolat első tagjának (V_1) átalakítása siklóhanggá (aminek következtében a siklóhang a szótag onszetjévé válik), c) a két magánhangzó összeolvadása pótlónyúlással, d) diftongusképzés (aminek révén a V_1 és V_2 egyetlen szótagban helyezkedő kettőshangzóvá változik), illetve e) egy mássalhangzó beszúrása (Siptár, 2005; Casali, 2011; Gósy, 2014). Ezen stratégiák mind a hiátushelyzet megszüntetésére törekednek, azonban kérdésként merülhet fel, hogy miért van szükség egyáltalán a hiátushelyzet feloldására? A hiátus feloldásának kérdését gyakran fonotaktikai szempontból szokták magyarázni, mégpedig azzal, hogy a VV-kapcsolat (és az onszet nélküli szótag) jelölt, míg a magánhangzókapcsolat feloldásával egy kanonikus, jelöletlen CV-szerkezetű szótag alakul ki (Brown, 1970; Pulleyblank, 1986; Balogné Bérces, 2006; Casali, 2011). Ehhez hasonló Haas (1988) magyarázata is, amely szerint a hiátustöltés a szótagok „rossz kapcsolódásából" fakad, ahol a magánhangzó-kapcsolat miatt szonoritási összeférhetetlenség keletkezik. Az érvelés szerint a szonoritási, azaz hangzósági összeférhetetlenséget az okozza, hogy a szomszédos magánhangzók általában megközelítőleg hasonló mértékben hangzósak, de a fonotaktikai szabályok megkövetelnének egyfajta hangzósságbeli „visszaesést" a szótagok között. (Ez a magyar nyelvben is megfigyelhető: itt a szótagmag a legmagasabb szonoritással rendelkező elem a szótagban, és ettől mindkét irányban távolodva egyre csökken a szegmentumok hangzóssága, vö. Siptár & Törkenczy, 2000, 107-110). Egy má-

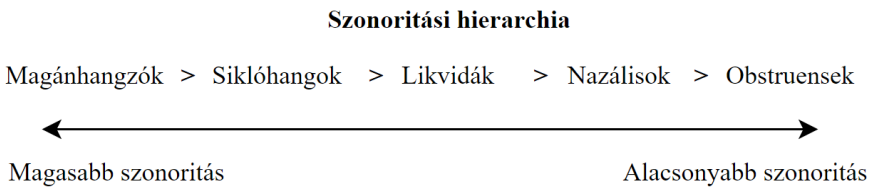
sik elgondolás szerint a hiátus feloldását nem a CV-szótagstruktúra kialakítása indukálja, hanem az, hogy a nyelvben alapvetően is elkerüljük a magánhangzó-kapcsolatokat azért, mert a két magánhangzó koartikulációs egymásra hatása mindkét magánhangzó minőségét befolyásolja, és így akadályozhatja az azonosításukat (Casali, 2011).

A jelen tanulmányban az előzőekben ismertetett stratégiák közül relevanciájánál fogva az utolsóval foglalkozunk: a mássalhangzó, egészen pontosan a [j] betoldásával. Picard (2003) és Uffmann (2007) szerint a hiátus kitöltése a leggyakrabban egy olyan siklóhang (pl. [j]) beszúrásával történik, amely homorgán vagy a V_1-gyel vagy a V_2-vel (pl. [i]). Egy homorgán siklóhang beszúrása gazdaságosabbnak és hatékonyabbnak tekinthető, mint egy eltérő képzési helyű szegmentum beszúrása, hiszen ezáltal megőrződnek a mögöttes magánhangzó-fonémák képzési jegyei: mivel a hiátustöltő [j] az [i]-vel lényegében azonos képzési jegyekkel rendelkezik, ezáltal a homorgán siklóhang beszúrása tulajdonképpen a meglévő fonológiai információ (jegy)terjedéseként értelmezhető (Uffmann, 2007, 458; Siptár & Törkenczy, 2000, 285).

A magyar nyelvben akkor, ha a heteroszillabikus magánhangzókapcsolat egyik (de nem mindkettő) eleme [i], ez a hiátus egy, az [i]-vel homorgánnak tekinthető [j] siklóhanggal töltődik ki. Siptár (2013), továbbá Siptár & Törkenczy (2000) elemzése szerint ez a [j] – a fentieknek megfelelően – a magyar nyelvben is siklóhang, azaz [–msh, +szon] jegyekkel rendelkezik, és abban az esetben, ha a magyar nyelvben olyan magánhangzó-kapcsolat jelenik meg, amelynek egyik eleme /i/, akkor a magas nyelvállású magánhangzó jegykötege átterjed a szomszédos szótag onszetjére. Ez Siptár & Törkenczy (2000, 285) *fiú* példája alapján azt jelenti, hogy az /i/ magánhangzó-csomópontja jobbra terjed, és hozzá tartozóan az időzítés tengelyén létrehoz egy plusz pozíciót, illetve asszociál hozzá egy onszet csomópontot is a következő szótag elején.

Az eddig leírtakat összegezve tehát a fonológiai szakirodalom szerint amíg a fonemikus /j/ realizációja a magyarban mássalhangzósabb természetű likvida [+msh, +szon], addig a hiátustöltőként megvalósuló [j]-realizáció magánhangzósabb természetű siklóhang [–msh, +szon].

A következőkben bemutatjuk, hogy a [j] fentebb ismertetett likvida és a siklóhang realizációinak milyen ejtési sajátosságai vannak. Ehhez azonban előbb részletesebben is be kell mutatnunk mit is értünk a beszédhangok szonoritásán, azaz hangzósságán, és azt, hogy ennek milyen szerepe van a szótagépítésben és fonotaktikában. A szonoritás a bináris fonológiai jegyekkel szemben egy viszonylagos, skaláris tulajdonság, amelyet fonetikai szempontból az intenzitáshoz és hangossághoz/hallhatósághoz kapcsolhatunk (Parker, 2011). Ladefoged (1975, 219) megfogalmazásában a szonoritás a beszédhangnak a többi, vele egyező időtartamú, azonos dallammenettel megvalósuló és azonos hangsúly-pozíciójú hanghoz képesti viszonylagos hangossága. Szonoritásuk szempontjából a beszédhangok öt nagy természetes osztályát az 1. ábrán látható hierarchiai sorba rendezhetjük egymáshoz képest (Clements, 1990).



1. ábra. A beszédhangok öt természetes osztályának szonoritási hierarchiája (Parker, 2011, 1198 nyomán).

A leginkább szonoráns beszédhangok az egyúttal legnyíltabb toldalékcsővel képzett magánhangzók. A hierarchiai sor másik végpontján az obstruensek helyezkednek el, ahol a toldalékcsőben a levegő útjában egy jelentős szűkület vagy zár áll. Ezen toldalékcső-beállítások közvetlen következménye, hogy az akusztikai szerkezetben a magánhangzók esetében inkább kváziperiodikus rezgést, míg az obstruensek esetében valamilyen aperiodikus zörejt várunk (potenciálisan kváziperiodikus rezgéssel, azaz zöngével együtt) (Stevens, 2000). A fent említett [j]-megvalósulások szempontjából releváns siklóhangok kategóriája a szonoritási hierarchiai sorban közelebb helyezkedik el a magánhangzókhoz, mint a likvidáké, tehát a siklóhangok esetében a likvidákhoz képest nagyobb szonoritást, illetve nagyobb toldalékcső-nyitottságot várunk az ejtésben.

Mint említettük, a jelen vizsgálatban a fonemikusan és hiátustöltőként megjelenő [j] variánsok elemzéséhez vezető vizsgálatsorozat egyik első lépéseként azt vizsgáltuk meg, hogy képesek-e a beszélők különbséget létrehozni olyan [j]-variánsok között, amelyek a helyesírás sugalmazása szerint önálló beszédhangként jelennek meg a hangsorban (ez az eset a fonemikus [j] esetével analóg), és amelyek nem (ez a hiátustöltőként megvalósuló [j]-vel hozható párhuzamba). Ezeket rendre az *ijá* /ija:/ [ija:] és az *iá* /ia:/ [ija:] szekvenciákban elemezzük (lentebb bemutatjuk, hogy ezeket a konkrét magánhangzókat azért választottuk, mert korábbi vizsgálatainkban is ezek jelentek meg nyelvközi összevetésben). Lássuk tehát ennek a hangsornak a képzését, fonetikai megvalósításának részleteit, amely részletek rámutatnak arra is, hogy miként lehetnek számszerűsíthetőek a különféle [j]-variánsok sajátosságai között feltételezett eltérések.

Az /i/ és /a:/ magánhangzók, valamint a közöttük megjelenő bármely feltételezett [j]-variáns képzésekor a levegő akadálymentesen áramlik ki a toldalékcsövön. Eközben a toldalékcső üregeinek eltérő beállításával és annak változtatásával létrejönnek a hangsorban a különböző minőségű beszédhangok, azon keresztül, hogy a toldalékcső alaki és méretbeli tulajdonságai révén megszűri a hanghullámokat, azaz felerősít és gyengít bizonyos frekvenciájú összetevőket a zöngében (Fant, 1960). A toldalékcső gerjesztésekor az általa felerősített frekvenciákat, energiacsúcsokat (illetve az ezeket kialakító üregi sajátfrekvenciákat) nevezzük formánsoknak (Fant, 1960; a részletes definícióért l. Deme, 2016, 26). Mind az approximánsokat, mind a magánhangzókat (együttesen: vokalikus elemeket) formánsos szerkezet jellemzi (Ladefoged, 1975).

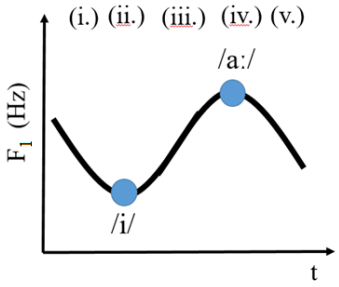
A magánhangzók és az approximánsok közötti különbség úgy ragadható meg, hogy az approximánsok képzésekor a toldalékcső elkeskenyedése (szűkülete) a magánhangzókhoz viszonyítva jelentősebb, de nem olyan mértékű, hogy a hatására turbulens zörej jöjjön létre (Trask, 1996). A magánhangzók és approximánsok produkcióját egy olyan akusztikus csőrendszer segítségével tudjuk modellezni, amelyben a nyelv által okozott szűkület a toldalékcsövet két nagyobb üregre bontja. Az ezáltal létrejövő hátulsó üreg a gégéhez, míg az elülső üreg a fogakhoz és az ajkakhoz áll közelebb. Ezt az akusztikai rendszert például

egy ajakkerekítéses beszédhang esetén egy további szűkülettel és üregekkel lehet kiegészíteni (Stevens, 2000). Mivel a rendszer egyes részei akusztikailag össze vannak kapcsolva, és hatással vannak egymásra, ezért ez alapján a modell alapján csak közelítő megállapítások tehetők a toldalékcső alakjának és a formánsok értékének összefüggéseiről (Stevens, 2000).

Ilyen közelítő, és a jelen vizsgálat szempontjából releváns megállapítás például az első formáns (F_1) frekvenciaértéke és a toldalékcsőben létrejövő szűkület jellemzői közötti összefüggés, miszerint az F_1 értéke a nyelvállás és/vagy állkapocsnyitás fokával fordított arányban áll. Így tehát az általunk vizsgált felső nyelvállású palatális /i/ magánhangzó esetében a magas nyelvállásból, azaz a relatíve kis átmérőjű szűkületből következően alacsony F_1-értékre, míg a legalsó nyelvállású /aː/ magánhangzó esetében a relatíve nagy átmérőjű szűkületből fakadóan magas F_1-értékre számíthatunk (Stevens, 2000). A két magánhangzó között tapasztalhatók pedig az ott (feltételezetten) megvalósuló [j] tulajdonságait mutatják a következőképpen.

Az ún. koartikuláció célja az, hogy a diszkrét (fonológiailag meghatározható és inherens tulajdonságokkal rendelkező) beszédhangok között tranzíció képződjön azért, hogy e diszkrét entitásokból egy összefüggő beszédfolyamot kapjunk (Daniloff & Hammarberg, 1973). A beszéd ugyanis folytonos, és ezzel összefüggésben a beszédhangjainkat alkotó artikulációs gesztusok átfedésben állnak egymással (Magen, 1997). Ennél fogva a beszédhangok artikulációs gesztusai (így például a beszédhangok képzési helye) hatással vannak az őket körülvevő beszédhangok megvalósulására is (vö. Öhman, 1966; Deme et al., 2022), így például a toldalékcső nyíltságára. A koartikulációból, illetve a beszédfolyam folytonosságából kiindulva azt mondhatjuk, hogy egy vokalikus beszédhangszekvencia esetében az akusztikai szerkezetben, pontosabban egy adott formáns frekvenciaértékének változásában megfigyelhető az egyes artikulációs-akusztikai célok megvalósulása. A formánsos akusztikai szerkezetű szegmentumok – elsősorban magánhangzók – esetében három fázist különíthetünk el az ejtés idői lefutásában. Ezeket a fázisokat a 2. ábra mutatja be a feltételezetten csak két

(mögöttes) artikulációs/akusztikai céllal rendelkező *iá* /ia:/ hangsor példáján (egy korábbi tanulmányunk eredményeit alapul véve, Juhász & Deme, 2022b).



2. ábra. Az *iá* /ia:/ vokalikus szekvencia sematikus F_1-menete az idő függvényében, az /i/ és /a:/ akusztikai céljait kiemelve.

A kezdő, rágördülő (onglide) (i.) fázisban jön létre az /i/ artikulációs-akusztikai céljának megközelítéséhez szükséges átmenet. A középső (ii.) fázisban történik meg az artikulációs/akusztikai cél elérése (ahol egy viszonylagosan stabil megvalósulást, ún. „tiszta fázist" is feltételezhetünk, vö. Gósy, 2004), ezután a záró, legördülő (offglide) (iii.) fázisban készül fel a beszélőszervrendszer a következő artikulációs/akusztikai cél (azaz az /a:/ konfigurációs céljának) elérésére, tehát ez egyben a hangsorban következő beszédhang rágördülő (onglide) fázisa is. Itt tehát a folyamat újraindul, a 2. ábrán szemléltetett (iv.) fázisban a képzési szervrendszer eléri a következő artikulációs/akusztikai célt, majd ismételten az előző cél lecsengését és a következő célra való felkészülést tapasztaljuk ((v.) fázis) (vö. Lehiste & Peterson, 1961; Deme et al., 2022).

A fentiekkel szemben az olyan beszédhangoknak a realizációja, mint amilyen a /j/ approximáns is, nem feltétlenül követeli meg a tiszta fázis jelenlétét, többek között azért, mert rövid időtartamú és gyors a megvalósulása. Így például szókezdő pozícióban a fentebbi fázisok közül a /j/ esetében csak a következő akusztikai célra való felkészülést szolgáló legördülő (offglide) tranzíció valósul meg (Catford, 1988). Ettől függetlenül a fonemikus /j/-realizáció egyértelműen azonosítható képzési hellyel rendelkezik, tehát a fentiek nem azt jelentik, hogy a /j/ képzési helye specifikálatlan, hanem pusztán azt, hogy az akuszti-

kai szerkezetére a világosan detektálható kvázistatikus szakasz („tiszta fázis") hiánya jellemző. Ezen a ponton fontos tisztáznunk, hogy képzési vagy artikulációs/akusztikai célon a beszédhang célkonfigurációjának elérését értjük, amit a jelen akusztikai elemzésben Gay (1978) nyomán határozunk meg. A vokalikus elemek esetében az artikulációs/akusztikai célt a formánsmenetek azon pontjának tekinthetjük, ahol a formánsgörbe pozitív vagy negatív irányban „kicsúcsosodik", és ez a maximális vagy minimális kitérés akár stabil fázisként is megjelenhet. Amennyiben a vokalikus fázisban több artikulációs/akusztikai cél is megjelenik, a graduális akusztikai változást a formánsfrekvencia-értékek több pontú mérésével tudjuk vizsgálni. A [j]-realizációk szegmentálása tehát magánhangzók között (és különösen [j] szomszédságában) a fentiekből következően nehéznek bizonyul (ezt ugyanis elsődlegesen a formánsmenetekre alapozhatnánk), különösen azért, mert a [j]-approximáns megvalósulása (a magánhangzókhoz képest) rövid és dinamikus (Jaggers, 2018).

A fonemikus /j/ realizációjáról érdemes még elmondani, hogy mivel egy önálló artikulációs/akusztikai céllal rendelkező (alveolo)palatális approximáns (Recasens, 2013), ezért a magánhangzókhoz (és kiemelten a magas nyelvállású [i] magánhangzóhoz) képest magasabb nyelvállás jellemzi. Ez azt jelenti, hogy a nyelv a [j] megvalósításakor jobban megközelíti a kemény szájpadot, és ennek az artikulációs gesztusnak az akusztikai következményeként a [j]-ben az [i]-hez képest alacsonyabb F_1-értéket várhatunk (Hunt, 2003).

Heselwood (2006), valamint Davidson & Erker (2014) szerint (legalábbis az angol nyelvben) a hiátustöltésben megjelenő [j] és [w] siklóhangok – az intruzív [ɹ] betoldásától eltérően – fonetikai szempontból pusztán csak koartikulációs tranzíciós jelenségek. Ezzel összefüggésben azt feltételezik (egyébként nem kifejezetten az angolra vonatkoztatva), hogy a hiátustöltőként megvalósuló mássalhangzók (így a magyarban a [j]) nem jelennek meg a mögöttes reprezentációban (Casali, 2011, 19), így nincs is önálló artikulációs/akusztikai céljuk sem – ellentétben a fonemikusan jelen lévő /j/ realizációjával. Ez – legalábbis a fonetikai megvalósítást illetően – párhuzamban látszik lenni Siptár & Törkenczy (2000) megállapításával, mely szerint a hiátustöltő [j] a magyarban siklóhang-

ként jelentkezik, melynek a megvalósulása gyengébb és átmenetibb jellegű, mint a mögöttes /j/ fonetikai realizációjáé (Siptár & Törkenczy, 2000, 91), még ha az utóbbi idézett műben a szerzők egyébként a hiátushelyzetben a mögöttes reprezentációban is (vélhetően) feltételezni látszanak a [j]-t. A jelen tanulmány szerzői a mögöttes reprezentációban való megjelenés ügyében nem kívánnak állást foglalni, de a vizsgálat szempontjából nem is tartják relevánsnak ezt a kérdést: azt teszteljük hogy a fonemikusan és hiátushelyzetben megjelenő [j]-k, illetve az ezekkel bizonyos tekintetben analógnak tekinthető helyzetekben megjelenő [j]-k esetében tapasztalhatók-e az egyöntetűnek látszó módon leírt különbségek, amelyek szerint hiátushelyzetben a fonemikus /j/ megvalósulásához képest „gyengébb" és magánhangzósabb elem jelentkezik. A várakozásunk a fonológiai és fonetikai leírások alapján egyaránt az, hogy a hiátustöltőként megjelenő [j]-realizáció (a siklóhangokra is jellemzően) mögöttes cél nélküli és magánhangzósabb természetű, mint a fonemikus /j/ megvalósulása. A jelen vizsgálatban ehhez kapcsolódóan azt tárjuk fel, hogy megfigyelhető-e ez az elkülönülés akkor, ha a beszélőket (szinte) minden lehetséges módon arra irányítjuk, hogy a hiátusos és fonemikus helyzetekhez hasonlóan, de a lehető legnagyobb mértékben különítsék el ezeket a potenciálisan különböző [j]-ket. A fentiek nyomán a hiátustöltőként megjelenő [j] megfelelőjeként itt elemzett, az ortográfia által nem jelölt [j] esetében (az *iá* hangsorban) az [i] és [aː] akusztikai céljai közötti akusztikai tranzícióban nem várunk az approximánsokra jellemző szűkületet a toldalékcsőben. Ez egészen konkrétan azt jelenti, hogy várakozásunk szerint az *iá* hangsorban a magas nyelvállású /i/ rendelkezik a legkisebb átmérőjű szűkülettel, és az /aː/ célját megközelítő koartikulációs tranzícióban ez a szűkület fokozatosan egyre nagyobb átmérőjűvé válik. Megjegyezzük ugyanakkor, hogy a kétféle vizsgált esetre az egyszerűség és az áttekinthetőség kedvéért így utalunk röviden: *ijá* /ijaː/ [ijaː] (fonemikusan megjelenő, illetve a helyesírásban jelölt [j]) és *iá* /iaː/ [ijaː] (hiátustöltőként megjelenő, illetve a helyesírásban nem jelölt [j]).

A fonemikus és hiátustöltőként megvalósuló [j]-realizációk közötti különbségek már számos elemzés és fonetikai vizsgálat középpontjába kerültek, így a

következőkben az ezzel kapcsolatos főbb eredményeket is áttekintjük: elsőként az időtartammal, majd a hangszínnel foglalkozunk. Menyhárt (2006) változatos kontextusokban elemezte az általa percepciós alapon azonosított és szegmentált hiátustöltő [j] megvalósulásait, és ezeket korábbi szakirodalmi adatokkal vetette össze. Arra jutott, hogy a fonemikus [j]-k tartama átlagosan mintegy 20 ms-mal nagyobb, mint a hiátustöltőként megjelenőké. A [j]-realizációk közötti időtartambeli eltérést Siptár (2011) egy „minivizsgálat" keretein belül vetette alá akusztikai mérésnek, ahol összevetett (vélekedése szerint) nyilvánvalóan hiátustöltőként (pl. *mánia*) és nyilvánvalóan fonemikusan (ott: lexikálisan) megjelenő [j]-megvalósulásokat (pl. *kölnije*), illetve olyan eseteket, ahol a [j] státusza ebben a tekintetben kérdéses lehet (pl. *állnia*). A „minivizsgálat" eredményei szerint a fonemikusan megjelenő [j] hosszabb időtartammal valósult meg, mint a hiátustöltő, a kérdéses esetek pedig ehhez képest köztes értékeket mutattak. Gósy (2014) akusztikai elemzésében többféle magánhangzó-kapcsolódást elemzett az időtartam szempontjából, és a fentiekkel megegyezően azt találta, hogy a fonemikus /j/ megvalósulásainak időtartama átlagosan hosszabb volt, mint a hiátustöltőé.

Az eddig említett fonetikai vizsgálatokban közös, hogy azokban a kutatók szegmentálták a [j] megvalósulásait, és így elemezték azokat, miközben – ahogyan fentebb utaltunk már rá – ismert, hogy a [j] hang határainak azonosítása egy vokalikus szekvenciában sok szempontból ütközik nehézségbe. Ennek a nehézségnek az áthidalása érdekében a közelmúltban úgy vizsgáltuk a kérdéses [j]-realizációkat, illetve az ezekkel analóg helyzeteket álszavakban, hogy ahhoz a teljes vokalikus szakaszt egyben elemeztük statikus és dinamikus módszerekkel a jelen tanulmányban is vizsgált hangsorokban (Juhász & Deme, 2022b). Az időtartam szerint az idézett tanulmányban – a korábbi eredményekkel egybehangzóan – azt az eredményt kaptuk, hogy a helyesírás által is jelölten (vö. fonemikusan) is /j/-t tartalmazó *ijá* /ija:/ [ija:] hangsor mássalhangzó-kontextustól függetlenül hosszabb volt, mint a helyesírás által nem jelölt [j] elemmel megvalósított (vö. hiátustöltővel megvalósuló) *iá* /ia:/ [ija:] (Juhász & Deme, 2022b, 305). A [s] kontextus esetében az /ia:/ hangsor időtartama 309 ms, az /ija:/ hangsoré

pedig 333 ms volt, míg a [ʃ] kontextus esetében az /iaː/ hangsor időtartama 311 ms, az /ijaː/ hangsoré pedig 328 ms volt. Ez az eredmény közvetetten azt is megerősíti, hogy a vokalikus szekvenciában jelenlévő akusztikai/artikulációs célok száma eltért: arra következtethetünk, hogy míg az /iaː/-ban csak kettő, addig az /ijaː/ esetében három cél realizálódott az akusztikai szerkezetben.

Az akusztikai szerkezetet illetően Menyhárt (2006) azt találta, hogy a hiátustöltőként megvalósuló [j] (korábbi adatokhoz viszonyítva) átlagosan nem tér el jelentősen, de változatosabbnak látszik, mint a fonemikus /j/ megvalósulásai mind nyíltság képzési jegye (azaz az F_1 formáns frekvenciája), mind pedig az elöl képzettség (azaz az F_2 formáns frekvenciája) mentén. Gósy (2014) elemzése azt mutatta, hogy a fonemikus /j/ megvalósulásának F_1-értéke alacsonyabb volt a hiátustöltő [j] esetében mértnél, bár ez az eltérés pusztán tendenciaszintű (statisztikailag nem alátámasztott) különbség volt. Ez azt jelenti, hogy Gósy (2014) vizsgálatában a fonemikus /j/-realizáció bizonyos mértékben alacsonyabb F_1-értékkel, tehát akusztikai szempontból valamivel zártabban (feltehetőleg magasabb nyelvállással) realizálódott a hiátustöltő [j]-t tartalmazó hangsorokhoz képest, összhangban a várakozásokkal. Gósy (2014) vizsgálata szerint továbbá a fonemikus /j/-realizáció a hiátustöltőhöz képest átlagosan magasabb F_2-értékkel is valósult meg. Ezt a különbséget később Juhász & Deme (2022b) vizsgálatai is megerősítették a fonemikus és hiátustöltős helyzettel analóg álszavas elemzésekben. Az idézett elemzésben ugyanis a megfelelő vokalikus szakaszok F_2-görbéit összevetve eltérést találtak a [j]-realizációk között, mégpedig úgy, hogy az írásban jelölt/fonemikus /j/ megvalósulásaként jelentkező [j] magasabb F_2-maximummal (tehát akusztikai szempontból palatalizáltabban, hátul képzettebben), illetve meredekebb tranzíciós fázissal valósult meg, mint az írásban nem jelölt/hiátustöltőként jelentkező [j]. Az utóbbi vizsgálatban emellett az eredményekből arra is lehetett következtetni, hogy mind az írásban jelölt/fonemikus, mind az írásban nem jelölt/hiátustöltő [j]-realizációk /i/-környezetben (azaz az *iá* /iaː/ és *ijá* /ijaː/ hangsorokban) – feltételezhetően a palatális /i/ koartikulációs hatására – magasabb F_2-értékekkel, tehát a nem-/i/ környezetű [j]

approximánsokhoz képest palatalizáltabban, hátul képzettebben valósultak meg (vö. Juhász & Deme, 2022b).

1.2. Hipotézis

A jelen kísérlet hipotézisének alapját a nemzetközi fonetikai és fonológiai szakirodalomnak a fonemikusan (vagy lexikálisan) és hiátustöltőként megjelenő [j]-re vonatkozó állításai, Siptár (2013), illetve Siptár & Törkenczy (2000) fonológiai elemzései, valamint Menyhárt (2006), Siptár (2011) és Gósy (2014) empirikus kutatásainak eredményei szolgáltatták. A magyar nyelvre vonatkozó fonológiai elemzések szerint – mint korábban említettük – a mögöttesen jelenlévő /j/ likvidaként osztályozandó [+msh, +szon], tehát mássalhangzós tulajdonságokkal rendelkezik, míg a hiátustöltő [j] siklóhang [–msh, +szon], amely a fonemikus [j] megvalósulásánál magánhangzósabb (Siptár, 2013, 6; vö. még Heselwood, 2006). Ennek megfelelően az feltételezhető, hogy a fonemikus /j/ ejtését egy önálló artikulációs/akusztikai cél, továbbá a magánhangzók (így pl. az /i/) ejtésénél jelentősebb szűkület, illetve magasabb nyelvállás (Hunt, 2003; Jaggers, 2018) jellemzi. A hiátustöltő [j]-realizáció ezzel szemben a szakirodalmi források és a várakozásaink szerint is egy, a fonemikusan megjelenő [j]-nél „gyengébb, átmenetibb", magánhangzó-természetű elem, ami lényegében koartikulációs tranzícióként jelentkezik. Ennélfogva a hiátustöltőként jelentkező [j] képzése során nem számíthatunk az /i/ után megjelenő, annál jelentősebb toldalékcső-szűkületet mutató elemre, illetve annak akusztikus lenyomatára.

A jelen vizsgálatban a fentiekből kiindulva azt a feltételezést teszteltük, hogy ha minden eszközzel (az ortográfiai megjelenítéssel, illetve egymás után ejtendő, lényegében minimális párként jelentkező hangsorok produkciójával) a lehető leginkább facilitáljuk a lehetséges [j]-realizációk potenciális akusztikai elkülönítését, akkor képesek-e arra a beszélők, hogy ténylegesen megvalósítsanak valamilyenfajta különbséget ezek között. A kérdésünk tehát az, hogy a beszélők eltérően ejtik-e azokat a [j]-megvalósulásokat, amelyeket az íráskép önálló beszédhangként jelöl, és amelyeket nem. Mivel az előbbi eseteket a fonemikus, utóbbiakat pedig a hiátustöltőként megjelenő [j] megvalósításokkal állíthatjuk

párhuzamba, így a jelen vizsgálat végeredményben arról szolgáltat információkat, hogy várható-e egyáltalán a feltételezetten kétféle [j] elkülönítése azokban az esetekben, ahol kétséget kizáróan fonemikusan vagy hiátustöltőként jelenik meg. Feltételezésünk szerint az írásban is jelölt (vö. a fonemikusan megjelenő) [j] mutat önálló artikulációs/akusztikai célt, szemben az írásban nem jelölt (vö. a hiátushelyzetben megjelenő) [j]-vel, amely mássalhangzósabb, tehát zártabb. Ezt az F_1 formáns elemzésével ellenőrizzük: a mássalhangzósabb, tehát zártabb (írásban jelölt/fonemikusan jelentkező [j] estében) alacsonyabb értéket vártunk a két magánhangzós cél közötti átmenetben.

2. Módszertan

A vokalikus hangsorok ejtését 14 magyar anyanyelvű nő (átlagéletkoruk 21,3 ± 1,8 év) felolvasásában rögzítettük. A felvételkészítésre az ELTE BTK Alkalmazott Nyelvészeti és Fonetikai Tanszékén egy csendesített szobában került sor. A hangfelvételeket az *Audacity* (The Audacity Team, 2023) programmal rögzítettük egy külső hangkártyával és egy omnidirekcionális kondenzátoros fejmikrofonnal. Vizsgáltukban az írásban nem jelölt (vö. hiátustöltő [j]-vel megvalósuló) *iá* /iaː/ szekvenciát, illetve az írásban jelölt (vö. fonemikus /j/-realizációval megjelenő) *ijá* /ijaː/ szekvenciát ejtették a beszélők egy szótagú álszavakban (ahol az onszet /s/ vagy /ʃ/ volt), izolált ejtésben, véletlenszerű sorrendben, fejenként nyolcszori ismétlésben. Referenciaként további két vokalikus hangsort is felvettünk ugyanezen két szibilánskörnyezetben: a C+/aː/ szekvenciát, illetve a nem /i/-kontextusú [j] approximánst tartalmazó C+/jaː/ kontextust (1. táblázat).

A kísérlet során az ejtés könnyítésének céljából a mássalhangzó-torlódásos (/sjaː/, /ʃjaː/) hangsorok bemutatásakor az orientálódást segítendő jelentéssel rendelkező szavak is láthatóak voltak a képernyőn (pl. *vasjáték* → *sjá* /ʃjaː/, *vészjáték* → *szjá* /sjaː/), azonban ezeket a szavakat a kísérleti személyeknek nem kellett felolvasniuk. Erre azért volt szükség, mert így biztosítottuk azt, hogy a beszélők a magyar nyelvre nem jellemző szó(tag)kezdő mássalhangzó-torlódást

1. táblázat. A vizsgált hangsorok, a szibilánsok és vokalikus hangsorok (azaz a környezet) szerint rendezve.

nem palatális /Caː/ szekvenciák	*sá* /ʃaː/
	szá /saː/
[j]-t tartalmazó /Cjaː/ szekvenciák	*sjá* /ʃjaː/
	szjá /sjaː/
írásban nem jelölt [j]-realizációt tartalmazó /Ciaː/ szekvenciák	*siá* /ʃiaː/
	sziá /siaː/
írásban jelölt [j]-realizációt tartalmazó /Cijaː/ szekvenciák	*sijá* /ʃijaː/
	szijá /sijaː/

ne oldják fel egy magánhangzó beékelésével (vö. **szjá* → *szijá*) (vö. Guevara-Rukoz, 2018). Összesen (2 szibiláns × 4 vokalikus környezet × 8 ismétlés × 14 kísérleti személy =) 896 hangsort vizsgáltunk.

A hangfelvételeket a *Praat* szoftverben (Boersma & Weenink, 2020) címkéztük és elemeztük. Az elemzés a teljes vokalikus hangsorra fókuszált, ezeket a teljes formánsszerkezet megjelenésétől a kváziperiodikus hullám megszűnéséig szegmentáltuk. Dinamikus elemzéssel elemeztük a szűkület keresztmetszetének alakulását, amit az F_1 értékével számszerűsítettünk: ehhez az F_1 formáns frekvenciájának értékét 5 ms-onként nyertük ki a szegmentált időtartamon belül a Burg-algoritmus segítségével.

A dinamikus elemzéshez általánosított additív kevert modelleket használtunk (generalised additive mixed model, GAMM) az mgcv csomag (Wood, 2017) segítségével az *R* programban (R Development Core Team, 2021). A GAMM-ok nemlineáris adatsorok elemzésére szolgálnak, ezért a formánsmenetek esetében – ahol görbék lefutását elemezzük – jobb illesztést biztosítanak a lineáris modelleknél (Wieling, 2018). Összesen két GAMM-modellt írtunk fel a két szibilánskontextusra külön-külön. Az alapmodellekben az F_1 frekvenciáját, mint függő változót vizsgáltuk a NORMALIZÁLT IDŐTARTAM függvényében. (A voka-

likus szakasz normalizált időtartamát úgy kaptuk meg, hogy az F_1-et kinyerő szkriptben 5 ms-ként növekvő időértékeket arányítottuk a teljes vokalikus szakasz időtartamával). Emellett minden esetben felírtunk egy bővített modellt is, melyben az adott alapmodellt egy parametrikus kifejezéssel, azaz egy sorba rendezett faktor változóval egészítettük ki, ez volt a vokalikus szakaszra utaló HANGSOR változó. Mivel külön modellekkel vizsgáltuk a dentialveoláris /s/ és a posztalveoláris /ʃ/-t követő hangsorokat, így ezekben a modellekben a HANGSOR 4 szinttel rendelkezett (/ija:/, /ia:/, /ja:/, /a:/). E változó első szintje minden esetben az /ija:/-hoz tartozó F_1-görbe volt, amit a modell referenciagörbének tekintett, és ehhez a görbéhez képest becsülte a változó többi szintjéhez tartozó differenciagörbéket. Az alap és bővített modelleket a χ^2-próbával vetettük össze az itsadug compareML() parancsának segítségével (van Rij et al., 2020). Minden esetben a bővített modell illesztése bizonyult szignifikánsan jobbnak. A bővített modellekbe random simítást illesztettünk tokenenként, majd az autokorreláció ellenőrzése (acf.resid()) után annak korrekcióját is elvégeztük. Végül az így kapott modellt a gam.check() paranccsal ellenőriztük. A modellek által becsült F_1-görbéket minden esetben 95%-os konfidencia-intervallummal ábrázoltuk.

3. Eredmények

Az eredmények szerint mindkét szibiláns környezetében a NORMALIZÁLT IDŐTARTAM változón túl a HANGSOR változót is tartalmazó bővített modell illeszkedett jobban az adatokra (/s/ = $\chi^2(11{,}00) = 162474{,}7$; $p < 0{,}001$; /ʃ/ = $\chi^2(11{,}00) = 163504{,}6$; $p < 0{,}001$). A simítás és az autokorreláció kezelése után az /s/ kontextusú modell az adatokban talált variancia 92,3%-át, a /ʃ/ kontextust vizsgáló modell pedig a 85,9%-át magyarázta (3. ábra). A GAMM-modell parametrikus együtthatói (2. táblázat) azt mutatják, hogy az *ijá* /ija:/ vokalikus fázisának F_1-görbéi rendelkeztek a legalacsonyabb átlagértékkel szibilánskontextustól függetlenül (*szijá* /sija:/ = 637,6 Hz, *sijá* /ʃija:/ = 644,2 Hz) a vokalikus hangsorok között, és ennél szignifikánsan magasabb átlagértékkel realizálódtak az /ia:/ (*sziá* /sia:/ = 656,4 Hz, *siá* /ʃia:/ = 665,8 Hz) megvaló-

sulásai. (A becsült különbségek az *ijá* /ija:/ és *iá* /ia:/ hangsorok F_1-görbéinek átlaga között az /s/-kontextus esetében: 18,8 Hz; $p < 0{,}01$; az /ʃ/-kontextus esetében: 21,6 Hz; $p < 0{,}01$). Ugyanezen parametrikus együtthatók, azaz a becsült átlagértékek szempontjából a *já* /ja:/, illetve az *á* /a:/ is magasabb F_1-értékkel rendelkezett mindkét szibilánskontextusban, mint az /i/-t tartalmazó *iá* /ia:/ és *ijá* /ija:/: ez az átlagérték a *szjá* /sja:/ esetében 739,5 Hz, a *sjá* /ʃja:/ esetében 724,9 Hz, míg az *szá* /sa:/ esetében 846,8 Hz és a *sá* /ʃa:/ esetében 813,3 Hz volt.

2. táblázat. A GAMM-ok parametrikus és smooth eredményei a vizsgált vokalikus hangsorokra vonatkozóan

(a)

A dentialveoláris /s/-környezetű vokalikus szekvenciák

parametrikus együtthatók

hangsor	becsült átlag	t	$Pr(>\|t\|)$
/sija:/	637,6	124,6	$< 0{,}001$
/sia:/	656,4	2,6	$< 0{,}01$
/sja:/	739,5	14,1	$< 0{,}001$
/sa:/	846,8	28,9	$< 0{,}001$

smooth együtthatók

hangsor	*EDF*	F	p
/sija:/	8,5	63,5	$< 0{,}001$
/sia:/	8,4	46,2	$< 0{,}001$
/sja:/	7,5	50,1	$< 0{,}001$
/sa:/	7,1	20,6	$< 0{,}001$
traj.	2964,4	7,9	$< 0{,}001$

(b)

A posztalveoláris /ʃ/-környezetű vokalikus szekvenciák

parametrikus együtthatók

hangsor	becsült átlag	t	$Pr(>\|t\|)$
/ʃija:/	644,2	113,4	$< 0{,}001$
/ʃia:/	665,8	2,7	$< 0{,}01$
/ʃja:/	724,9	10,0	$< 0{,}001$
/ʃa:/	813,3	20,9	$< 0{,}001$

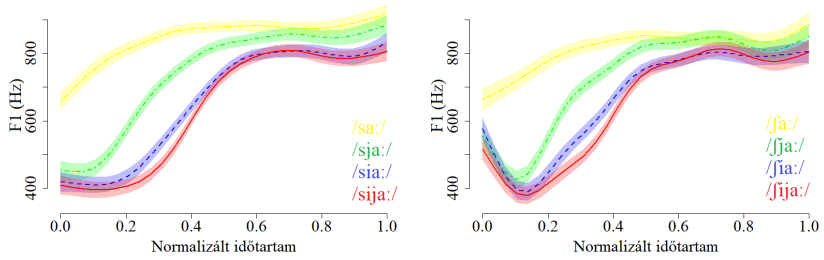
smooth együtthatók

hangsor	*EDF*	F	p
/ʃija:/	8,3	14,6	$< 0{,}001$
/ʃia:/	8,5	19,6	$< 0{,}001$
/ʃja:/	7,6	23,1	$< 0{,}001$
/ʃa:/	3,1	16,1	$< 0{,}001$
traj.	2844,1	6,446	$< 0{,}001$

A referenciagörbeként meghatározott *szijá* /sija:/ és *sijá* /ʃija:/ vokalikus szakaszainak F_1-értéke eltérést mutatott az azonos szibilánskontextusú *iá* /ia:/, *já* /ja:/ és *á* /a:/ hangsoroktól a normalizált időtartamon belül (l. smooth együtthatók esetében mindig $p < 0{,}001$) (2. táblázat). A hangsorok között az

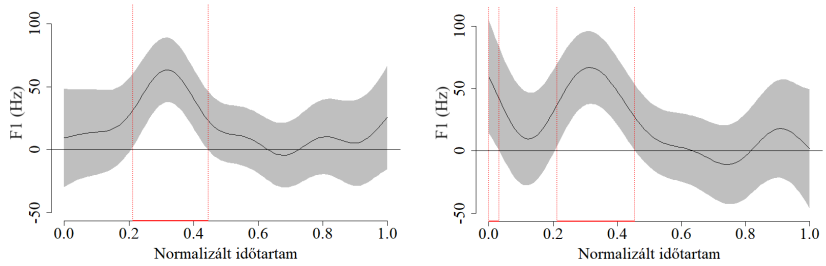
ijá /ija:/ hangsor F_1-görbéjének modellezéséhez volt szükség a legtöbb különböző függvény használatára (ennek volt a legmagasabb az *EDF*-, azaz *Effective Degrees of Freedom* értéke), ami azt mutatja, hogy ebben az esetben várható a normalizált időtartam szempontjából a leginkább „hullámzó", tehát a legkevésbé lineáris görbe (Chuang et al., 2020, 3), azaz minél magasabb az *EDF* értéke, annál jelentősebb kitérések várhatók a görbe alakjában. Mindezek alapján, illetve a 2. táblázatban szereplő adatok alapján ennélfogva azt a következtetést is levonhatjuk, hogy minél magasabb az *EDF* értéke, annál több artikulációs/akusztikai cél jelenik meg a hangsorban, illetve ezek között a célok között annál nagyobb a nyelvállásfokbeli eltérés.

Az F_1-görbéket (3. ábra) szemügyre véve azt láthatjuk, hogy az /s/ kontextus esetében a normalizált időtartam kezdeti szakaszában mind az *iá* /ia:/, mind az *ijá* /ija:/ hangsor alacsony F_1-értékkel valósult meg, amit a normalizált időtartam 20%-a körül egy magasabb frekvenciaértékek felé irányuló tranzíció követ. Itt, a vokalikus szakasz első 10% és 20%-a között azonosítható tehát az első akusztikai cél, az [i]-é. Az ezt követően, 20% környékén induló átmenet (emelkedés az F_1 értékében) megközelítőleg a normalizált időtartam 60%-áig tart, ami után az F_1-értékek viszonylag konstans, magas értéket vesznek fel. Ez utóbbi pontot azonosíthatjuk az [a:] akusztikai céljaként. Tehát: míg a normalizált időtartam elején az /i/ magánhangzó magas nyelvállása alacsony F_1-értéket eredményezett a hangsorok vokalikus szakaszában, addig a hangsor végén ejtett legalsó nyelvállású /a:/ magas F_1-értékkel valósult meg (3. ábra, bal oldal). Ehhez hasonló mintázatok figyelhetőek meg a posztalveoláris /ʃ/-kontextus esetében is (3. ábra, jobb oldal). Ez utóbbi kontextusban azonban az is szembetűnő, hogy a normalizált időtartam kezdetén az /i/-hez tartozó akusztikai cél egy rágördülő fázissal (onglide-dal) realizálódott, azaz a minimális F_1-értéket egy lefelé ívelő tranzíciós fázis előzte meg, amit az /s/ kontextus esetében nem tapasztaltunk. A 4. ábrán látható páros összevetést előrevetítve ezzel kapcsolatosan az is elmondható, hogy a normalizált időtartam 0–4%-ig tartó időablakában az /ia:/ F_1-görbéje szignifikánsan magasabb értékekkel, azaz valószínűsíthetően zártabban realizálódott, mint az /ija:/ F_1-menete.



3. ábra. Az /s/-kontextusú vokalikus szakaszok (balra), valamint a /ʃ/-kontextusú vokalikus szakaszok (jobbra) becsült F_1-értékei a normalizált időtartam függvényében.

Az /ija:/ és /ia:/ vokalikus szakaszok páros összevetését mássalhangzó-kontextusonként a 4. ábra mutatja be. Ebben az összehasonlításban a modell a két görbe közötti eltérést becsüli úgy, hogy a hangsorpár egyik elemét tekinti referenciának, ez a jelen esetben az /ija:/ volt. Ebből következően, ha a referenciagörbe értékei alacsonyabbak, mint a vele összevetett görbe értékei, akkor negatív becsült (Hz-ben mért) különbségértéket kapunk.



4. ábra. A /sija:/ és /sia:/ (balra), valamint a /ʃija:/ és /ʃia:/ (jobbra) hangsorok F_1-görbéi közötti becsült eltérés, ahol a C/ija:/ hangsor F_1-görbéje a referencia, és a piros szín azt jelöli, ha és ahol a két görbe a normalizált időtartam valamely szakaszában eltér egymástól.

A páros összevetések szignifikáns különbséget mutattak az *iá* /ia:/ és *ijá* /ija:/ F_1-görbéi között, mégpedig úgy, hogy mindkét szibiláns kontextusában az *ijá* /ija:/-hoz tartozó görbe realizálódott az *iá* /ia:/-hoz képest alacsonyabb F_1-értékekkel. A szignifikáns eltérés időzítése a két kontextus esetében lényegé-

ben azonos volt: az /s/-kontextus esetében a normalizált időtartam 21%-a és 44,5%-a között volt megfigyelhető, míg az /ʃ/ kontextus esetében 21% és 45% között (4. ábra). Mindezt összevetve a 4. ábrán láthatóakkal, illetve az ott azonosítható /i/-hez és /aː/-hoz tartozó artikulációs/akusztikai célokkal arra jutunk, hogy a [j] megvalósulásai erre, a kb. 20% és 45% közötti szakaszra tehetők a vizsgált hangsorok vokalikus szakaszában, amely szakaszok az /i/ és az /aː/ akusztikai célok közötti átmenetben, tranzícióban valósulnak meg. Ezzel a tranzícióval, illetve a [j] megvalósulásaival kapcsolatosan pedig azt állapíthatjuk meg, hogy az *ijá* /ijaː/ szekvenciát alacsonyabb F_1, azaz vélhetően magasabb nyelvállás/zártabb toldalékcső jellemzi, mint az *iá* /iaː/ szekvenciát.

Érdemes kiemelni ennek a 20% és 45% közötti szakasznak az alaki jellemzőit is: míg az *iá* /iaː/ hangsor esetében (szibilánskontextustól függetlenül) lineáris tranzíciót látunk itt, addig az *ijá* /ijaː/ homorúbb átmenettel valósult meg az /i/ és az /aː/ artikulációs/akusztikai céljai között. Továbbá fontos kiemelni azt is, hogy a normalizált időtartam kezdetén, azaz az /i/ akusztikai céljának elérésekor tendencia szintjén az /ijaː/ hangsorban megvalósuló /i/ alacsonyabb F_1-értékkel rendelkezett, mint az /iaː/ hangsor első eleme.

A tanulmány tárgyához, azaz a [j]-megvalósulások kérdéséhez csak közvetetten kapcsolódik az a megfigyelés, mely szerint a /saː/ és /ʃaː/ esetében (szibilánskörnyezettől függetlenül) kevésbé meredek tranzíciót találtunk a vokalikus hangsor első felében a mássalhangzó és az /aː/ feltételezett céljának elérése között, amelyet nem is jellemzett alacsonyabb kezdeti, illetve átlagos F_1-érték (3. ábra). Ez annak köszönhető, hogy nem felső nyelvállású beszédhanggal kezdődött a szakasz és így nem is jelentkezett jelentősebb változás a nyelvállás fokában az /aː/ feltételezett céljának eléréséig. A /jaː/ hangsorokban az F_1-érték emelkedésével járó tranzíció – szibilánskontextustól függetlenül – korábban elindult a normalizált időtartamban, valamint szignifikánsan magasabb F_1-értékekkel realizálódott, mint az /iaː/ és /ijaː/ vokalikus fázisokban. Szignifikáns különbséget a /sjaː/ és /sijaː/ pár esetében a normalizált időtartam 6% és 100%-a között, míg a /ʃjaː/ és /ʃijaː/ pár esetében a normalizált időtartam 10,1% és 74,7%-a között találtunk.

4. Következtetések

Az itt bemutatott vizsgálatunk a hiátustöltőként és fonemikusan megjelenő [j] realizációk potenciális különbségeinek leírásához kíván hozzájárulni. A korábbi fonetikai és fonológiai szakirodalom alapján az feltételezhető, hogy a fonemikus /j/ ejtését egy önálló artikulációs/akusztikai cél és a magánhangzókénál (így pl. az /i/) ejtésénél zártabb toldalékcső, azaz jelentősebb szűkület, illetve magasabb nyelvállás jellemzi, hiszen a fonemikus /j/-realizáció egy mássalhangzósabb természetű likvida. A hiátustöltő [j]-realizáció ezzel szemben a leírások szerint a fonemikus /j/-realizációnál magánhangzósabb természetű. Ennek megfelelően a hiátustöltő [j] realizációjában nem várható az /i/ után és az /aː/ előtt megjelenő, az /i/-ben tapasztaltnál jelentősebb toldalékcső-szűkületet mutató elem vagy szakasz, illetve annak akusztikus lenyomata.

A jelen tanulmányban bemutatott kísérletben dinamikus elemzéssel hasonlítottuk össze az *iá* /iaː/ és az *ijá* /ijaː/ vokalikus szekvenciák megvalósulását magyar álszavakban a létrejövő szájüregi szűkület akusztikai vetületének, azaz az első formánsfrekvencia menetének vizsgálata szempontjából. Ennek a vizsgálatnak a célja az volt, hogy fényt derítsünk arra a kérdésre, hogy a kísérletben résztvevők képesek-e egyáltalán elkülöníteni ezeket a hangsorokat és az ezekben megvalósuló [j]-t úgy, hogy erre minden körülmény a lehető leginkább facilitáló hatással van. A kísérlet távlatilag a fentebb említett szembenállás elemzését, azaz a hiátustöltőként és fonemikusan megjelenő [j] összehasonlítását alapozza meg, melyek a jelen esetben vizsgált párokkal analógiában jellemzően rendre jelöltek vagy sem a helyesírásban.

Az itt elemzett *ijá* /ijaː/ hangsorban azt vártuk, hogy a beszélők az írásban jelölt, fonemikusan megjelenő [j]-t egy különálló artikulációs/akusztikai céllal és zártabban ejtik, mint az *iá* /iaː/ hangsorban megjelenő [j]-t. Utóbbi esetben ugyanis az íráskép (a hiátus jelöléséhez hasonlóan) pusztán két beszédhang jelentkezését sugallja, amit a beszélők várakozásaink szerint két akusztikai-artikulációs céllal (azaz, kizárólag az [i] és az [aː] ejtésével) valósítanak meg. A toldalékcső nyíltságát, tehát a nyelvállás fokát (ill. az állkapocsnyitás mérté-

két) az első formáns frekvenciaértékével (F_1), és annak dinamikus változásával ragadtuk meg a kérdéses hangsorok teljes vokalikus szakaszának vizsgálatával (tehát nem kísérletük meg hanghatárok megállapítását a [j] megvalósulásaiban).

A kísérlet eredményei szerint a kérdéses hangsorok vokalikus szakaszai (azaz az *iá* /ia:/ és az *ijá* /ija:/) szibilánskontextustól függetlenül a normalizált időtartam ugyanazon szakaszában tértek el egymástól, mégpedig a hipotézisekben leírt irányokban: az /ija:/ adott szakasza az /ia:/-hoz képest alacsonyabb F_1-értékekkel, tehát akusztikai szempontból zártabban valósult meg. Mindebből arra következtethetünk, hogy az álszavak vokalikus szakaszában szignifikánsan eltérő szakasz az írásban nem jelölt/hiátustöltő és az írásban jelölt/fonemikus [j] megvalósulásának akusztikus lenyomata, melyek valóban eltérnek úgy, hogy az írásban jelölt/fonemikusan megjelenő [j] esetében akusztikai vetületében kisebb átmérőjű a toldalékcsőben jelentkező szűkület, azaz mássalhangzósabb az ejtés. Ezek az eredmények arra utalnak, hogy ha minden körülmény ezt támogatja, akkor a beszélők az íráskép alapján potenciálisan elkülönítik azokat a [j]-variánsokat, amelyeket az ortográfia önálló beszédhangként jelenít meg a hangsorban, és amelyeket nem. Megerősítettük tehát azt a feltételezést, hogy létrejöhet a korábbi szakirodalom alapján sugallt eltérés a hiátustöltő és fonemikus [j] megvalósulása között. Kutatásunk következő lépésében tehát arra kereshetjük majd a választ, hogy az ebben az esetben, álszavak ejtésében tapasztalt eltérés megjelenik-e olyan valódi szavak esetében, a) ahol a helyesírás megjelenít vagy nem jelenít meg [j]-t olyan helyzetekben, amikben egyébként az ejtésben mindenképpen feltételezzük a megjelenését (vö. /i/+V vagy V+/i/ kapcsolatok), vagy b) ahol akár a fonológiai elemzés felől is egyértelműnek látszik, hogy a helyesírásban nem megjelenő [j] hiátustöltő elem. Ezekre a kérdésekre jelenleg is zajló kísérleteinkben keressük a választ.

Az itt vizsgált hangsorok megvalósításának eltéréseivel kapcsolatban érdemes megjegyezni azt is, hogy míg az *iá* /ia:/-ban jelentkező [j]-realizáció esetében egy lineáris átmenetet láthattunk a szótagkező szibiláns mássalhangzó és a szóvégi /a:/ között, addig az *ijá* /ija:/ esetében (főleg posztalveoláris szibilánskontextusban) a tranzíció inkább domború mintázatot mutatott. Habár az

ez utóbbi esetben tapasztalt kitérés nem látszik egyértelműen például a magánhangzók esetében is megjelenő, az artikulációs/akusztikai célt jelző kitérésnek, mégis azt sejteti, hogy ebben az esetben nem pusztán egy, a beszédhangok közötti tranzíciós szakaszról van szó. Az *iá* /iaː/ esetében azonban erről van szó: a lineáris átmenet ugyanis egyértelműen a beszédszerveknek a lehető leggazdaságosabb átrendeződését sejteti a két magánhangzós artikulációs/akusztikai cél között. Megjegyzendő, hogy a jelen vizsgálatban kizárólag olyan magánhangzókat elemeztünk [j]-kontextusokként, amelyek egymástól artikulációs és akusztikai értelemben is a legmesszebb helyezkednek el azért, mert itt detektálhatók a legjobban a kérdéses [j] ejtési sajátosságai.

A fonemikus /j/ approximánst (melyet a jelen elemzésben az írásban megjelenített [j] analógiája képezett le) olyan beszédhangként írtuk le, amelyet az /i/ magánhangzóhoz képest jelentősebb szűkület jellemez. Ennek látszólag ellentmond az az eredmény, hogy a *szjá* /sjaː/ és *sjá* /ʃjaː/ hangsorokban megjelenő /j/ mind az írásban nem jelölt/hiátustöltőként jelentkező [j]-t *sziá* /siaː/ és *siá* /ʃiaː/, mind pedig az írásban jelölt/fonemikusan jelentkező [j]-t tartalmazó *szijá* /sijaː/ és *sijá* /ʃijaː/ hangsor [j] eleméhez képest magasabb F_1-értékkel rendelkezett, tehát ezeknél akusztikai vetületében nagyobb átmérőjű toldalékcső-szűkülettel valósult meg. Ez az ellentmondás két körülmény összjátékával magyarázható: a beszédhangkörnyezet koartikulációs hatásával, illetve a /j/ approximáns rövid időtartamú, dinamikus megvalósulásával. A beszédhangkörnyezet koartikulációs hatása vélekedésünk szerint abban nyilvánulhatott meg, hogy a C/jaː/-típusú hangsorokban mind az alveoláris/posztalveoláris mássalhangzós kontextus (tehát az itt jelentkező nyelvvel képzett akadály), mind (és különösen) a követő /aː/ ejtésére jellemző alsóbb nyelvállás ellene hathatott az (alveolo)palatális régióban megvalósuló nyelvemelkedésnek (ami a /j/ mássalhangzósabb ejtéséhez lenne szükséges), és a koartikuláció révén ezek „lejjebb húzhatták" a /j/-nek az idealisztikus célkonfigurációját jobban közelítő magasabb nyelvállásfokot. Ezt a hatást pedig erősíthette a /j/ approximáns rövid és dinamikus ejtése is tehát az a tény, hogy a fonemikus [j]-ben az artikulá-

ciós/akusztikai cél megközelítésére rövid idő állt rendelkezésre, ami nagyobb mértékű célalulmúláshoz vezethet.

A fentiekkel ellentétben az *ijá* /ija:/ hangsor esetében azt feltételezhetjük, hogy az /i// magánhangzó a progresszív koartikuláció révén segítette a nyelv emelkedését a /j/-ben az idealisztikus célkonfigurációhoz közelebb megvalósulni: mivel a palatális régió megközelítése már az /i/ ejtésekor megtörtént, így az /ija:/ hangsorban megjelenő /j/-realizáció esetében a toldalékcső szűkítése és a nyelvtest megemelése is hatékonyabban mehetett végbe (a rövid időtartamú megvalósulás dacára), mint akkor, amikor az approximáns a szótagkezdő szibiláns után állt közvetlenül. Ha a koartikulációs folyamatokat az ellenkező irányból, azaz regresszíve szemléljük, akkor arra is magyarázatot találhatunk, hogy az /i/ megvalósulása miért tért el (a tendencia szintjén) ugyanezen hangsorokban úgy, hogy az *ijá* /ija:/ hangsorban megjelenő /i/ alacsonyabb F_1-értékkel, azaz akusztikailag zártabban, feltehetőleg magasabb nyelvállással valósult meg, mint az *iá* /ia:/ hangsorban. Itt ugyanis arra következtethetünk, hogy az /i/-re az őt követő, írásban jelölve jelentkező/fonemikus /j/-realizáció megvalósulása hatott: mivel a /j/-realizációjának az ejtése magasabb nyelvállást és kisebb átmérőjű szűkület létrehozását kívánta meg, mint ami az /i/-t jellemzi, ezek koartikulációs hatásként az /i/ nyelvemelkedését is feljebb húzhatták az *ijá* /ija:/ hangsorokban ahhoz képest, mint amit az /ia:/-ban tapasztalhattunk.

Az előzőekben bemutatott gondolatmenet kapcsán érdemes érinteni egy terminológiai kérdést is a fonemikus /j/-realizáció képzési helyével kapcsolatban. A szakirodalomban széles körben alkalmazott szóhasználatban a /j/ approximáns – az /i/ magánhangzóhoz hasonlóan – palatális képzési helyű. Azonban egyes források (pl. Recasens, 2013) szerint a /j/ egy, a palatális régiónál előrébb képzett, ún. alveolopalatális képzési helyű beszédhang. Ez azt jelenti, hogy az /i/ és /j/ képzési helye eltér, ennélfogva az [i] környezetében álló [j] ejtése – a koartikuláció révén – is eltérhet a nem [i] környezetében álló [j] ejtésétől is: míg a nem [i] környezetében álló [j] alveolopalatális; addig az [i] környezetében álló [j] ennél hátrébb képzett, palatális beszédhang. Ezt a magyar /j/ elemzéséből származó korábbi kísérleteink közvetetten megerősítik: az itt is elemzett

hangsorokat vizsgálva azt találtuk korábban, hogy a nem [i]-kontextusú írásban megjelenő/fonemikus /j/ megvalósulása (pl. a *szjá* /sjaː/ hangsorban) akusztikai szempontból előrébb képzett (alveolopalatálisabb), és a nyelvállás fokát tekintve alacsonyabb nyelvállású/nyíltabb, mint az [i]-kontextusú /j/ megvalósulása (pl. a *szijá* /sijaː/ hangsorban) (Juhász & Deme, 2022b).

A kísérlet további eredményeit tekintve fontos említést tennünk arról, hogy a vokalikus fázis elején miért találtunk szignifikáns eltérést a posztalveoláris /ʃ/-kontextusú *iá* /iaː/ és *ijá* /ijaː/ hangsorok között, és a dentialveoláris /s/-kontextus esetében miért nem. Erre a kérdésre véleményünk szerint a két szibiláns eltérő mértékű és jellegű palatalizációja ad választ: ugyanezen hangsorok szibiláns beszédhangjait vizsgálva egy korábbi kísérletben (Juhász & Deme, 2022a) azt találtuk, hogy míg a dentialveoláris /s/ aperiodikus zörejének záró fázisában a palatális /i/ kontextus hatására frekvenciacsökkenés mutatkozott, azaz a képzési hely relatív pozíciója a szájüregben vélhetően hátrafelé mozdult, addig a posztalveoláris /ʃ/ esetében nem. Ez más szavakkal azt jelenti, hogy míg az artikulátorok a dentialveoláris /s/-ben már a szibiláns ejtése alatt megközelítették a következő palatális artikulációs/akusztikai célt, addig a posztalveoláris /ʃ/ esetében a palatális képzési helyet megközelítő koartikulációs tranzíció csak „késve", a szibilánst követő vokalikus fázisban ment végbe. Vélekedésünk szerint ebből következően láthattuk azt a jelen adatokban, hogy a dentialveoláris *sziá* /siaː/ és *szijá* /sijaː/ F_1-görbéje egy stabilabb, konstans fázissal indult, míg a posztalveoláris *siá* /ʃiaː/ és *sijá* /ʃijaː/ esetében az F_1 értéke meredekebb csökkenést mutatott a görbe kezdeti szakaszában. Továbbá ezen okból fakadóan találhattunk szignifikáns eltérést a *siá* /ʃiaː/ és *sijá* /ʃijaː/ hangsorok kezdeti fázisában is, hiszen mint fentebb említettük, az *ijá* /ijaː/ /i/-je az *iá* /iaː/ /i/-jéhez képest tendenciózusan alacsonyabb F_1-gyel (nyíltabban) realizálódott.

Összegezve, a jelen tanulmányban megerősítettük azt, hogy a magyar anyanyelvű beszélők képesek elkülöníteni az írásban jelölt és nem jelölt (de az ejtésben a hangsor sajátosságaiból fakadóan mindenképpen jelentkező) [j]-ket, tehát potenciálisan képesek lehetnek elkülöníteni a hiátustöltőként és a fonemikusan megjelenő [j]-realizációkat is. A hiátustöltő és fonemikus [j]-kkel analógiában

itt elemzett (rendre) írásban jelölt és nem jelölt [j]-realizációk akusztikailag úgy tértek el, hogy az írásban jelölt [j] az írásban nem jelölthez képest akusztikai szempontból zártabban, tehát magasabb nyelvállással, illetve kisebb átmérőjű szűkülettel valósul meg. Ebből arra következtettünk, hogy az írásban is jelentkező [j] az írásban nem jelöltnél mássalhangzósabb ejtésű, amely különbséget egyébként a fonemikusan és hiátustöltőként megjelenő [j]-k között is várhatnánk a szakirodalom alapján. Mindezek alapján a jövőben arra a kérdésre is megpróbálhatunk választ találni, hogy az itt kimutatott különbségek valóban jellemzik-e a fonemikus és hiátustöltő [j]-realizációkat is (valódi szavakban, a túlartikulációt kevésbé facilitáló környezetekben). Erre a kérdésre jelenleg is zajló kísérleteinkben keressük a választ. A jelen kísérlet eredményei hozzájárulnak a [j] különböző megvalósulásainak, illetve a hiátustöltés folyamatának mélyebb megértéséhez, valamint bővítik az ismereteinket arról, hogy a beszéd ortográfiai megjelenítése miként hat a beszédhangok megvalósítására, illetve hogyan függ össze azzal.

Köszönetnyilvánítás

A kutatás a Kulturális és Innovációs Minisztérium EKÖP-24 kódszámú Egyetemi Kiválósági Ösztöndíj Programjának a Nemzeti Kutatás, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült, valamint a TKA–DAAD 177375. sz. projektje, illetve az NKFIH FK128814 sz. projektje támogatta. Köszönjük névtelen bírálóinknak, hogy észrevételeikkel segítették a tanulmány érvelésének (reményeink szerint) meggyőzőbb felépítését.

Hivatkozások

Balogné Bérces, K. (2006). What's wrong with vowel-initial syllables? *SOAS Working Papers in Linguistics*, *14*, 15–21.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [computer program]. 6.1.15-ös verzió. (Letöltés ideje: 2019. november 4.).

Brown, G. (1970). Syllables and redundancy rules in generative phonology. *Journal of Linguistics*, *6*, 1–17. URL: https://doi.org/10.1017/S0022226700002322. doi:10.1017/S0022226700002322.

Casali, R. (2011). Hiatus resolution. In M. Oostendorp (Ed.), *The Blackwell Companion to Phonology* (pp. 1469–1496). Malden: Wiley-Blackwell. URL: https://doi.org/10.1002/9781444335262.wbctp0061. doi:10.1002/9781444335262.wbctp0061.

Catford, J. (1988). *A practical introduction to phonetics*. Oxford: Clarendon Press.

Chuang, Y., Fon, J., & Baayen, R. (2020). Analyzing phonetic data with generalized additive mixed models. *PsyArXiv*, (pp. 1–27). URL: https://doi.org/10.31234/osf.io/bd3r4. doi:10.31234/osf.io/bd3r4.

Clements, G. (1990). The role of the sonority cycle in core syllabification. In J. Kingston, & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 283–333). Cambridge: Cambridge University Press. URL: https://doi.org/10.1017/CBO9780511627736.017. doi:10.1017/CBO9780511627736.017.

Daniloff, R., & Hammarberg, R. (1973). On defining coarticulation. *Journal of Phonetics*, *1*, 239–248. URL: https://doi.org/10.1016/S0095-4470(19)31388-9. doi:10.1016/S0095-4470(19)31388-9.

Davidson, L., & Erker, D. (2014). Hiatus resolution in American English: The case against glide insertion. *Language*, *90*, 482–514. URL: https://doi.org/10.1353/lan.2014.0028. doi:10.1353/lan.2014.0028.

Deme, A. (2016). *A magánhangzók ejtése és észlelése a szopránéneklésben*. Budapest: ELTE Eötvös Kiadó.

Deme, A., Bartók, M., Csapó, T. G., Gráczi, T. E., Juhász, K., & Markó, A. (2022). A magánhangzók centralizációja és produkciós homogenitása az

előrefelé és hátrafelé ható magánhangzók közti koartikulációban – artikulációs és akusztikai adatok. *Általános Nyelvészeti Tanulmányok*, *34*, 15–49.

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

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America*, *63*, 223–230. URL: `https://doi.org/10.1121/1.381717`. doi:`10.1121/1.381717`.

Guevara-Rukoz, A. (2018). Decoding perceptual vowel epenthesis: Experiments & modelling. PhD disszertáció.

Gósy, M. (2004). *Fonetika, a beszéd tudománya*. Budapest: Osiris.

Gósy, M. (2014). A palatális közelítőhang kétféle funkcióban. *Beszédkutatás*, *22*, 17–40.

Haas, W. G. d. (1988). A formal theory of vowel coalescence: A case study of ancient Greek. PhD disszertáció.

Heselwood, B. (2006). Final schwa and r-sandhi in RP English. *Leeds Working Papers in Linguistics and Phonetics*, (pp. 78–95).

Hunt, E. (2003). Acoustic characterization of the glides /j/ and /w/ in American English. PhD disszertáció.

Jaggers, Z. (2018). Evidence and characterization of a glide-vowel distinction in American English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*, 1–27. URL: `https://doi.org/10.5334/labphon.36`. doi:`10.5334/labphon.36`.

Juhász, K., & Deme, A. (2022a). Mandarin kínai és magyar szibilánsok palatalizációja kínaiul tanuló magyar anyanyelvűek ejtésében. *Alkalmazott Nyelvtudomány*, *XXII*, 64–89.

Juhász, K., & Deme, A. (2022b). Palatális approximánsok a kínaiban és a magyarban – a kínai alveolopalatális /ɕ/ szibilánst követő vokális szakasz

produkciója kínaiul tanuló magyar anyanyelvűeknél. In K. Mády, & A. Markó (Eds.), *Általános Nyelvészeti Tanulmányok XXXIV* (pp. 287–332). Budapest: Akadémiai Kiadó.

Ladefoged, P. (1975). *A Course in Phonetics*. New York: Harcourt.

Lehiste, L., & Peterson, G. (1961). Transitions, glides, and diphthongs. *The Journal of the Acoustical Society of America*, *33*, 268–277. URL: https://doi.org/10.1121/1.1908638. doi:10.1121/1.1908638.

Magen, H. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *25*, 187–206. URL: https://doi.org/10.1006/jpho.1996.0041. doi:10.1006/jpho.1996.0041.

Markó, A. (2012). Boundary marking in Hungarian v(#)v clusters with special regard to the role of irregular phonation. *The Phonetician*, *105-106*, 7–26.

Menyhárt, K. (2006). Koartikulációs folyamatok két magánhangzó kapcsolatában. *Beszédkutatás*, (pp. 44–56).

Öhman, S. (1966). Coarticulation in vcv utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*, 151–168. URL: https://doi.org/10.1121/1.1909864. doi:10.1121/1.1909864.

Parker, S. (2011). Sonority. In M. Oostendorp (Ed.), *The Blackwell companion to phonology* (pp. 1195–1220). Malden: Wiley-Blackwell. URL: https://doi.org/10.1002/9781444335262.wbctp0049. doi:10.1002/9781444335262.wbctp0049.

Picard, M. (2003). On the emergence and resolution of hiatus. *Folia Linguistica Historica*, *24*, 44–57. URL: https://doi.org/10.1515/flih.2003.24.1-2.47. doi:10.1515/flih.2003.24.1-2.47.

Pulleyblank, D. (1986). Underspecification and low vowel harmony in Okpe. *Studies in African Linguistics*, *17*, 119–153. URL: https://doi.org/10.32473/sal.v17i2.107490. doi:10.32473/sal.v17i2.107490.

R Development Core Team (2021). R foundation for statistical computing, 4.0.5-ös verzió. URL: https://www.R-project.org/.

Recasens, D. (2013). On the articulatory classification of (alveolo)palatal consonants. *Journal of the International Phonetic Association*, *43*, 1–22. URL: https://doi.org/10.1017/S0025100312000199. doi:10.1017/S0025100312000199.

van Rij, J., Wieling, M., Baayen, H., & van Rijn, D. (2020). itsadug: Interpreting time series and autocorrelated data using GAMMs. 4.1.2-es R csomagverzió. (Letöltés ideje: 2021. november 1.).

Siptár, P. (2002). Hiátus. In L. Hunyadi (Ed.), *Kísérleti fonetika, laboratóriumi fonológia* (pp. 85–98). Debrecen: Debreceni Egyetem Kossuth Egyetemi Kiadója.

Siptár, P. (2005). A magánhangzó-kapcsolatok fonológiájából. *Magyar Nyelv*, *101*, 282–304.

Siptár, P. (2011). Alakváltozatok, allomorfok, alternációk. *Magyar Nyelv*, *107*, 147–160.

Siptár, P. (2013). Palatálisok. In A. Benő, E. Fazakas, & E. Kádár (Eds.), *„...hogy legyen a víznek lefolyása...": Köszöntő kötet Szilágyi N. Sándor tiszteletére* (pp. 433–448). Kolozsvár: Erdélyi Múzeum Egyesület. URL: http://archive.nytud.hu/oszt/elmnyelv/siptar/publ/palatalisok.pdf.

Siptár, P. (2016). A mássalhangzók. In F. Kiefer (Ed.), *Strukturális Magyar Nyelvtan II. kötet – Fonológia [Digitális kiadás*. Budapest: Akadémiai Kiadó. URL: https://mersz.hu/dokumentum/m26smny2__113/ elérhető:.

Siptár, P., & Törkenczy, M. (2000). *The phonology of Hungarian*. Oxford: Oxford University Press.

Stevens, K. (2000). *Acoustic Phonetics*. Massachusetts: MIT Press. URL: https://doi.org/10.7551/mitpress/1072.001.0001. doi:10.7551/mitpress/1072.001.0001.

The Audacity Team (2023). Audacity. 2.4.-es verzió. URL: http://audacityteam.org/ (Legutolsó hozzáférés: 2023. december 14.).

Trask, R. (1996). *A Dictionary of Phonetics and Phonology*. London: Routledge.

Uffmann, C. (2007). Intrusive [r] and optimal epenthetic consonants. *Language Sciences*, *29*, 451–476.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. URL: https://doi.org/10.1016/j.wocn.2018.03.002. doi:10.1016/j.wocn.2018.03.002.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R*. Boca Raton: Chapman and Hall. URL: https://doi.org/10.1201/9781315370279. doi:10.1201/9781315370279.

Az anyanyelvre gyakorolt célnyelvi hatás gyengülésének kérdése a növekvő számú anyanyelvi ingerek hatására

Juhász Kornélia[1,2]

[1] *HUN–REN Nyelvtudományi Kutatóközpont*
[2] *Eötvös Loránd Tudományegyetem*

Abstract

This acoustic analysis is focused on how an atonal L1 and a tonal L2 interact in the case of Hungarian learners of Mandarin Chinese. In particular, this experiment intends to shed light on whether L2 Chinese tonal patterns' effect weakens on L1 Hungarian intonation contours throughout the experiment, as the production of L1 utterances increases. It was hypothesized that at the beginning of the Hungarian L1 recordings, language learners' production is primarily shaped by the L2-dominant bilingual mode, thus L1 Hungarian intonation patterns approximate the L2 Chinese tonal curves in their shape. However, throughout the recording, as language learners produce more L1 utterances, their production is hypothesized to approach the standard native L1 patterns gradually due to the weakening of the L2 tonal effect. Since we expected that L2 tonal effects also depend on language learners' L2 experience, we analysed two speaker groups with different levels of L2 experience. The effect of the L2 tones was analysed by the f_0 curve and the duration of the vocalic section of the monosyllabic utterances was recorded in four different L1 tunes: declarative, imperative, and two interrogative intonation patterns. Statistical analysis was submitted to GAMMs, where the f_0 change was analysed along the vocalic section's normalized duration, as well as throughout the recording by aligning the utterances along their ordinal number. Our results did not confirm the gradual weakening of the L2 effect on L1 intonation patterns but rather suggest that the sudden change between L1 and L2 induces a more dynamic excursion towards the L1 language mode, which is followed by a return to the L2-dominated language mode approximating L2 tonal patterns. Regarding these results, questions arise whether longer recordings with more utterances would show a different outcome regarding the weakening of the L2 effect on L1 intonation patterns. The results of the experiment also contribute to the deeper understanding of which acoustic features Hungarian native speakers enhance along repetitions of the same L1 sentence type in monosyllabic utterances.

1. Bevezetés

A tanulmány középpontjában a célnyelv (L2) anyanyelvre (L1-re) gyakorolt hatásának kérdése áll. A mandarin kínaiul tanuló magyar anyanyelvűek ejtésé-

Email address: juhasz.kornelia@nytud.hun-ren.hu (Juhász Kornélia)

Beszédtudomány – Speech Science 2024/1. 42–84. DOI: 10.15775/Besztud.2024.1.42-84

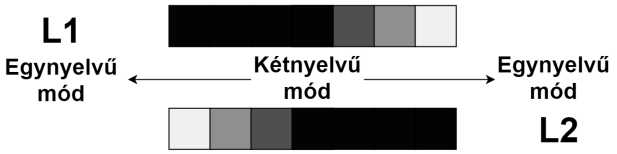
ben vizsgálom azt, hogy a kínai tónusprodukció anyanyelvre gyakorolt hatása változik-e annak függvényében, hogy a kísérleti személyek a felvétel rögzítése során mennyi anyanyelvi (azaz magyar) ingert produkálnak. Ez azt jelenti, hogy magyar intonációs dallamkontúrokon a normalizált időtartamban bekövetkezett f_0-változást vizsgálom az idő előrehaladásának függvényében. A kísérletben négy különböző magyar dallam realizációját figyelem meg: az ereszkedő dallamok esetében a kijelentő és felszólító dallamokat, míg az emelkedő dallamok közül pedig az egy szótagú kérdést és az alternatív kérdés emelkedő fázisát elemzem (Varga, 1994). A vizsgálatban az L2 nyelvi tapasztalatnak is jelentőséget tulajdonítok, ezért két különböző, egy kezdő és egy haladó nyelvtanuló csoport produkcióját is összevetem. A fentebbi kérdés alapjait, azaz a dallammenetek megvalósulásában megfigyelhető célnyelvi hatást már egy korábbi tanulmányban statikus szempontból vizsgáltam és megerősítettem (l. Juhász, megj.). Azonban a korábbi eredmények értelmezésekor nem vettem figyelembe azt a dinamikus szempontot, hogy a célnyelvi hatás az anyanyelvi ingerek befolyására gyengülhet. Ezért e tanulmány célja, hogy a korábbi elemzést a fentebbi új szemponttal, azaz a kísérletben eltelt idővel, tehát a növekvő számú anyanyelvi megnyilatkozások hatásával egészítse ki. A tanulmány jelentősége továbbá az is, hogy a magyar kontrollcsoport esetében betekintést nyerhetünk abba, hogy a vizsgált monoszillabikus intonációs kontúrok esetében a beszélők milyen akusztikai tulajdonságokra erősítenek rá az ismétlések mentén.

1.1. Szakirodalmi háttér

E tanulmány előzményeként – a jelen esetben is elemzett adatsoron – statikus szempontból vizsgáltam azt, hogy a kétnyelvű személyek elméjében megjelenő két nyelv kapcsolata és interakciója hogyan befolyásolja a produkciót, pontosabban abból a hipotézisből indultam ki, hogy nemcsak az anyanyelv lehet hatással a célnyelvi ejtésre, hanem ez a kölcsönhatás fordított irányban is fennáll, azaz a célnyelv is befolyásolhatja az anyanyelvi produkciót. A hipotézis tesztelésekor az L2 L1-re gyakorolt hatásának vizsgálatában a Cook (2006) által meghatározott multikompetens kognitív nyelvi rendszerből, illetve a Grosjean-féle nyelvimód-

hipotézisből indultam ki. Cook (2006) szerint a többnyelvűek esetében (amely terminus magába foglalja a nyelvtanulókat is) megjelenő kognitív rendszer nyelvi szempontból multikompetensnek tekinthető, tehát a nyelvtanuló elméjében a nyelvek, az L1 és az L2 (pontosabban az L2 esetében a köztes nyelv) nem egymástól elkülönülten jelenik meg, hanem ugyanazon kognitív rendszer részeként össze vannak kapcsolva és hatással vannak egymásra (Cook, 2006). Ebben a specifikus esetben az L1 és az L2 kapcsolatát kései nyelvtanulók, azaz olyan fiatal felnőttek esetében vizsgálom, akiknek az L1-ük már kiforrott belső struktúrával rendelkezik a nyelv minden szintjén, és ez után következik be az L2 elsajátítása. Ebből fakadóan a nyelvtanulók esetében a két nyelv közötti kompetencia nincs egyensúlyban, ennek ellenére az ő esetükben is kétnyelvűségről beszélünk, ahol a köztes nyelvet (azaz a nyelvtanuló saját elméjében megjelenő L2-t) mind a nyelvtanuló L1-e, mind a környezeti L2-ingerek együtt alakítják (Major, 2001; Cook, 2006). Tehát a nyelvek a kétnyelvű személy elméjében kapcsolatban állnak, azonban az aktivációs szintjük nem tekinthető állandónak, hanem a környező behatásokhoz adaptálódva folyamatosan változik (Green, 1986). Ezért az összekapcsolt rendszerből fakadóan az L1 és az L2 kölcsönösen befolyásolhatja egymás nyelvi feldolgozását, percepcióját és produkcióját (Pavlenko, 2000; Grosjean, 2001, 2008; Cook, 2003; Hammarberg, 2014). E kölcsönös egymásra hatást az előzménykutatásomban (és a jelen tanulmány esetében is) a nyelvi mód fogalmán keresztül ragadom meg. Grosjean (1998; 2001) nyelvimód-hipotézise szerint a kétnyelvű személy adott időpillanatban a környező pszichológiai és nyelvi tényezőktől függően eltérő szinteken aktiválhatja az L1-et és az L2-t az elméjében (Grosjean, 2001, 3). Grosjean (2001) hipotézise alapján a nyelvi mód egy kontinuumként képzelhető el, ami két abszolút végponttal rendelkezik: ezek a végpontok az egynyelvű módok. Az egynyelvű módok esetén az adott személy elméjében csak az egyik nyelv, azaz vagy csak az L1, vagy csak az L2 aktív, míg a másik nyelv deaktivált állapotban van (1. ábra). A környezet ingereinek hatására a két nyelv egyszerre is aktiválódhat: ekkor a beszélő elméje kétnyelvű módba vált. Azonban a két nyelv aktivitása nem tekinthető egyensúlyi helyzetnek: Grosjean szerint a kétnyelvű módban az egyik nyelv mindig dominánsabban

jelenik meg a másikhoz képest, és a domináns nyelvet mátrixnyelvnek nevezzük. A két nyelv együttes aktiválása (azaz a kétnyelvű mód) olyan köztes értékeket eredményez a percepcióban és a produkcióban, melyek az L1-ben és az L2-ben (egynyelvű módban) tapasztaltak között helyezkednek el (Grosjean, 2001). Ez azt jelenti, hogy ha a kísérleti személy kevert nyelvi ingereket észlel, azaz az elméjében mindkét nyelv aktiválódik (azaz az elméje kétnyelvű módba vált), akkor az L1-hez és az L2-höz képest köztes, átmeneti értékeket kaphatunk.



1. ábra. A Grosjean által javasolt nyelvi módok (ahol az aktivált állapotot a sötétebb, míg a deaktivált állapotot a világos szín jeleníti meg, Grosjean, 1998, 136 alapján).

Tehát a jelen kísérlet célja az volt, hogy a Grosjean-féle nyelvimód-hipotézis paradigmáját alapul véve a nyelvi módok kontrollálásával facilitáljuk azt, hogy a nyelvtanuló kísérleti személyek elméjében az anyanyelvi produkció ideje alatt az L2 dominálta kétnyelvű mód jelenjen meg. Tehát ezáltal az L1 mintázatok megvalósulásán keresztül vizsgálhatóvá váljon az L2 hatása. A kísérletben résztvevő nyelvtanulók kínaiul tanuló magyar anyanyelvű beszélők, így a magyar nyelv tekintendő L1-nek, míg a mandarin kínai L2-nek. Ebből következtetve a mandarin kínai L2 tónusok hatását vizsgálom a magyar L1 monoszillabikus intonációs kontúrok realizációjában. Habár a mandarin kínai nyelvben a szótaghoz rendelt hangmagasság-változás (avagy a tónus) a szótag argumentuma és lexikai szinten határozza meg a jelentést, a magyar nyelvben a hangmagasság-változás, avagy intonáció révén szándékot vagy érzelmet tudunk kifejezni (Chao, 1948/1963; Gósy, 2004). Annak ellenére, hogy a kínai lexikai tónusok és a magyar intonáció nyelvi funkciójukban eltérnek, mégis ugyanazon fiziológiai folyamat révén jönnek létre: a hangszalagok vibrációjának eredményeképpen. A hangszalagok rezgése az alapfrekvenciával (f_0-val) jellemezhető, ami az észlelt hangmagas-

sággal áll logaritmikus összefüggésben, és ezért a Hertz-értékeket félhangokká szokás konvertálni ('t Hart et al., 1990). Emellett megjegyzendő az is, hogy a mandarin kínai esetében a lexikai tónusok realizációját az intonáció is befolyásolja (Shen, 1990), azonban e tanulmány esetében a kínai intonáció hatása nem tekinthető relevánsnak. Tehát mint fentebb említettem, a kínai tónusok és a magyar intonáció két különböző nyelvi funkciót lát el, mégis ugyanazon artikulációs-motoros mechanizmus (tehát a hangszalagrezgés frekvenciájának modulációja) révén jönnek létre. Mint az széles körben ismert, a beszédprodukció motoros beállításai a beszélő elméjében kódolva vannak és a különböző nyelvekben eltérő artikulációs mintázatok produkciójára lehet szükség. Ezen különböző artikulációs mintázatok közötti váltás, például az adott L2-re jellemző artikulációs beállítás elérése, vagy a hirtelen váltás egyik nyelvről a másikra – főleg abban az esetben, ha ez az artikulációs folyamat nincs kellően automatizálva – problémát okozhat a nyelvtanulók számára (vö. Leather & James, 1991). Tehát ezek alapján habár a tónusok és az intonáció produkciója két különböző nyelvi működésnek tűnhet, fiziológiai-motoros szempontból azonban mégis ugyanazon artikulációs folyamat modulálásával áll összefüggésben, ami nyelvspecifikus kihasználtságából fakadóan szoros összefüggésben áll a nyelvi módokkal.

A tanulmány szempontjából érdemes arra is kitérni, hogy a kísérlet felépítése, azaz az ismétlések produkciója hogyan befolyásolják az ejtés akusztikai tulajdonságait, és néhány szót szólnunk kell az ismétlési redukció fogalmáról is. Abban az esetben, ha adott megnyilatkozás többször hangzik el, akkor az első ejtéshez képest a második ismétlésnél már redukció léphet fel a beszédhangok időtartamában, akusztikai tulajdonságaikban (periferikusságában), illetve dallamívében is (vö. Jacobs et al., 2015). Ezért a produkciós feladatokban a megnyilatkozások ismételt produkciójával, azaz „begyakorlásával" párhuzamosan arra számíthatunk, hogy a feladat egyre könnyebbé és egyértelműbbé válik a beszélő számára, így egyre nagyobb eséllyel számíthatunk redukcióra is az ejtésben (ebben a specifikus esetben a dallamív realizációja szempontjából) (Gahl et al., 2012). Azt azonban fontos kiemelni, hogy a jelen kísérletben redukció

esetén nem várhatjuk a kontraszt teljes hiányát az eltérő jelentéssel rendelkező megnyilatkozások között, hiszen ebben az esetben a kísérlet módszertana erősen facilitálja az olyan minimális párok produkcióját, amik kizárólag intonációs tónusukban, azaz dallammenetükben térnek el egymástól. Ebből fakadóan arra számíthatunk, hogy a redukció elsősorban azon akusztikai jegyeket érinti, amelyek kevésbé relevánsak a kontraszt megvalósításában. Tehát ebből kiindulva a kísérlet eredményein keresztül információt nyerhetünk arról is, hogy a magyar kontrollcsoport esetében – akiknek az ejtését nem befolyásolja idegen nyelvi hatás – milyen folyamatok játszódnak le az ismétlések hatására, és a magyar intonációs kontúrok mely jegyei válnak prominensebbé, és melyek esnek áldozatul a redukciónak.

A tanulmány középpontjában – mint fentebb említettem – az L2 L1-re gyakorolt hatása áll a nyelvimód-hipotézis szemszögéből. A nemzetközi szakirodalomban az L1 és az L2 egymásra hatását érintő kísérletek között az L1 L2-re gyakorolt hatását érintő diskurzusok vannak többségben, azonban az ennek a folyamatnak a fordítottja (tehát a jelen tanulmány fókuszában álló folyamat, azaz az L2 hatása) is relatíve kutatott területté vált az utóbbi évtizedekben (de Leeuw et al., 2011). Azonban azt fontos kiemelni, hogy ez a téma elsősorban olyan esetekben kerül az elemzések középpontjába, ahol az L2-vel való kapcsolat hosszú távú hatását vizsgálják, azaz ahol az L2 válik a beszélők elsődleges, minden nap használt nyelv(változatá)vá, aminek hatására az L1 használati gyakorisága visszaszorul (de Leeuw et al., 2011). Továbbá ha a nyelvi mód hatását vizsgáló pillanatnyi nyelvi aktivitásra vonatkozó eredményeket vesszük középpontba, akkor azok elsősorban retrospektíve fogalmaznak meg kritikákat a nem megfelelő kísérleti módszertanokkal kapcsolatban (pl. Elman et al., 1977; Grosjean, 2001; Schwartz et al., 2015). Továbbá a nyelvi módot vizsgáló kutatások elsősorban pszicholingvisztikai vagy percepciós vizsgálatokat foglalnak magukba, azaz lexikai szinten szótalálási feladatokat, beszédhang-identifikációs kísérleteket vagy produkciós szempontból kódváltási szituációkat elemeznek (vö. Grosjean, 2001; Wu et al., 2018; Yu & Schwieter, 2018). A szerző tudomása szerint nem született még a jelen tanulmány középpontjában álló szupraszegmentális vizsgálathoz ha-

sonló elemzés, azaz a kínai L2 tónus-kontúrok visszahatása valamilyen atonális L1 nyelv intonációs dallamainak megvalósulására.

1.2. A tanulmány előzményei

Ebben a kísérletben a nyelvi módok kontrollálása, avagy az L2-dominálta kétnyelvű nyelvi mód előhívása az L1 produkcióban a felvételek sorrendjével állt szoros összefüggésben. A kísérlet első részében az volt a célom, hogy egy kínai nyelvű produkciós feladat révén a nyelvtanulók elméjét lehetőség szerint legjobban ráhangoljam a kínai L2 egynyelvű módra, amit egy kínai anyag felolvasásával igyekeztem elérni. Tehát a kísérlet ezen részében azt igyekeztem megvalósítani, hogy a nyelvtanulók elméjében az L2 legyen a mátrixnyelv, ez vezérelje a produkciót és ezáltal a nyelvtanulók ejtése a lehető legjobban megközelítse a natív mintázatot. Ezután a kísérlet második részében a kínai nyelvi produkció után átmenet nélkül rögtön a magyar nyelvű anyag felolvasása következett, ahol azt vártam, hogy a nyelvtanulók elméje olyan kétnyelvű módba kerül, ahol az L2 vezérli elsősorban a produkciót, de az L1 is aktivált ejtésben jelenik meg. Tehát összegezve a kísérlet alaphipotézise az volt, hogy ha a kísérletben kontrolláljuk a nyelvi módokat, akkor a nyelvtanulók elméjében megjelenő kétnyelvű mód (amelyben az L2 a domináns mátrixnyelv) hatással van az anyanyelvi produkcióra, mégpedig úgy, hogy a nyelvtanulók a magyar intonációs kontúrok ejtésekor a kínai tónusok mintázatait közelítik meg ejtésükben. Továbbá azt is feltételeztem, hogy a nyelvi tapasztalat befolyásolja a kétnyelvű mód, azaz a célnyelvi hatás megjelenését: arra számítottam, hogy a haladó nyelvtanulók jobban eltérnek a sztenderd magyar ejtéstől a kínai tónusok irányába, mint a kezdők. Ezen alhipotézis alapjául az szolgált, hogy a kezdő nyelvtanulók ejtését elsősorban az L1-transzfer határozza meg (Major, 2001; Flege, 2022), ebből fakadóan az anyanyelvi ejtésükben sem várható jelentős célnyelvi hatás, hiszen az anyanyelvi és célnyelvi mintázatok között nincs nagy eltérés. Ezzel szemben a haladók produkciójában már jelentős kontrasztot vártam a célnyelvi és anyanyelvi mintázatok között, hiszen az L1-transzfer már kevéssé alakítja az ejtést (Major, 2001; Flege, 2022), ezért az ő esetükben

a kezdőknél jelentősebb célnyelvi hatásra számítottam, mégpedig úgy, hogy a haladó nyelvtanulók magyar intonációs kontúrjai a kínai tónusokat megközelítő, sztenderd magyar ejtéstől jobban eltérő mintázatokkal realizálódnak.

Arra vonatkozóan, hogy a nyelvtanulók magyar intonációs kontúrjain milyen természetű L2-hatást vártam, azaz hogy a magyar intonációs kontúrok milyen akusztikai tulajdonságaikban közelítik meg a kínai tónusokat, egy korábbi kísérletem eredményeit használtam fel (l. Juhász, 2023), és ezen eredmények alapján állítottam fel hipotéziseket a nyelvtanulók kétnyelvű módú ejtéséről. A kínai tónusok és a magyar intonációs kontúrok összehasonlításában a magyar egy szótagú eldöntendő kérdést és az alternatív kérdés emelkedő monoszillabikus tagját a kínai emelkedő 2. tónussal vetettem össze, valamint az egy szótagú felszólító és a kijelentő dallamot pedig a kínai 4. tónussal hasonlítottam össze. Ezért a következőkben először bemutatom a vizsgált magyar és kínai dallamok akusztikai jellemzőit. Majd ezután összefoglalom, hogy a kínai és magyar kontrollcsoportok ejtésében a kínai és magyar dallamívek hogyan realizálódtak egymáshoz képest, tehát miben hasonlítanak és miben térnek el egymáshoz képest (Juhász, 2023 alapján). Ezt követően végül azt foglalom össze, hogy a magyar kontrollcsoporthoz képest a nyelvtanulók produkciójában milyen eltérésekre számítottam, és ezeket a feltevéseket az eredmények megerősítették-e (Juhász, megj. alapján).

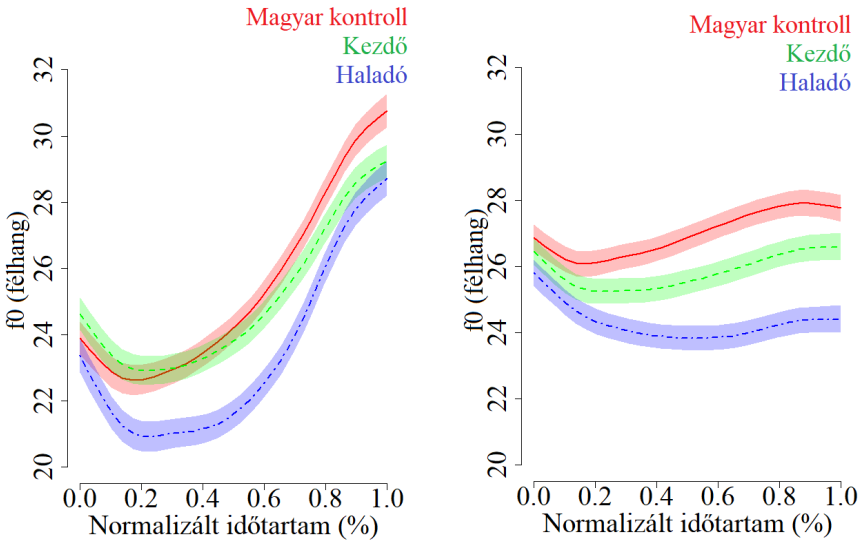
A vizsgált kínai és magyar dallamok akusztikai tulajdonságait középpontba véve az emelkedő dallamok közül a mandarin kínai 2. tónust, a magyar egy szótagú eldöntendő kérdést és a magyar alternatív kérdés emelkedő fázisát vizsgáltam. E dallamok mind mögöttes LH célokkal rendelkeznek, amely struktúra a magyar egy szótagú eldöntendő kérdés esetében az emelkedő-ereszkedő szerkezet trunkációjával valósul meg (Varga, 1994). Ehhez hasonlóan az alternatív kérdés első tagjának dallamát is emelkedő fázis jellemzi, azonban ebben az esetben egy kérdőszavas konstrukció részeként megjelenő realizációban ennek megvalósulását az egy szótagú eldöntendő kérdéstől eltérő mintázattal várhatjuk, hiszen ebben az esetben nem kizárólag prozódiai eszközök állnak rendelkezésre a kérdő szándék kifejezésére (vö. Olaszy, 2002; Juhász, 2023). A kínai 4. tónus és a magyar kijelentő és felszólító dallam egyaránt ereszkedve valósul meg (HL),

azonban Fónagy és Magdics (1967) leírása alapján a felszólító dallam a magyar kijelentőhöz képest egy szekunddal magasabb hangmagasságból indul ki. Ez azt jelenti, hogy a felszólító dallam kezdeti fázisában prominensen magasabb (H) akusztikai célra számíthatunk, míg a kijelentő dallam realizációjában inkább az alacsony f_0-értékek dominálnak, tehát az alacsony (L) tonális cél kap jelentősebb szerepet (Juhász, 2023). Tehát ebből kiindulva a kérdő dallamok megvalósulásában az emelkedés H-s specifikációja, míg az ereszkedő dallamok esetében a felszólítás magasabb tartományból való ereszkedése, és a kijelentés alacsonyabb f_0-tartományban való megvalósulása látszik a dallammenetek közötti szembenállás elsődleges akusztikai kifejeződésének lenni.

A kínai és magyar kontrollcsoportok ejtésében összehasonlítva a sztenderd kínai és magyar dallamokat a következő eredmények születtek: a kínai 2. emelkedő tónust a magyar egy szótagú kérdés emelkedő dallamával összevetve, a kínai 2. tónus mintázata a magyar egy szótagú kérdéshez képest homorúbb dallamívvel és magasabb maximális f_0-lal realizálódott. Ezért a nyelvtanulók ejtésében a kínai 2. tónus hatására bekövetkező akusztikai változást a sztenderd magyar ejtéshez képest homorúbb dallamívben és alacsonyabb maximális f_0-jában vártam. E hipotézis a dallammenet alakját tekintve megerősítést nyert, hiszen mindkét kínaiul tanuló csoport a sztenderd magyar ejtésnél homorúbb, öblösebb f_0-görbét produkált, ami azt jelenti, hogy a normalizált időtartamban az emelkedő fázis csak késleltetve jelent meg a sztenderd ejtéshez képest, és a görbék jelentősebb része realizálódott alacsony f_0-tartományban (2. ábra, bal). A nyelvi tapasztalat hatása e dallam produkciója esetében megerősítést nyert oly módon, hogy a kezdő nyelvtanuló csoporthoz képest a haladók mind az f_0-ban talált differenciában, mind a görbék átfedésének mértékében jobban eltértek a magyar kontrollcsoport ejtésétől.

A kínai 2. tónus a magyar alternatív kérdés emelkedő fázisához képest homorúbb f_0-menettel és magasabb f_0-minimummal valósult meg. Ebből kiindulva a nyelvtanulók ejtésében a 2. tónus hatását a sztenderd magyar ejtésnél homorúbb dallamívben és alacsonyabb minimális f_0-értékben vártam. E hipotézis mindkét nyelvtanuló csoport ejtésében megerősítést nyert, azaz mindkét nyelvtanuló

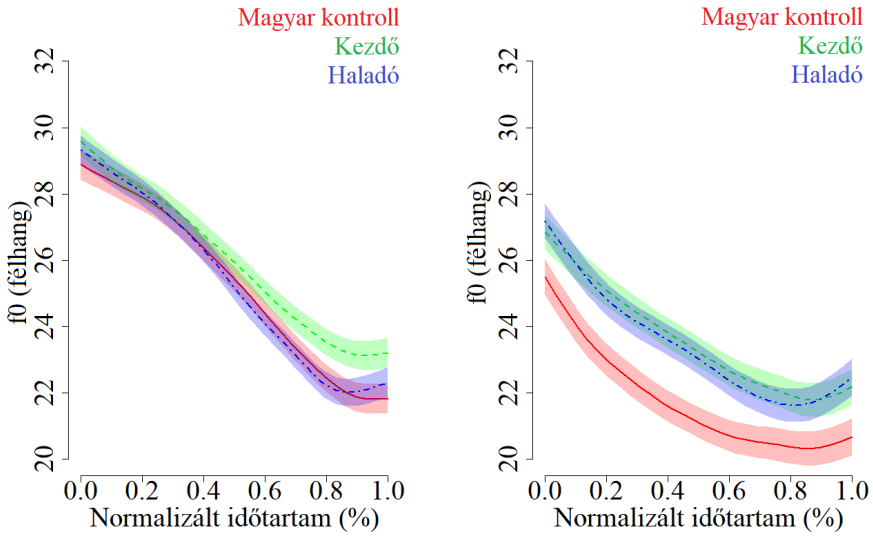
csoport a kontrollcsoportnál homorúbb görbét produkált alacsonyabb minimális f_0-értékkel (2. ábra, jobb). A nyelvi tapasztalat hatása ebben az esetben is megerősítettnek tekinthető: a kezdők jobban megközelítették a sztenderd magyar ejtést, és csak a normalizált időtartam legelején produkáltak átfedést a kontrollcsoport ejtésével. Míg ezzel szemben a haladók semmilyen egyezést nem mutattak a kontrollcsoport f_0-kontúrjával, és az ő produkciójuk rendelkezett a csoportok között a legalacsonyabb minimális f_0-értékkel.



2. ábra. A magyar egy szótagú eldöntendő kérdő (balra), alternatív kérdő (jobbra) dallamok f_0-görbéi a magyar kontrollcsoport (piros), a kezdő (zöld) és a haladó (kék) kínaiul tanulók ejtésében (Juhász, megj.).

A felszólító dallamot a kínai ereszkedő 4. tónussal natív beszélők ejtésében összehasonlítva, a maximális és minimális f_0-értékében nem, kizárólag a dallam ívében találtam eltérést e két dallam között. A kínai 4. tónus kizárólag a domborúbb dallamívében tért el a magyar felszólító dallamhoz képest. Ebből következően a nyelvtanulók ejtésében nem vártam célnyelvi hatást megjelenni a felszólító dallam minimum és maximum f_0-értékében – hiszen ezekben az akusztikai tulajdonságokban a magyar és a kínai dallamok nem térnek el. Azonban a dallam ívét a nyelvtanulók ejtésében a sztenderd magyar ejtéshez viszonyítva

domborúbbnak vártam. A nyelvtanulók a minimális és maximális f_0-értékekben – a vártakkal megegyezően – nem tértek el a sztenderd magyar ejtéstől, azonban a hipotézis domborúbb dallammenetre vonatkozó része nem nyert megerősítést, hiszen mindhárom vizsgált beszélői csoport átfedő, lineáris görbéket produkált (3. ábra, bal).



3. ábra. A magyar felszólító (balra) és kijelentő (jobbra) dallamok f_0-görbéi a magyar kontrollcsoport (piros), a kezdő (zöld) és a haladó (kék) kínaiul tanulók ejtésében (Juhász, megj.).

A kijelentő dallam a kínai ereszkedő 4. tónushoz képest alacsonyabb f_0-tartományban valósult meg, tehát alacsonyabb maximális és minimális f_0-értékkel, valamint a 4. tónusnál homorúbb dallamívvel. Ezért a nyelvtanulók esetében a kétnyelvű mód hatására a sztenderd magyar ejtéshez képest magasabb maximális és minimális f_0-értéket és domborúbb dallamívet vártam. E hipotézis a 4. tónus hatásáról megerősítést nyert: a nyelvtanulók kijelentő módú dallamívei a sztenderd magyar ejtéshez képest magasabb f_0-tartományban, tehát magasabb maximális és minimális f_0-értékkel valósultak meg (3. ábra, jobb). A nyelvtanulók lineáris görbéi tulajdonképpen a kijelentés homorú dallamívéhez képest „domborúbbnak” tekinthetők, tehát ebből a szempontból a 4. tónus

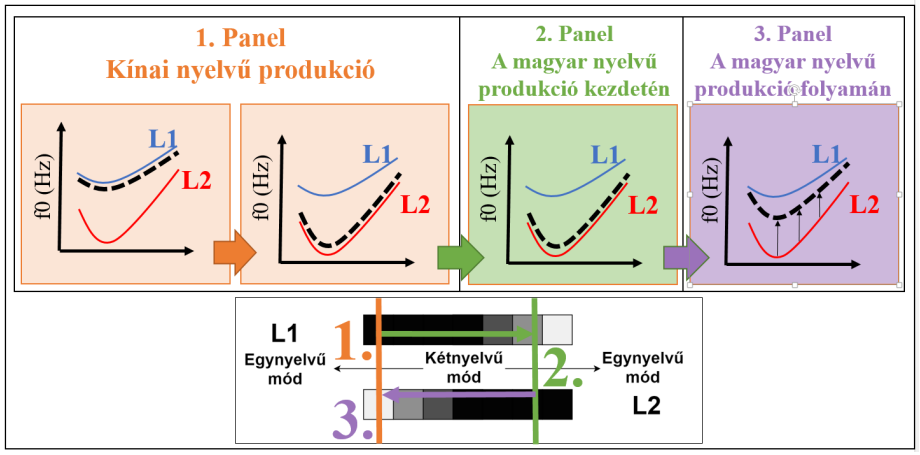
hatását megerősítettem. Azonban a nyelvi tapasztalat hatásával kapcsolatos feltételezések ebben az esetben nem bizonyultak helytállónak, mert a két nyelvtanuló csoport átfedő görbéket produkált, tehát a kezdők nem közelítették meg jobban a sztenderd magyar ejtést, mint a haladók.

1.3. Hipotézisek

A bemutatott eredmények alapján az L2-dominálta kétnyelvű mód hatása megerősítést nyert a kínaiul tanuló magyar anyanyelvűek anyanyelvi intonációskontúr-produkciójában, tehát a kísérlet módszertanával sikerült a kínaiul tanulók elméjét a kínai nyelvű produkciós feladattal a kínai egynyelvű mód felé mozdítani (4. ábra 1. panel). Továbbá ezután a dallamívek megvalósulása megerősítette azt a nyelvimód-beállítást, hogy a nyelvtanulók elméjében sikerült egy olyan kétnyelvű módot előhívni, ahol mind az L1, mind az L2 aktivált állapotban mutatkozik, de az L2 vezérli dominánsan a produkciót, amely célnyelvi hatás révén az ejtésükben a kínai tónusok dallammintázatait közelítik meg. Emellett fontos kiemelni, hogy a célnyelvi hatás csak abban az esetben jelentkezhetett, ha az L1 és L2 mintázat között különbség volt felfedezhető, valamint a realizációját a nyelvi tapasztalat is befolyásolta. Ezen eredményekben azonban az L2-dominálta kétnyelvű módú produkciót egy statikus szempontból ragadtam meg, azaz a kísérlet összes megnyilatkozását egyetlen dallamkontúr formájában vizsgáltam, és nem vettem figyelembe azt a tényezőt, hogy a nyelvi módok változása az elmében egy környezeti ingerekhez kötött adaptív folyamat. Ezért a jelen elemzés célja az, hogy az adatok elemzését az eddigi vizsgálati szempontok mellett egy dinamikus szemponttal is kiegészítsem, ezáltal figyelembe vegyem és megragadjam a kísérlet előre haladását, és a nyelvtanulók által produkált növekvő számú anyanyelvi magyar megnyilatkozást, és ennek nyelvi módokra gyakorolt hatását. A kísérletben tehát arra a kérdésre keresem a választ, hogy a növekvő számú magyar megnyilatkozás hatására a kínaiul tanuló magyar anyanyelvűek produkciójában a célnyelvi hatás gyengülést mutat-e.

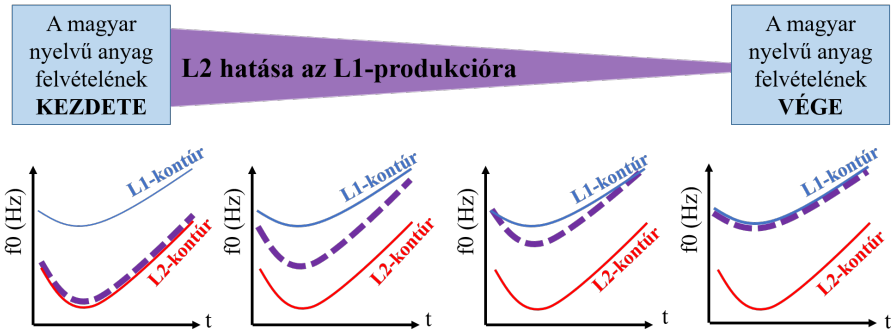
A jelen vizsgálatban tehát azt várom, hogy a magyar nyelvű produkciós feladat kezdetén a nyelvtanulók ejtésében a magyar dallamkontúrok a kínai tó-

nusokra jellemző mintázatokkal jelennek meg (4. ábra. 2. panel). És ezt követően – ahogy egyre több magyar nyelvű megnyilatkozást produkálnak, tehát ahogy egyre inkább hangolódnak „vissza” az L1-re – az ejtésükben annál inkább megközelítik az anyanyelvre jellemző sztenderd dallammintázatokat (4. ábra 3. panel). Tehát összegezve arra számítok, hogy a kínaiul tanulók esetében időben minél több L1-produkció előzi meg az adott L1 megnyilatkozást, az L1-es ingerek hatására az L1 aktivációja egyre nő, és ezért az L1 egyre dominánsabban szabályozza a produkciót, ami az L2-hatás gyengülését eredményezve a sztenderd anyanyelvi mintázatok megközelítésében nyilvánul meg (5. ábra).



4. ábra. A felvételkészítés főbb lépései, illetve a hozzájuk kapcsolódó nyelvi módokban várt változás a nyelvimód-kontinuumon (lent), és ehhez kapcsolva a dallammenetben (fent, ahol a szaggatott fekete vonal jelöli az L1-hez és L2-höz képest megvalósuló f_0-kontúrt).

A nyelvi tapasztalat szempontjából az L2-hatás gyengülését a haladók esetében várom jelentősnek, hiszen az ő elméjükben jelenik meg nagyobb különbség az L1- és L2-mintázat között, így az ő esetükben várható a legjelentősebb nyelvimódbeli áthangolódás is. A kezdők esetében is az L2-hatás gyengülését várom, azaz azt, hogy az anyanyelvi produkciós feladatban előrehaladva az ejtésük egyre inkább a sztenderd L1-es mintázatokat közelítik meg, azonban náluk a haladókhoz képest visszafogottabb hatást várok, mivel a kezdők elméjében



5. ábra. A célnyelvi (L2-)hatás feltételezett csökkenése a kísérlet előrehaladásával, azaz a növekvő számú L1-megnyilatkozás produkciójával (fent), illetve az f_0-kontúrok változása ennek függvényében (lent).

az L1- és L2-es mintázatok közötti különbség elmarad a haladóknál megjelenő differenciától.

A vizsgált négy magyar dallam esetében az L2-hatás gyengülését a következő akusztikai paraméterekben várom megnyilvánulni. Ezeket a paramétereket az 1.1. alfejezetben leírtak alapján egyrészről a kínai tónusok és a magyar intonációs dallamkontúrok közötti különbségből, másrészről az ott bemutatott dallammintázatok közötti eltérésből következtetem. A magyar egy szótagú kérdés esetében azt várom, hogy a magyar nyelvű produkció kezdetéhez képest a kísérlet előrehaladásával párhuzamosan (azaz az anyanyelvi megnyilatkozások számának növekedésével) mindkét nyelvtanuló csoport f_0-görbéjében a maximális f_0-érték emelkedik, illetve a haladók esetében emellett a minimális f_0 is magasabb értéket vesz fel. Az alternatív kérdés emelkedő fázisa esetében is a minimális f_0-érték emelkedésére és a dallammenet egyre lineárisabb, emelkedőbb mintázatára számítok, tehát a normalizált időtartam záró szakaszában magasabb maximális f_0-értéket várok. A felszólító dallam esetében nem találtam szignifikáns eltérést a csoportok között, tehát az L2-hatás nem torzította a nyelvtanulók produkcióját a sztenderd magyar ejtéshez képest, ezért az L2-hatás gyengülését sem várom. A kijelentések esetében mind a maximális, mind a minimális f_0-érték csökkenésére számítok az idő előrehaladásának függvényében.

A magyar kontrollcsoport esetében az L2 hatása nem releváns változó, azonban e beszélői csoport produkciójában azt várom, hogy az ismétlések számával párhuzamosan – az ismétlési redukcióból, és az ejtés gazdaságosságából fakadóan – a dallamok legrelevánsabb akusztikai tulajdonságai erősödnek fel, amely akusztikai tulajdonságok lehetővé teszik a vizsgált dallamok közötti szembenállások nyilvánvalóvá tételét. Az egy szótagú eldöntendő kérdő dallam emelkedése esetében ez azt jelenti, hogy az idő előrehaladásával a dallamzáró magas (H) akusztikai cél, tehát az f_0 maximum prominens emelkedésére számíthatunk, amit az ismétlési redukcióból fakadóan az f_0-tartomány kompressziója kísér. A kísérlet előrehaladásával párhuzamosan az alternatív kérdés emelkedő fázisának megvalósulásában is ugyanerre a tendenciára számítok, még akkor is, ha e dallam esetében a prozódiai eszközök mellett lexikai elemek is rendelkezésre állnak a kérdő szándék kifejezésére. A felszólító dallam esetében a dallam kezdeti fázisában – az eltelt idővel párhuzamosan – szintén a H-s specifikáció prominensebbé válását, azaz az f_0 maximum emelkedését várom az f_0-tartomány kompressziójával. Valamint a kijelentő dallam esetében a kísérlet előrehaladásával párhuzamosan a normalizált időtartamon belül az alacsony (L) tónusú fázis prominensebb megjelenésére számítok, ami a minimális f_0-értékek csökkenésében valósul meg és amit szintén az f_0-terjedelem összeszűkülése kísér.

2. Módszertan

A kísérletben két eltérő nyelvi tapasztalattal rendelkező magyar anyanyelvű kínaiul tanuló beszélői csoport ejtését hasonlítottam össze egy magyar anyanyelvű beszélői csoport produkciójával. Mindhárom beszélői csoport egyetemista hallgatókból állt, csoportonként 7-7 (összesen 21) kísérleti személyből, akik valamennyien nők. A magyar kontrollcsoport átlagéletkora $26,3 \pm 2,81$ év és nem rendelkeztek semmilyen kínai nyelvi tapasztalattal. A kezdő kínaiul tanulók kínai alapszakos egyetemisták voltak, akik legalább 2 éve tanulnak kínaiul (átlagéletkoruk $22 \pm 1,14$ év volt). A haladó nyelvtanuló csoport tagjai kínai mesterképzéses hallgatók voltak, akik legalább 3 éve tanulnak kínaiul, és a

hét személyből négyen töltöttek egy évet kínai nyelvterületen (átlagéletkoruk $24,2 \pm 6,21$ év). A felvételkészítésre az ELTE BTK Alkalmazott Nyelvészeti és Fonetikai Tanszéken került sor. A hangfelvételeket 16 bit-en, 44,1 kHz-en digitaliztálva rögzítettem egy külső hangkártyával és egy omnidirekcionális kondenzátoros fejmikrofonnal.

Ahogy azt már a hipotézisekben említettem, a kísérlet több különböző nyelvi anyag felvételéből állt össze, azért hogy a lehető leginkább kontrolláljuk a nyelvi módokat. A felvétel legelején a kísérleti személyeknek egy relatíve hosszú, kb. 10-12 perces kínai nyelvű anyagot kellett felolvasniuk, melyben kérdő és kijelentő kínai mondatokat kellett szembeállítaniuk egymással. Ezután következett a kínai monoszillabikus 2. és 4. tónusú, a jelen vizsgálatban elemzett magyar célszavakkal közel megegyező szegmentumokból álló CV szerkezetű, jelentéssel rendelkező szavak felolvasása izolált ejtésben, amely kísérlet eredményeit a Bevezetésben mutattam be (a teljes elemzést lásd: Juhász, megjelenőben). Ezután a felvételeket mindenféle szünet vagy átmenet nélkül a magyar anyag felolvasásával folytattam azért, hogy a kínai felvételek alatt behangolt nyelvi módot a lehetőség szerinti legkevesebb olyan magyar anyanyelvi inger érje, ami nem a kísérlethez kapcsolódik.

A magyar kísérlet anyagát 5 magyar CV-szerkezetű, jelentéssel rendelkező, egy szótagú szó adta, melyeket 5 ismétléssel kellett felolvasni. A magyar anyagban a beszélők négyféle dallamtípust produkáltak: az emelkedő dallammenetek esetében egy szótagú és alternatív kérdést, az ereszkedő dallamok esetében pedig kijelentést és felszólítást. A magyar célszavakat a kísérleti személyeknek rövid párbeszédekbe foglalva, de önálló megnyilatkozásként kellett felolvasniuk, amelyeket kontextusba helyezve hívtam elő a megfelelő intonációs séma produkálásával. A felvétel során megjelenített magyar párbeszédeket bemutató példa az 1. táblázatban látható. A kísérletben felolvasott összes megnyilatkozás a tanulmány végén, a Függelék 6. táblázatában található. A kísérletben a magyar egy szótagú szavak vokalikus részét elemeztem, mely egy középső nyelvállású, elöl képzett, labiális [øː] szegmentum volt, amit minden esetben egy zöngét-

len aspirálatlan obstruens előzött meg. A vizsgált célszavak az 1. táblázatban kiemeléssel láthatók.

1. táblázat. A kísérletben felolvasott emelkedő és ereszkedő dallamú közlések példái.

Emelkedő dallamok	**Ereszkedő dallamok**
Alternatív kérdés (ebben az esetben mindig az első tag a vizsgált elem) és egy szótagú jelöletlen kérdés: – ***Cső*** [tʃøː] *vagy csá?* – *Nem tudom.* ***Cső?*** [tʃøː]	Kijelentés: *Mi az a henger, amiben folyik a víz?* – ***Cső.*** [tʃøː]
	Felszólítás: *A vízvezetékszerelő tíz óra munka után így szólítja fel a csövet:* –***Cső!*** [tʃøː] *Nehogy kilyukadj nekem!*

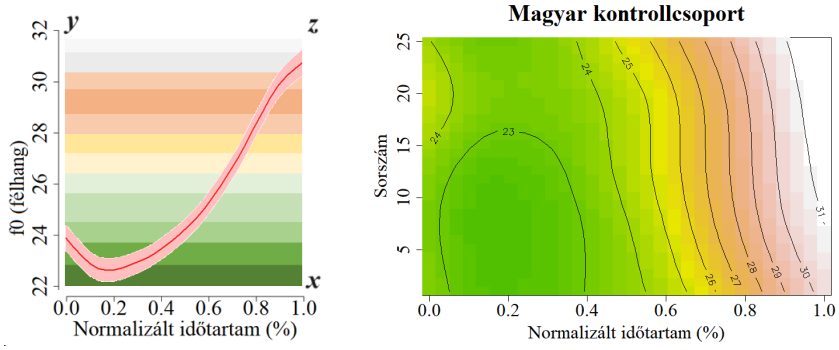
A hangfelvételeket a *Praat* szoftverben (Boersma & Weenink, 2019) címkéztem és elemeztem: minden elemzést a vokalikus szakaszokon végeztem. A vokalikus szakaszokat a kváziperiodikus hullám megjelenésétől annak megszűnéséig szegmentáltam, és ezt az intervallumot elemeztem a szakasz időtartamaként is. A kinyert f_0-értékeket minden esetben félhangokká konvertáltam az *R* programban (R Core Team, 2021) a *hqmisc* (Quené, 2014) csomag segítségével, minden esetben 50 Hz-es referenciaértékkel. Az f_0-görbék elemzéséhez az f_0 értékét 5 ms-onként nyertem ki adott időpillanatban, automatikusan, a szegmentált időtartamon belül.

Az f_0-görbék elemzésére generalizált additív modelleket (GAMM) használtam, külön modellel vizsgálva négy közléstípust. A GAMM nemlineáris adatsorok elemzésére szolgál, és a becsült görbét alapfunkciók (függvények) és az alapfunkciókhoz rendelt különböző súlyok kombinációjával állítja elő (Wood, 2017). Az alapmodellben két független változó interakcióját elemeztem: egyrészről az (5 milliszekundumonként kinyert és félhangokká konvertált) f_0 függő változóra a normalizált időtartam független változó hatását vizsgáltam, ami megadja, hogy az f_0 értéke az időtartamra simítva hogyan változik a normalizált időtartamon belül. Másrészről a normalizált időtartamban megjelenő f_0-változást a magyar

nyelvű kísérlet kezdetétől eltelt idő függvényében is vizsgáltam, amit a felvétel során rögzített megnyilatkozások sorszámával számszerűsítettem. Tehát a GAMM-ban az f_0 változása e két független változó tenzor-interakciójában lett elemezve. Továbbá a modellt a beszélői csoport háromszintű sorrendbe állított faktorváltozójával egészítettem ki (kezdő, haladó, kontroll), valamint mintánkénti random simítással bővítettem. A sorrendbe állított faktor változói közül mindig a magyar kontrollcsoport ejtése volt az első szint, tehát ez szolgált a referenciagörbének, amelyhez képest a modell a differenciagörbéket számolta. A becsült görbék statisztikai elemzését az R-ben (R Core Team, 2021) az *mgcv* (Wood, 2017) csomaggal, míg a görbék ábrázolását az *itsadug* (van Rij et al., 2020) csomag segítségével végeztem. A statisztikai próba eredményei közvetlenül eredménnyel szolgálnak az f_0-görbék minimális és maximális f_0-értékéről is, így a dallammenetek ezen szempontok szerint is vizsgálhatók.

Az adatábrázolást illetően tehát ez azt jelenti, hogy az f_0 változását három dimenzió mentén hasonlítjuk össze, amely összehasonlítás eredményét a 6. ábra jobb panelje mutatja be. Az f_0 függő változó értékeit a z-tengely mentén a színárnyalatok mélysége kódolja. Ez azt jelenti, hogy a 6. ábra bal paneljén látható magyar kontrollcsoport által ejtett egy szótagú kérdés emelkedő mintázatában az f_0 minél magasabb értéket vesz fel, annál melegebb (narancssárgább∼fehérebb), illetve minél alacsonyabb, annál hidegebb (zöldebb) színárnyalatot kapunk (6. ábra, bal). Az x-tengely a 6. ábrán minden esetben a megnyilatkozás vokalikus szakaszának normalizált időtartamát hivatott jelezni, azonban az y-tengely a jobb panel esetében eltér: ebben az esetben az y-tengely a sorszámot jeleníti meg, tehát azt, hogy az adott megnyilatkozás a magyar nyelvű kísérletben hányadikként jelenik meg (a vizsgált mondattípuson belül). Tehát a sorszám emelkedésével (az ábrán felfelé haladva az y-tengely mentén) távolodunk a kínai nyelvi ingerektől, azaz a nyelvtanulók egyre több magyar nyelvű megnyilatkozást produkáltak. Ebben az esetben tehát az adott megvalósulású dallamkontúr az ábra egy horizontális szeletében jelenik meg. Továbbá abban az esetben, ha a sorszám változónak nincs hatása az f_0 alakulására, akkor az f_0-értékeket jelző szintvonalak az y-tengellyel párhuzamos vertikális egyeneseket vesznek fel. Fon-

tos megemlíteni, hogy a színárnyalatok kódolása mindig határértékek függvényében történik, amely határértékeket minden ábra esetében külön-külön az ábra címében mellékelek. Valamint a színárnyalatok és az f_0-értékek közötti összefüggés az eredményekben bemutatott differencia-plotok esetében is alkalmazható, azonban a differencia-plotok esetében minél melegebb (pirosas) a színárnyalat, annál jelentősebb a különbség az összehasonlított dallamok között.



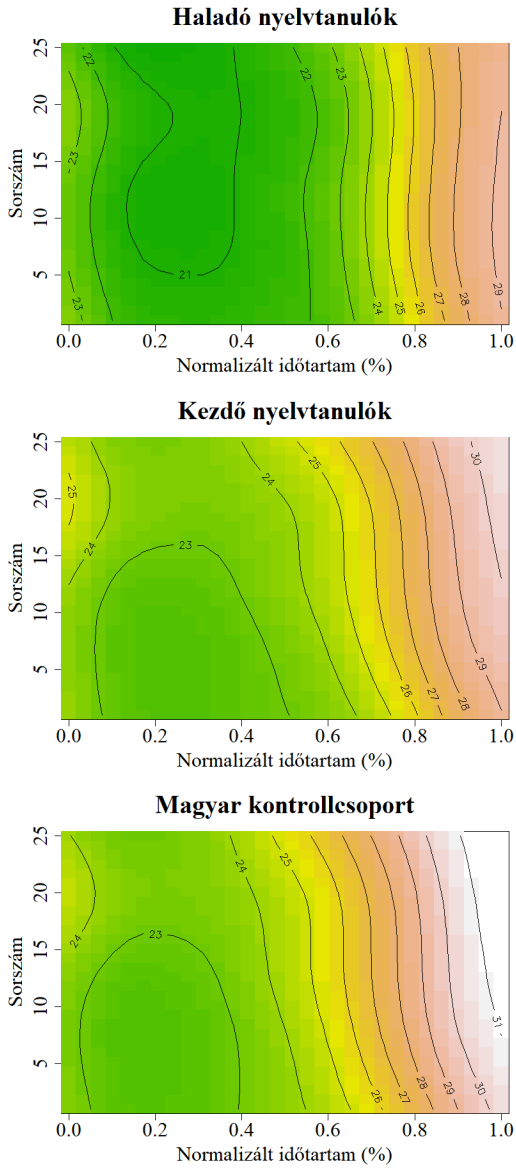
6. ábra. Az f_0 változása a normalizált időtartam függvényében és a színárnyalatok szempontjából (bal), és a normalizált időtartam és a sorszám interakciójában (jobb).

3. Eredmények

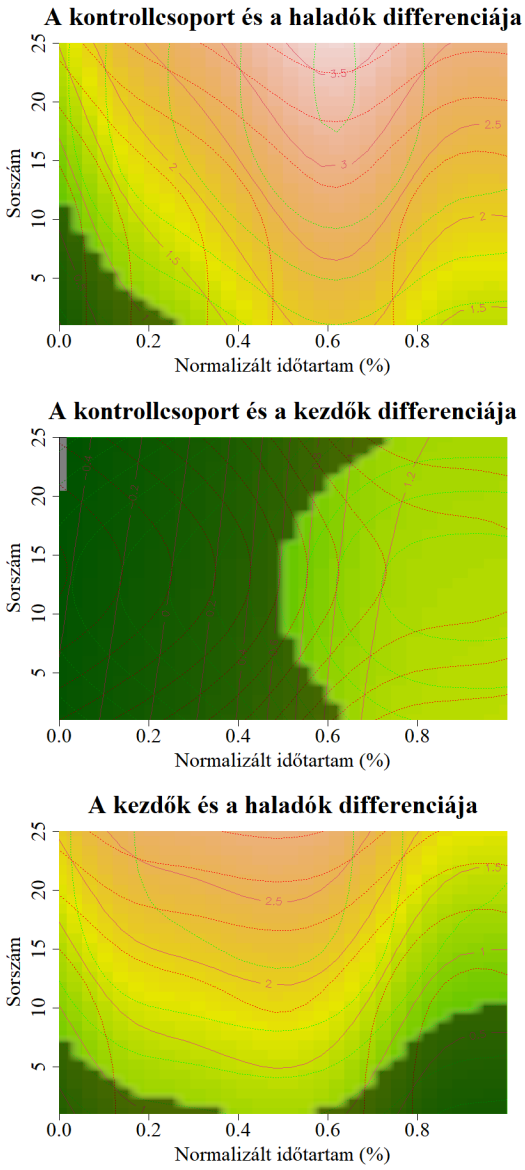
3.1. Az egy szótagú eldöntendő kérdő dallam

A monoszillabikus kérdés esetében az f_0 változására szignifikáns hatást gyakorolt a normalizált időtartam és a megnyilatkozás sorszámának interakciója ($EDF = 14,7$; $F = 13,4$; $p < 0,001$; $R^2 = 84,7\%$). A magyar kontrollcsoport ejtésében az látható, hogy az idő előrehaladtával mind a maximális, mind a minimális f_0-érték emelkedik, ami azt jelenti, hogy míg a kísérlet elején a dallam átlagosan 23 st-ról (félhangról) 30 st magasságra emelkedett a csoport beszélőinek ejtésében, addig az f_0-érték a kísérlet végére már átlagosan 24 st-ról indulva érte el a 31 st-gal jellemezhető maximális f_0-értéket (7. ábra). Tehát az f_0-terjedelem a kísérlet egészében átlagosan 7 félhangnak tekinthető, azonban az idő előrehaladásával a dallam egy félhanggal magasabb f_0-tartományba

emelkedett (2. táblázat). Hasonló mintázatokat mutatnak a kezdő nyelvtanulók is: körülbelül a 15. sorszámú megnyilatkozás után kezd megjelenni az egy félhanggal magasabb minimális f_0 (23 st-ről emelkedik 24 st-re), valamint ugyanígy a 15. sorszám esetében kezd a maximális f_0 is egy félhanggal magasabban megvalósulni (29 st-ról emelkedik 30 st-re). Tehát ebben az esetben a kontrollcsoporthoz hasonlóan az egy szótagú eldöntendő kérdő dallam a kísérlet végére a kezdeti értékekhez képest egy félhanggal magasabb f_0-tartományban realizálódik, azonban a kontrollcsoporthoz képest kisebb f_0-terjedelmű, csak 6 félhang emelkedéssel. A haladó nyelvtanulók ejtésében az eddigiekkel pont ellentétes tendenciákat láthatunk, ami azt jelenti, hogy a kísérlet kezdetéhez képest éppen ereszkedett mind a minimális, mind a maximális f_0-értéke. Ez azt jelenti, hogy míg a kontrollcsoport és a kezdők esetében a dallamív megjelenése magasabb f_0-tartomány felé emelkedett, a haladók ejtésében az f_0-kontúr éppen ellenkezőleg, alacsonyabb f_0-tartomány felé mozdult el. Azonban a haladók f_0-terjedelemében jobban megközelítették a kontrollcsoport produkcióját, hiszen mindkét csoport megközelítőleg 7 st f_0-terjedelemmel ejtette a magyar egy szótagú kérdés dallamát. A differencia-plotokat szemügyre véve (8. ábra) azt láthatjuk, hogy a kontrollcsoport és a haladók f_0-kontúrjai között a kísérlet végére a normalizált időtartam kétharmadánál jelentkezik a legnagyobb – több mint 3 félhangnyi – különbség. Ez az idő előrehaladásával növekvő eltérés a két csoport között abból fakadhat, hogy a kontrollcsoport az idő előrehaladásával egyre meredekebb és a normalizált időtartamban kiterjedtebb emelkedést produkál, ami magas maximális f_0-val rendelkezik, ezzel szemben a haladó nyelvtanulók görbéje esetében éppen az alacsony frekvenciasávban megvalósuló homorú, öblös fázis terjed ki a normalizált időtartam adott szakaszára. A kontrollcsoport és a kezdők produkciója közötti különbség pedig abból fakad, hogy a kezdők a kontrollcsoportnál 1 félhanggal alacsonyabb maximális f_0-értékkel rendelkeztek.



7. ábra. Az egy szótagú kérdés dallamának f_0-változása a normalizált időtartam és a megnyilatkozás sorszáma függvényében, a három beszélői csoport ejtésében (ahol a színárnyalatok határértékei: 20 st és 31 st).



8. ábra. Az egy szótagú kérdés f_0-változásának páros összehasonlítása, azaz differenciái a beszélői csoportok között (ahol a nem szignifikáns különbséget a szürke sáv, míg a szignifikánsan eltérő tartomány minőségét a színárnyalatok jelölik, amelynek határértékei: $-0,5$ st és 4 st, ahol a hidegebb (zöldebb) színárnyalat relatíve kicsi eltérést; míg a melegebb (pirosas) színárnyalat a jelentős (4 st-t megközelítő) eltérést jelöli.

2. táblázat. Az egy szótagú kérdés becsült minimális és maximális átlagos f_0-értéke a produkciós feladat kezdetén és végén (a GAMM-ok ábrái alapján kiemelve az átlagos becsült határértékeket) a három beszélői csoport ejtésében

	Időpont	**f_0-érték**	**f_0-terjedelem**
Haladók	*kezdet (min – max)*	22 st – 29 st	7 st
	vég (min – max)	21 st – 28 st	7 st
Kezdők	*kezdet (min – max)*	23 st – 29 st	6 st
	vég (min – max)	24 st – 30 st	6 st
Kontroll	*kezdet (min – max)*	23 st – 30 st	7 st
	vég (min – max)	24 st – 31 st	7 st

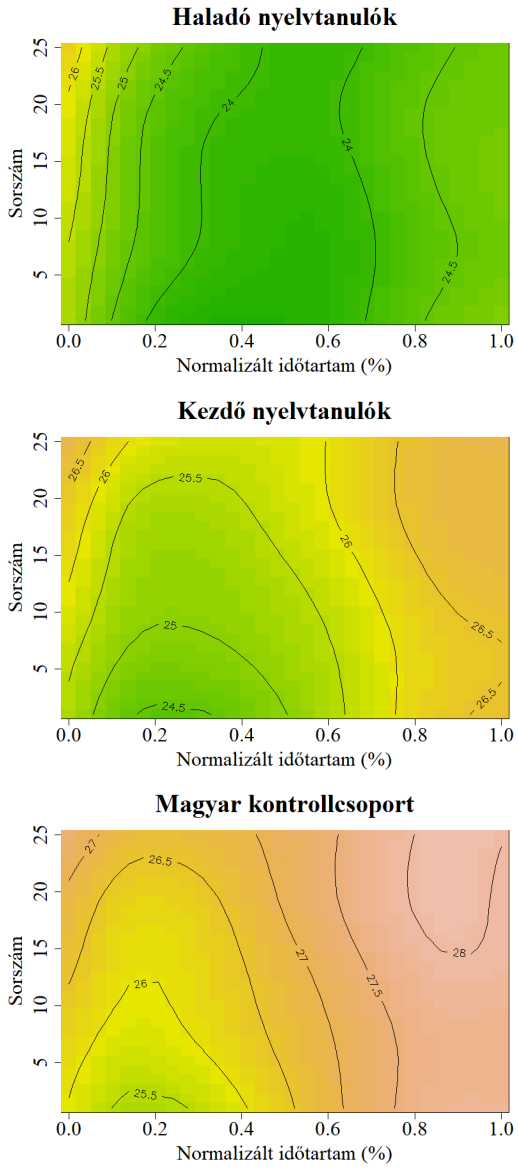
3.2. Az alternatív kérdés emelkedő fázisa

Az alternatív kérdés emelkedő fázisát vizsgálva az f_0 változására szintén szignifikáns interakciós hatást gyakorolt a normalizált időtartam és a megnyilatkozás sorszáma ($EDF = 10,4$; $F = 10,1$; $p < 0,001$; $R^2 = 88,0\%$). A magyar kontrollcsoport esetében itt is megfigyelhető az egy szótagú kérdéshez hasonló mintázat, azaz a dallamkontúr magasabb f_0-tartományba emelése a megnyilatkozások sorszáma mentén, azonban itt ez a folyamat az f_0-terjedelmet is befolyásolja: míg a kísérlet elején a dallamívben 2 st-nyi emelkedés volt megfigyelhető, az f_0-terjedelem a kísérlet végére megközelítőleg 1,5 st-ra csökkent (9. ábra, 3. táblázat). A kezdők esetében nem jelenik meg a maximális f_0-érték emelkedése, pusztán a dallamív a normalizált időtartamban korábban éri el a maximumát, ezért az f_0-görbe terjedelmesebb szakasza realizálódik magas f_0-tartományban. A kezdők minimális f_0-ját középpontba véve azonban emelkedés figyelhető meg az idő előrehaladtával, ami azt eredményezi, hogy a dallam f_0-terjedelme a kezdeti 2 st-nyi emelkedéshez képest megközelítőleg 1,5 st-re csökken. A haladók ejtésében az figyelhető meg, hogy az idő előrehaladásával a görbe egyre kisebb hányada realizálódik alacsony f_0-tartományban, tehát a formája egyre „V-alakúbbá” válik, azonban emelkedő mintázatot csak nagyon visszafogottan, mindössze fél félhangnyi terjedelemben mutat. A haladók produkciójában az

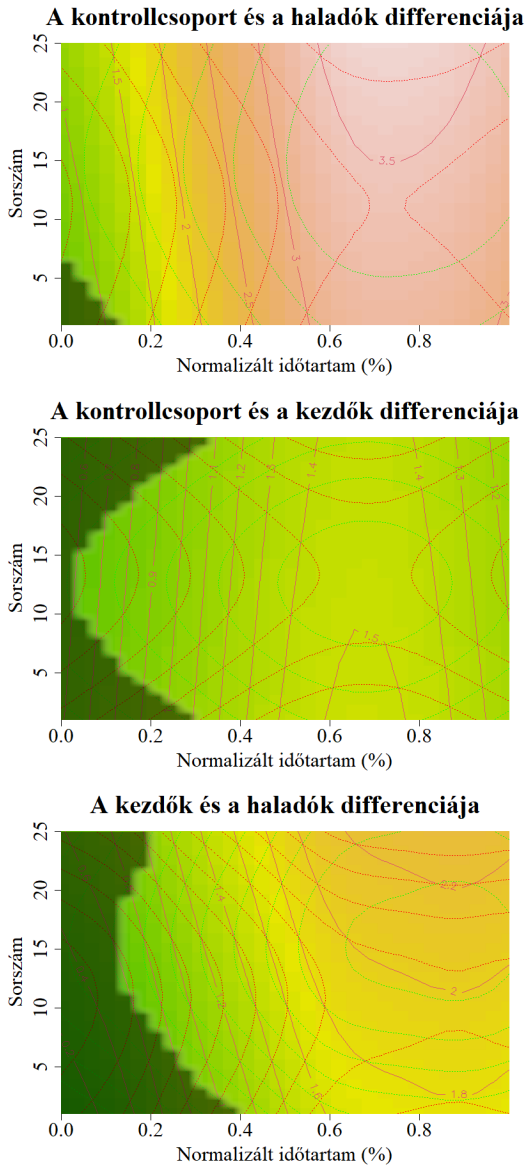
alternatív kérdés emelkedő mintázata helyett a normalizált időtartam elején inkább egy meredekebb ereszkedés figyelhető meg a kísérlet végére. Ezért az emelkedő fázis hiányából eredeztethető a jelentős differencia a kontrollcsoport és a haladók ejtése között is (10. ábra). A kezdők esetében a kontrollcsoporttól való szignifikáns eltérés pedig az f_0-görbe jelentősen alacsonyabb f_0-tartományú realizációjából fakad.

3. táblázat. Az alternatív kérdés becsült minimális és maximális f_0-átlagértéke a produkciós feladat kezdetén és végén (a GAMM-ok ábrái alapján kiemelve a becsült határértékeket) a beszélői csoportok ejtésében.

	Időpont	**f_0-érték**	**f_0-terjedelem**
Haladók	*kezdet (min – max)*	24 st – 24,5 st	0,5 st
	vég (min – max)	24 st – 24,5 st	0,5 st
Kezdők	*kezdet (min – max)*	24,5 st – 26,5 st	2 st
	vég (min – max)	25,5 st – 26,5 st	1 st
Kontroll	*kezdet (min – max)*	25,5 st – 27,5 st	2 st
	vég (min – max)	26,5 st – 28 st	1,5 st



9. ábra. Az alternatív kérdés dallamának f_0-változása a normalizált időtartam és a megnyilatkozás sorszáma függvényében, a három beszélői csoport ejtésében (ahol a színárnyalatok határértékei: 23 st és 29 st).



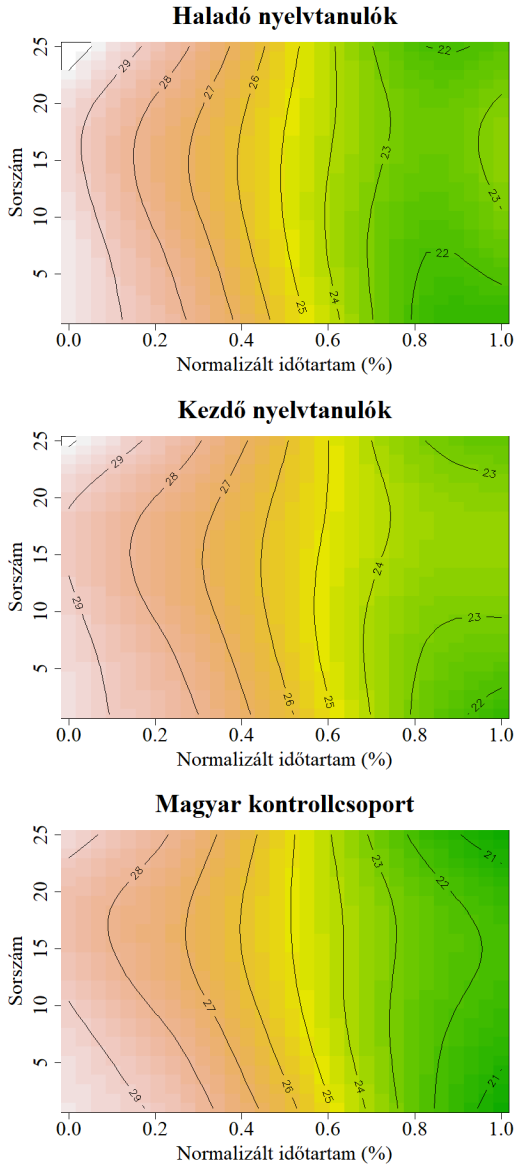
10. ábra. Az alternatív kérdés f_0-változásának páros összehasonlítása, azaz differenciái a beszélői csoportok között (ahol a nem szignifikáns különbséget a szürke sáv, míg a szignifikánsan eltérő tartomány minőségét a színárnyalatok jelölik, amelynek határértékei: 0,5 st és 4 st, tehát a hidegebb (zöldebb) színárnyalat relatíve kicsi (0,5 st közeli) eltérést; míg a melegebb (pirosas) színárnyalat a jelentős (4 st-t megközelítő) eltérést jelöli.

3.3. A felszólító dallam

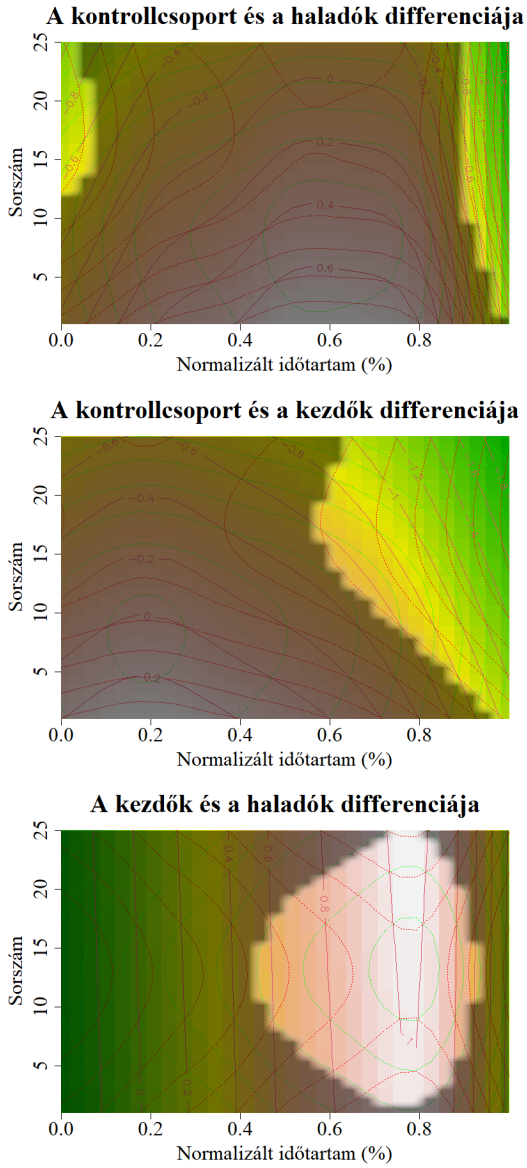
A felszólító dallam esetében is a GAMM szignifikáns interakciós hatást mutatott a normalizált időtartam és a megnyilatkozások sorszáma független változók között ($EDF = 11,9$; $F = 41,3$; $p < 0,001$; $R^2 = 81,7\%$). Azonban a felszólító dallam esetében a három vizsgált csoport nem különbözött jelentősen, a megközelítőleg 7-8 félhangnyi ereszkedés mind a kontrollcsoportra, mind a nyelvtanuló csoportokra jellemző volt (11. ábra, 4. táblázat). A rögzített megnyilatkozások sorszámának függvényében is ugyanazt a mintázatot tudjuk megfigyelni mindhárom csoport esetében: körülbelül a 15. sorszámú megnyilatkozás esetében a maximum f_0 a normalizált időtartam korábbi pontján kezd ereszkedni, valamint a normalizált időtartam végén a görbe záró fázisa kiterjedtebb fázist mutat ugyanabban az alacsonyabb frekvenciatartományban. A haladók szempontjából a kontrollcsoporttól való szignifikáns eltérés elsősorban a görbe záró fázisában jelenik meg, ami a haladók szignifikánsan magasabb f_0-minimumából következik (12. ábra). Hasonlóképpen a kezdők esetében a megnyilatkozások sorszámnövekedésével párhuzamosan egyre jelentősebbnek mutatkozik a szignifikánsan differens tartomány a kontrollcsoporthoz képest, szintén a kontrollcsoportnál alacsonyabb minimális f_0-ból következően.

4. táblázat. A felszólítás becsült minimális és maximális f_0-átlagértéke a produkciós feladat kezdetén és végén (a GAMM-ok ábrái alapján kiemelve a becsült határértékeket) a beszélői csoportok ejtésében.

	Időpont	**f_0-érték**	**f_0-terjedelem**
Haladók	*kezdet (max – min)*	29 st – 22 st	7 st
	vég (max – min)	30 st – 22 st	8 st
Kezdők	*kezdet (max – min)*	29 st – 22 st	7 st
	vég (max – min)	30 st – 23 st	7 st
Kontroll	*kezdet (max – min)*	29 st – 20 st	8 st
	vég (max – min)	29 st – 21 st	8 st



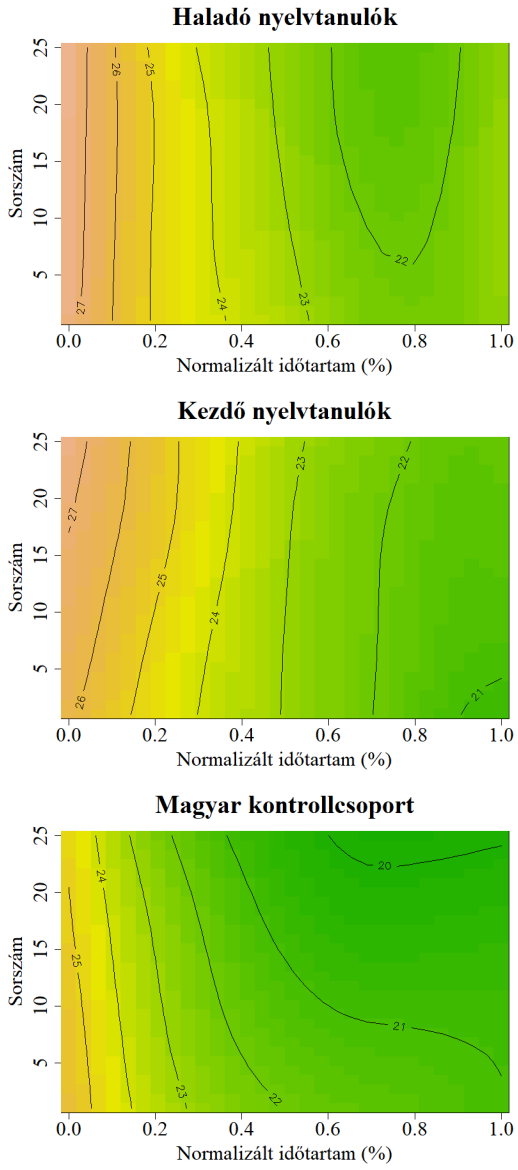
11. ábra. A felszólítás dallamának f_0-változása a normalizált időtartam és a megnyilatkozás sorszáma függvényében, a három beszélői csoport ejtésében (ahol a színárnyalatok határértékei: 20 st és 30 st).



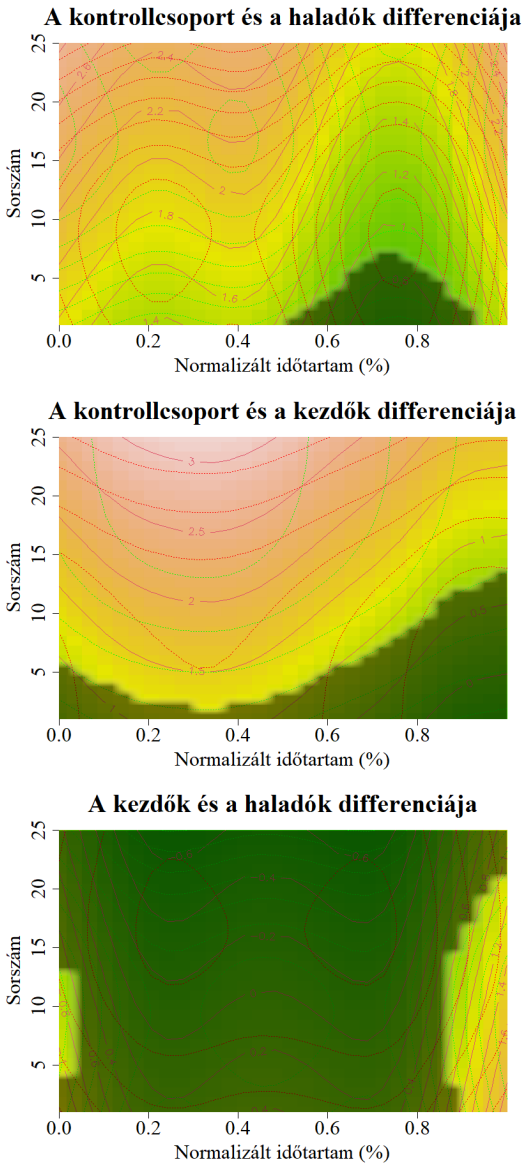
12. ábra. A felszólítás f_0-változásának páros összehasonlítása, azaz differenciái a beszélői csoportok között (ahol a nem szignifikáns különbséget a szürke sáv, míg a szignifikánsan eltérő tartomány minőségét a színárnyalatok jelölik, amelynek határértékei: -2 st és 1 st).

3.4. A kijelentő dallam

A kijelentések tekintetében a két független változó, azaz a normalizált időtartam és a sorszám szignifikáns hatást gyakorolt az f_0 változására ($EDF = 7,6$; $F = 7,1$; $p < 0,001$; $R^2 = 82,1\%$). A kontrollcsoport ejtésében az f_0-változást tekintve azt láthatjuk, hogy a megnyilatkozások sorszámával párhuzamosan a normalizált időtartam egyre jelentősebb hányada realizálódik alacsony f_0-tartományban és viszonylag konstans értékekkel, továbbá a kijelentő f_0-görbe mind minimális és maximális f_0-értéke csökkenő mintázatot mutat (13. ábra). Ezzel szemben a nyelvtanuló csoportok esetében az látható, hogy a görbe ereszkedése a kontrollcsoportnál magasabb f_0-tartományból indul és a dallam minimális f_0 tekintetében is a kontrollhoz képest magasabb értéket vesz fel. Mind a kezdők, mind a haladók egy fokozatosan ereszkedő görbét produkálnak, azaz a nyelvtanulók ejtésében nem mutatkozik a kontrollcsoporthoz hasonló stagnálás a görbe záró szakaszában, illetve a haladók esetében az f_0-kontúr a normalizált időtartam végére egy enyhe emelkedéssel is kiegészül. Fontos kiemelni, hogy a kontrollcsoport esetében az idő előrehaladtával az f_0-kontúr egyre alacsonyabb f_0-tartományba ereszkedve jelenik meg, továbbá a dallamív f_0-tartománya is az alacsonyabb frekvenciasávok felé tágul (3 st-ről 4 st-re), azonban ezek a jellemzők egyik nyelvtanuló csoportra sem jellemzőek (5. táblázat). A nyelvtanuló csoportok f_0-terjedelme a kontrollcsoportokhoz képest jelentősebb, 4-5 st-ra tehető, ami a haladók esetében a sorszám növekedése mentén egy félhanggal kiterjedtebbé válik úgy, hogy habár a maximális f_0-érték az idő előrehaladtával nem változik, az f_0 minimumértéke azonban csökkenést mutat. Ezzel ellentétben a kezdők produkciójában a kijelentő dallam görbéje megőrzi az 5 st-nyi terjedelmét, azonban az eddig bemutatottakkal ellentétesen a dallamív magasabb f_0-tartományba emelkedik. A kontrollcsoporthoz képest számított differencia esetében – meglepő módon – azt láthatjuk, hogy a kísérlet előrehaladásával a kontrollcsoport és a nyelvtanuló csoportok közötti dallamívbeli különbség egyre inkább növekszik (14. ábra). Ez a jelenség egyrészről a nyelvtanulók magasabb f_0-tartományban megjelenő kijelentő görbéjéből fakad, másrészről pedig a kontrollcsoporthoz képest eltérő irányú f_0-változásokból.



13. ábra. Az alternatív kérdés dallamának f_0-változása a normalizált időtartam és a megnyilatkozás sorszáma függvényében, a három beszélői csoport ejtésében (ahol a színárnyalatok határértékei: 19 st és 30 st).



14. ábra. Az alternatív kérdés f_0-változásának páros összehasonlítása, azaz differenciái a beszélői csoportok között (ahol a nem szignifikáns különbséget a szürke sáv, míg a szignifikánsan eltérő tartomány minőségét a színárnyalatok jelölik, amelynek határértékei: −1 st és 3,5 st).

5. táblázat. A kijelentés becsült minimális és maximális f_0-átlagértéke a produkciós feladat kezdetén és végén (a GAMM-ok ábrái alapján kiemelve a becsült határértékeket) a három beszélői csoport ejtésében.

	Időpont	**f_0-érték**	**f_0-terjedelem**
Haladók	*kezdet (max – min)*	27 st – 23 st	4 st
	vég (max – min)	27 st – 22 st	5 st
Kezdők	*kezdet (max – min)*	26 st – 21 st	5 st
	vég (max – min)	27 st – 22 st	5 st
Kontroll	*kezdet (max – min)*	25 st – 22 st	3 st
	vég (max – min)	24 st – 20 st	4 st

4. Következtetések

A jelen akusztikai elemzés középpontjában a nyelvi módok interakciója, pontosabban azon kérdés vizsgálata állt, hogy a kínai tónusprodukció anyanyelvre gyakorolt hatása változik-e annak függvényében, hogy a kísérleti személyek a felvétel rögzítése során mennyi anyanyelvi ingert produkálnak. Tehát azt a kérdést vizsgáltam, hogy a nyelvtanulók a növekvő számú anyanyelvi megnyilatkozás hatására graduálisan a sztenderd anyanyelvi ejtést közelítik-e meg. Emellett a kísérlet további eredményeit illetően arról is fogalmazhatunk meg következtetéseket, hogy a magyar kontrollcsoport ejtésében az ismétlések révén a felolvasott anyag megismerése hogyan befolyásolja a magyar intonációs kontúrok megvalósulását, azaz hogy az egyes dallammenetek esetében a beszélők milyen akusztikai tulajdonságokra erősítenek rá az ismétlések mentén. Ezért minden mondattípus esetében első körben a magyar kontrollcsoport ejtéséről teszek megállapításokat, és ezután térek rá a nyelvtanulók produkciójára.

4.1. Az egy szótagú eldöntendő kérdő dallam produkciója

Először az egy szótagú kérdéseket véve középpontba, a magyar kontrollcsoport ejtésében a kísérlet előrehaladása mentén az f_0-terjedelem nem változott jelentősen, azonban maga az f_0-kontúr az idő előrehaladtával egy félhanggal

magasabb tartományba emelkedett. Tehát a várakozásaimmal részleges ellentétben az egy szótagú eldöntendő kérdő dallam esetében a magyar anyanyelvűek a kísérlet előrehaladásával párhuzamosan nem csak magasabb f_0 maximumértékkel (azaz a H cél emelésével és prominensebbé tételével) produkálták ezt a dallamot, hanem az egész dallamkontúrt magasabb f_0-tartományba emelték, azaz az f_0-terjedelem kompressziója nem valósult meg. E stratégia feltételezhetően elsősorban pragmatikai okokra vezethető vissza, amelyek megkövetelik az állandó f_0-terjedelem megőrzését, kódolva a semleges eldöntendő kérdő dallam jelentését. Tehát ezekből az eredményekből arra következtethetünk, hogy az ismétlések hatására az egy szótagú eldöntendő dallam esetében mind a magas f_0-tartományban megvalósuló f_0 maximum, mind a kérdő dallam f_0-tartománya is szükségszerű a kérdő szándék kifejezésére.

Ehhez képest a nyelvtanulók ejtésében a magyar egy szótagú kérdés esetében azt vártam, hogy a kísérlet kezdetéhez képest az idő előrehaladtával mindkét nyelvtanuló csoport esetében a maximális f_0-érték emelkedik, illetve a haladók esetében emellett a minimális f_0 is magasabb értéket vesz fel. A haladó nyelvtanulók esetében e hipotézis egyik szempont szerint sem nyert megerősítést, mert mind a maximális, mind a minimális f_0 az idő előrehaladásával csökkenő mintázatot mutatott. Az f_0-terjedelemmel kapcsolatban a haladók a kontrollcsoporttal megegyező f_0-terjedelemmel produkálták az egy szótagú eldöntendő kérdő dallamot végig a produkciós kísérlet alatt, azonban míg az f_0-kontúr a kontrollcsoport esetében az időben előrehaladva magasabb f_0-tartományba emelkedett, addig a haladók esetében éppen alacsonyabb f_0-tartományba süllyedt. Tehát a hipotézisben megfogalmazottak nemcsak nem nyertek megerősítést, hanem az adatok éppen egy ellentétes mintázatot látszanak megerősíteni. Ez azt jelenti, hogy a haladók ejtésében az idő előrehaladásával nemcsak hogy nem közelítették meg a sztenderd magyar ejtést, hanem éppen a kínai tónusok irányába tolódott az ejtésük. A kínai 2. tónus megközelítése abban nyilvánult meg, hogy a kísérlet második felében a görbe jelentősebb hányadát töltötte alacsonyabb frekvenciasávban, valamint a kezdeti mintázathoz képest visszafogottabb emelkedést is mutatott. Ezen mintázatok alapján tehát kérdésként merül fel, hogy vajon a

nyelvváltásból (L2-ből hirtelen L1-re váltásból) fakadó interferencia valóban a hipotézisekben leírtak szerint – graduálisan – megy-e végbe. Nem lehetséges-e vajon az, hogy az anyanyelvre való visszaváltás hirtelen és jelentősen az L1 egynyelvű mód felé lendíti ki a nyelvtanuló elméjét (mint amikor a mérlegre hirtelen jelentős súlyt helyezve a mérleg nyelve eleinte túllő a tárgy súlyán), majd ezután a nyelvtanulók elméje az L2 nyelvi mód aktiváltságából fakadóan visszatér az L2-dominálta kétnyelvű módba (ahogy a mérleg nyelve is a pozitív kilengés után alulmúlja célját)? Továbbá az is elképzelhető vajon, hogy kizárólag ezen kilengést követően várható csak a hipotézisekben leírt folyamat, miszerint a nyelvtanuló elméje fokozatosan közelíti meg a környezeti ingerek kiváltotta nyelvi módot, és az ejtésre jellemző akusztikai tulajdonságokat? Az ezen felvetések mentén leírt folyamat magyarázattal szolgálna a bemutatott eredményekre a haladók ejtését illetően. A kezdők egy szótagú eldöntendő kérdő dallamát tekintve azt láthattuk, hogy a kísérlet előrehaladása a kontrollcsoporttal megegyező mintázatokat eredményezett a kezdő nyelvtanulók produkciójában. Az egyedüli különbség a kezdők és a magyar kontrollcsoport között pusztán a maximális f_0-értékben jelentkezett, ami a kezdők esetében a sztenderd magyar ejtésnél alacsonyabb értékkel valósult meg. Ebben az esetben tekinthetnénk úgy, hogy a hipotézis kezdőkre vonatkozó része megerősítést nyert, azonban nem lehetünk benne biztosak, hogy a maximális f_0-érték emelkedése valóban a célnyelvi hatás gyengüléséből fakad, és nem pedig az anyanyelvi nyelvi mód hatásából következik, aminek a révén a sztenderd magyar ejtésre jellemző séma érvényesült.

4.2. Az alternatív kérdés emelkedő fázisának produkciója

Az alternatív kérdés esetében a magyar kontrollcsoport az egy szótagú eldöntendő kérdő dallamhoz hasonló mintázatot mutatott a kísérlet előrehaladása mentén: az f_0-kontúr egyre magasabb f_0-tartományban valósult meg, ami mind a görbe maximális, mind a minimális f_0-jának emelkedését jelentette, emellett azonban az f_0-terjedelem egy árnyalatnyit összeszűkült. Ezek az eredmények megerősítették hipotézisemet, miszerint az ismétlések hatására a maximális f_0-érték magasabb f_0-tartományba emelkedik, és az f_0-terjedelem is összeszűkül

(abból következően, hogy a minimális f_0-érték jelentősebbet emelkedik a maximális f_0-értékhez képest). Ezekből az eredményekből kiindulva azt mondhatjuk, hogy az egy szótagú eldöntendő kérdéshez viszonyítva az ismétlési redukció hatása jelentősebb befolyással bír az alternatív kérdés realizációjára és az f_0-terjedelem kompressziójára. E hatás feltételezhetően abból eredeztethető, hogy az alternatív kérdés esetében a lexikai eszközök alapvetően meghatározzák a kérdő szándékot, ezért a dallam akusztikai megvalósulását érintő jelentősebb redukció sem akadályozza a megértést.

A nyelvtanulók ejtését illetően az alternatív kérdés emelkedő fázisában a minimális f_0-érték emelkedésére és a dallammenet egyre lineárisabb, emelkedőbb mintázatára számítottam, tehát magasabb maximális f_0-értéket vártam a normalizált időtartam záró szakaszában. A haladók ejtésében e hipotézis egyik szempontból sem nyert megerősítést, hiszen az emelkedő fázis teljesen elmaradt, a minimum f_0-érték sem emelkedett, egyedül a görbe alakja mutatott változást: a kezdeti lapos, alacsony frekvenciatartományra koncentrált görbéhez képest a kísérlet végére egy erős ereszkedéssel induló, inkább homorú V-alakú mintázatot kaptunk, amely habár nem mutatott emelkedést, mégis jelentősebb hányada valósult meg magasabb frekvenciatartományban. Ezen eredmények alapján tehát nem állíthatjuk, hogy a célnyelvi hatás gyengült volna: a haladók nem közelítették meg a kísérlet végére a sztenderd magyar ejtést. Tulajdonképpen azt is feltételezhetnénk, hogy éppen a kínai 2. emelkedő tónus homorú dallamíve az, ami a kísérlet végére a V-alakú homorúbb mintázatot előidézi, azonban ebben az esetben vissza kell térnünk az egy szótagú kérdésnél bemutatott kérdéskörhöz, azaz hogy a nyelvi módokban való áthangolódás valóban graduális-e. A kezdő nyelvtanulók esetében a hipotézis magasabb minimális és maximális f_0-értékére vonatkozó rész megerősítést nyert, azonban ebben az esetben is fennáll a kérdés, hogy ezen mintázatok valóban a célnyelvi hatás gyengüléséből eredeztethetőek, vagy csak a magyar kontrollcsoporthoz hasonlóan az ismétlések és az anyag ismerete hívja elő a magasabb minimális és maximális f_0-értékeket.

4.3. Az felszólító dallam produkciója

A felszólító dallam esetében a dallam kezdeti fázisában – az eltelt idővel párhuzamosan – szintén a H-s specifikáció prominensebbé válását, azaz az f_0-maximum emelkedését vártam az f_0-tartomány kompressziójával. E hipotézisem részlegesen nyert csak megerősítést, mert egyrészről, habár a felszólító dallam f_0-terjedelme az eltelt idővel párhuzamosan csökkent, a kísérlet végére újra visszatért a kezdeti f_0-terjedelemhez. Ehhez hasonló változást mutatott a felszólító dallam maximális és minimális f_0-értéke is, amelyek esetében a maximális f_0 a kísérlet előrehaladásával párhuzamosan csökkent, a minimális f_0-értéke pedig növekedett, de a kísérlet legvégére visszatértek a kezdeti értékükhöz. Tehát összegezve, habár az ismétlési redukció megfigyelhető a felszólító dallam esetében is, ami az f_0-terjedelem kompresszióját eredményezte, azonban az akusztikai szerkezet kezdeti tulajdonságai mégis visszaálltak a kísérlet végére. Ha a kísérlet végét befolyásoló hatástól eltekintünk, akkor érdemes megjegyezni, hogy az egy szótagú eldöntendő kérdő dallamtól eltérően ebben az esetben rugalmasabban változhat az f_0-terjedelem. Ez a flexibilitás feltételezhetően azért is engedhető meg, mert a felszólító dallam 8-9 félhangnyi ereszkedése jelentősen elkülönül a kijelentés 3-4 félhangnyi ereszkedő dallamától, valamint a felszólító dallamot a kísérletben egészében egy magas f_0-tartományból meredeken ereszkedő görbe jellemezte, szemben a kijelentés alacsony f_0-értékek dominálta megvalósulásához.

A nyelvtanulók ejtését véve középpontba, e dallamtípus esetében nem vártam célnyelvi hatást a nyelvtanulók ejtésében, és e hipotézisem megerősítést is nyert. Az egyetlen fennmaradó kérdés a felszólító dallam minimális f_0-értékére vonatkozik, ami a nyelvtanulók esetében, főleg a kezdő nyelvtanulóknál a kontrollcsoporthoz képest magasabb f_0-lal realizálódott. Ennek a különbségnek az is lehet feltételezhetően az oka, hogy a kezdők – jobban kimerülve a kínai produkciós feladatoktól – kisebb f_0-terjedelmeket produkáltak. De itt sem zárható ki a fentebbiekben bemutatott „késleltetett" célnyelvi hatás megjelenése, hiszen a kínai ereszkedő 4. tónus dallamíve – tendencia szintjén legalábbis – mind minimum, mind maximum értékében magasabb f_0-értékekkel valósult meg a fel-

szólító dallamhoz képest (l. Juhász, 2023). Így tehát ezen hatás felelős lehet a görbe minimumának megemeléséért, továbbá eredményezheti a kísérlet végén az utolsó néhány megnyilatkozás esetében kialakuló rendkívül magas maximális f_0-értékeket is. Ugyanezen mintázat a haladóknál is megjelenik, és az ő esetükben is a kezdőkhöz hasonló kiváltó okokra következtethetünk.

4.4. A kijelentő dallam produkciója

A kijelentő dallam esetében a magyar kontrollcsoport ejtésében a megnyilatkozások sorszámának emelkedése mentén – a várakozásaimmal megegyezően – egyre alacsonyabb maximális f_0-t és egyre kiterjedtebb alacsony frekvenciájú, viszonylag stagnáló fázist láthattunk. Tehát nem meglepő módon a kérdések magas maximális f_0-jával és a felszólítás jelentős f_0-tartományú ereszkedésével ellentétben a kijelentő dallamot leginkább egy visszafogott ereszkedésű, relatíve alacsony frekvenciasávban stagnáló mintázat jellemezte. Emellett a feltételezett f_0-terjedelem kompressziója helyett inkább egy félhangnyi növekedés mutatkozott az f_0-terjedelemben, tehát az ismétlések hatása nem összenyomta, hanem inkább kitágította az alacsonyabb frekvenciaértékek felé a görbe megvalósulását.

A nyelvtanulók esetében tapasztalt célnyelvi hatást a kijelentések esetében mind a maximális, mind a minimális f_0-érték csökkenésében vártam az idő előrehaladása mentén. A haladók ejtése a minimális f_0-ban csökkenést mutatott, ezzel kitágítva a dallam f_0-terjedelmét, tehát a hipotézis ezen része megerősítést nyert, azonban a maximális f_0-értékre vonatkozó hipotézis nem bizonyult helytállónak, mert az nem változott. Ebben az esetben is ugyanúgy fennáll a kérdés, hogy a kontrollcsoport ejtéséhez képest az idő előrehaladásában növekvő differencia nem a célnyelvi hatás késleltetett megjelenéséből fakad-e, hiszen a görbe megvalósulása a haladók ejtésében sokkal lineárisabb és magasabb f_0-tartományból ereszkedik és egyre jelentősebb f_0-terjedelmet vesz fel a kísérlet második felében. A kezdők esetében is hasonló mintázat figyelhető meg, annyi kiegészítéssel, hogy ebben az esetben – a kontrollcsoporttal teljesen ellentétes realizációként – mind a maximális, mind a minimális f_0-érték a vártakkal szem-

ben nem csökken, hanem éppen emelkedik, ezzel is jobban megközelítve a kínai 4. tónus magas frekvenciasávból ereszkedő mintázatát.

Összegezve a kísérlet eredményeit, a magyar kontrollcsoport produkciójában az ismétlési redukció hatása részleges megerősítést nyert, azonban nem minden vizsgált magyar dallam esetében mutatott azonos befolyást. A nyelvtanulók esetében pedig az anyanyelvi intonációs kontúrokra gyakorolt célnyelvi hatást vizsgáló hipotézis alapvetően nem nyert megerősítést, a nyelvtanulók egyik csoportja sem közelítette meg az idő előrehaladásával, azaz a növekvő mennyiségű anyanyelvi produkció révén a sztenderd magyar ejtésre jellemző mintázatokat. Azonban mindenképpen további vizsgálatok szükségesek a felvétel módszertanával kapcsolatban, ahol legelőször a felvétel időtartama merül fel kérdésként. Ebben a kísérletben összesen 100 megnyilatkozást vizsgáltam, azonban nem lehetünk benne biztosak, hogy ezen gyakran ismétlődő, rendkívül egyszerű párbeszédek elég hatékonyan és gyorsan hívják elő az L1 nyelvi módot. Tehát a jövőben érdemes megvizsgálni, hogy vajon, ha a kísérlet során több megnyilatkozást elemzünk hosszabb időtartamban, akkor megkapjuk-e a hipotézisekben leírt natív L1-ejtést megközelítő mintázatokat. Továbbá a kísérlet eredményei azt a kérdést is felvetik, hogy a hipotézisekben feltételezett graduális visszahangolódás az anyanyelvre valóban fokozatos-e, vagy ehelyett inkább a nyelvváltás hatására a nyelvtanuló elméje (és ezáltal a produkció) hirtelen az anyanyelvi nyelvi mód irányába lendül ki, és csak mindezek után késleltetve jelenik meg a fokozatos visszahangolódás az L2-dominálta kétnyelvű módból az L1 egynyelvű mód felé. Emellett nem utolsó sorban azt is meg kell jegyeznünk, hogy a célnyelvi és az anyanyelvi hatás interakciójának megfigyelését az is nehezíti, hogy a nyelvtanuló csoportok ejtése valószínűsíthetően nemcsak a nyelvi mód, hanem az ismétlések hatására is változik, tehát e kérdés a jövőben több vizsgálatot igényel. Összegezve a tanulmány eredményei alapul szolgálhatnak egy olyan jövőbeni kísérletnek, amely több megnyilatkozás produkcióját vizsgálva elemzi a célnyelvi hatást az anyanyelvi szupraszegmentális mintázatok realizációján, továbbá vizsgálja az anyanyelv és a célnyelv egymásra hatását.

Köszönetnyilvánítás

Köszönettel tartozom Gráczi Tekla Etelkának a statisztikai elemzés elméleti megalapozásáért. A kutatás a Kulturális és Innovációs Minisztérium EKÖP-24-es, valamint az NKFIH FK128814-es pályázat támogatásával készült.

Hivatkozások

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [computer program]. 6.1.15-ös verzió.

Chao, Y. R. (1948/1963). *Mandarin Primer*. Cambridge: Harvard University Press.

Cook, V. J. (2003). Introduction: the changing L1 in the L2 user's mind. In V. J. Cook (Ed.), *Effects of the Second Language on the First* (pp. 1–18). Clevedon: Multilingual Matters.

Cook, V. J. (2006). Interlanguage, multi-competence and the problem of the 'second' language. *Rivista di psicolinguistica applicata*, *6*, 39–52.

Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *The Journal of the Acoustical Society of America*, *62*, 971–977.

Flege, J. (2022). A distributional learning account of L2 speech learning. Előadás, 10th International Symposium on the Acquisition of Second Language Speech, 2022. április 22., Barcelona, Spanyolország.

Fónagy, I., & Magdics, K. (1967). *A magyar beszéd dallama*. Budapest: Akadémiai Kiadó.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*, 789–806.

Green, D. W. (1986). Control, activation and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language*, *27*, 210–223.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*, 131–149.

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (Ed.), *Explaining linguistics. One mind, two languages: Bilingual language processing* (pp. 1–22). Oxford: Blackwell.

Grosjean, F. (2008). *Studying bilinguals*. Oxford: Oxford University Press.

Gósy, M. (2004). *Fonetika, a beszéd tudománya*. Budapest: Osiris Kiadó.

Hammarberg, B. (2014). Problems in defining the concepts of L1, L2 and L3. In A. Otwinowska, & D. A. G. (Eds.), *Teaching and learning in multilingual contexts* (pp. 3–18). Toronto: Multilingual Matters.

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

Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? evidence from inner speech and homophone production. *Journal of Memory and Language*, *84*, 37–48.

Juhász, K. (2023). Atonális és tonális nyelvek dallammeneteinek összehasonlítása. *Alkalmazott Nyelvtudomány (Különszám)*, *2*, 21–46.

Juhász, K. (megj.). Az anyanyelv és célnyelv egymásra hatása a beszéd prozódiai megvalósításában.

Leather, J., & James, A. (1991). The acquisition of second language speech. *Studies in Second Language Acquisition*, *13*, 305–341.

de Leeuw, E., Mennen, I., & Scobbie, J. M. (2011). Singing a different tune in your native language: first language attrition of prosody. *International Journal of Bilingualism*, *16*, 101–116.

Major, R. (2001). *Foreign Accent*. New York: Routhledge.

Olaszy, G. (2002). A magyar kérdés dallamformáinak és intenzitásszerkezetének fonetikai vizsgálata. *Beszédkutatás*, *2002*, 83–99.

Pavlenko, A. (2000). L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics*, *11*, 1050–4273.

Quené, H. (2014). hqmisc: Miscellaneous convenience functions and dataset. 0.1-1-es r csomag-verzió.

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna: Foundation for Statistical Computing. URL: `http://www.R-project.org`.

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2020). itsadug: Interpreting time series and autocorrelated data using gamms. 2.4-es r csomag-verzió.

Schwartz, G., Balas, A., & Rojczyk, A. (2015). Phonological factors affecting L1 phonetic realization of proficient polish users of english. *Research in Language*, *13*, 181–198.

Shen, X. S. (1990). *The prosody of Mandarin Chinese*. California: California University Press.

Varga, L. (1994). A hanglejtés. In F. Kiefer (Ed.), *Strukturális Magyar Nyelvtan, 2, Fonológia* (pp. 468–549). Budapest: Akadémiai Kiadó.

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

Wu, J., Chen, Y., Heuven, V. J. v., & Schiller, N. O. (2018). Dynamic effect of tonal similarity in bilingual auditory lexical processing. *Language, Cognition and Neuroscience*, *34*, 580–598.

Yu, Z., & Schwieter, J. W. (2018). Recognizing the effects of language mode on the cognitive advantages of bilingualism. *Frontiers in Psychology*, *9*, 1–6.

Függelék

6. táblázat. A felolvasott magyar megnyilatkozások párbeszédbe ágyazva

A vizsgált magyar hangsorok (helyesírásban és fonetikus lejegyzésben)

	Alternatív kérdés részeként, illetve egy szótagú kérdésként	Felszólításként	Kijelentésként
tő **[tøː]**	*– **Tő** vagy tó?* *– Nem tudom. **Tő?***	*Olaszországban a következő jókívánsággal szokták bátorítani a tőkéket: **Tő!** Váljék belőled jó bor!*	*– Hogy hívják a szőlő szárát?* *– **Tő.***
sző **[søː]**	*– **Sző** vagy fon?* *– Nem tudom. **Sző?***	*Egy távoli országban a halált jelentő szó hasonlít a magyar „sző" szóhoz. Ezért gyakran felszólítják a halált, hogy menjen el messzire: **Sző!** Távozz tőlünk!*	*– Mit csinál Peti a szövőszéken?* *– **Sző.***
kő **[køː]**	*– **Kő** vagy lő?* *– Nem tudom. **Kő?***	*Pali legjobb barátja egy kavics, akit Kőnek hívnak. Gyakran így szól hozzá: **Kő!** Gyere ide hozzám!*	*– Mi az a kemény anyag?* *– **Kő.***
hő **[høː]**	*– **Hő** vagy hó?* *- Nem tudom. **Hő?***	*Amikor nagyon meleg van, az emberek felszólítják az időjárást: **Hő!** Légy egy kicsit alacsonyabb!*	*– Milyen mérővel szoktunk lázat mérni?* *– **Hő.***
cső **[tʃøː]**	*– **Cső** vagy csá?* *– Nem tudom. **Cső?***	*A vízvezetékszerelő 10 óra munka után így szól a vízvezetékhez: **Cső!** Most már nehogy kilyukadj nekem!*	*– Mi az a henger, amiben folyik a víz?* *– **Cső.***

Egy beszédkutatási kísérlet a *hát* diskurzusjelölő típusainak feltárására

Gocsál Ákos[1,2], Szeteli Anna[3], Szente Gábor[3], Alberti Gábor[3]

[1] *HUN–REN Nyelvtudományi Kutatóközpont*
[2] *Pécsi Tudományegyetem Művészeti Kar*
[3] *PTE Nyelvtudományi Tanszék ReALIS Elméleti, Számítógépes és Kognitív Nyelvészeti Kutatócsoport*

Abstract

Although many believe that the frequently occurring Hungarian discourse marker *hát* ('well/so') is a superfluous filler, previous research has pointed out its multifunctional nature. It harmonizes the communicating minds, preparing the listener for receiving new information. *Hát* therefore works like a semaphore and has a crucial role in human communication. This paper examines whether there are differences in duration parameters between *hát*s expressing different meanings. In this study, 53 speakers (28 females, 25 males, all university students, speakers of standard Hungarian) read a text, imitating spontaneous speech. In the utterances, *hát* appeared in ten different functions (h1 straightforward, h2 uncertain, h3 uneasy, h4 teasing, h5 resigning, h6 introductory, h7 summative, h8 evaluative, h9 sentence-final inferring, hf sentence-final confirming). The duration of *hát*s and of the following pauses, where *hát* was not in a final position, were measured. The duration of *hát*s with summative, straightforward, introductory, or evaluative functions was shorter, while those expressing negative attitudes or uncertainty were longer. Teasing represented a separate category, as well as the two sentence-final types. Differences in the duration of the pauses following the *hát*s were also found. Pauses following the uneasy *hát*s were significantly longer than those following the straightforward and summative ones. The results confirm the significance of prosody in different uses of *hát*. Possible uses of the results in speech technology and language teaching are raised, and it is also highlighted that *hát* is a natural element of the Hungarian language.

1. Bevezetés

A nyelv tudományos kutatása során rengeteg olyan nyelvi adat és hasznos megfigyelés keletkezik, ami nem kerül ki az adott kutatók által választott leírási keretet preferáló szűkebb tudományos fórumok világából. Sajátos problémaorientációja, területspecifikus terminológiája, esetleg „nem bölcsészeknek való"

Email addresses: gocsal.akos@pte.hu (Gocsál Ákos), anna.szeteli@gmail.com (Szeteli Anna), szenteg8@gmail.com (Szente Gábor), alberti.gabor@pte.hu (Alberti Gábor)

Beszédtudomány – Speech Science 2024/1. 85–117. DOI: 10.15775/Besztud.2024.1.85-117

formális-logikai reprezentációs módszertana miatt ott marad az „elefántcsonttoronyban", nem válik közkincsévé egy szélesebb kutatói és oktatói körnek. Pedig az érdekes adatsorokat és új megfigyeléseket más tudományos paradigmák képviselői is tudnák hasznosítani, sőt a nyelvoktatás gyakorlatában is lehetne őket kamatoztatni. A magyar generatív szintaxis eredményeinek közkinccsé tétele is megkezdődött például ebben a szellemben (Alberti & Laczkó, 2018). Jelen munkánkban pedig egy olyan tanulmányunk (Szeteli et al., 2022) releváns eredményeit kívánjuk közkinccsé tenni a kísérleti beszédkutatás művelői számára – beteljesítendő a 2020-as Szeteli-előadás célkitűzését (Szeteli et al., 2020), amelyik a *Beszédkutatás* hasábjain elsősorban Dér Csilla (2010; 2012; 2017) által rendszeresen tárgyalt *hát* diskurzusjelölőről szolgál újdonságokkal (l. még Dér, 2022), akinek a vizsgálatai korpuszbeli mintavételen alapulnak. Maga a Szeteli-cikk (Szeteli et al., 2022) több jelenséget vizsgál a magyar transzformációs generatív szintaxiselmélet szempontrendszere alapján (É. Kiss, 2002), Varga (2016) erre alapuló fonológiai megközelítését továbbfejlesztve, kísérletes megközelítést választva.

Meggyőződésünk, hogy ennek hozadékai a korpuszbeli mintavételen alapuló vizsgálatokhoz képest számot tarthatnak a *Beszédtudomány* olvasóinak érdeklődésére is, ha a releváns eredményeket és megfigyeléseket kiemeljük az eredeti tudományos kontextusból és összevetjük a korpusz alapú megközelítésekkel (Dér 2010; 2012; 2017; 2022; Németh, 2020) – amire önálló szakaszt szentelünk (4. szakasz) a kísérleteink módszertanának (2. szakasz) és eredményeinek (3. szakasz) a bemutatását követően, az összegzést megelőzően (5. szakasz).

A téma iránti érdeklődésünket Schirm Anita munkái ébresztették fel. Schirm (2008; 2011a; 2011b) a mellett, hogy nyelvtörténeti alapokon végigvezeti a *hát* útját a hely- és időviszonyok kifejezésén át a diskurzusjelölővé válásig, arra mutat rá, hogy a külvilágra közvetlenül utaló, vagy legalábbis viszonylag „könnyen kategorizálható" – például jól meghatározható mondattani kategóriával rendelkező – nyelvi elemekről sokat tudunk, illetve tudásunkat bevett módszertanokra támaszkodva növelhetjük. Ezzel ellentétben a hétköznapi kommunikáció – főként a spontán beszéd – szervezésében kulcsfontosságú szerepet betöltő diskurzus-

jelölőkről a formális nyelvészet szempontjából nehezen tudunk új megállapításokkal előállni. Ez egyrészt azoknak a dekategorizációs folyamatoknak tudható be, amelyeken a diskurzusjelölők jellemzően átesnek, nem hagyva morfológiai és mondattani támpontot a vizsgálódónak. Másrészt (történeti) poliszém rendszert építenek ki, ezáltal szemantikai és pragmatikai megítélésük is gyakran homályos. Dér & Markó (2017) a *hát* diskurzusjelölőt fokozottan multifunkcionális jelölőként határozzák meg; ami abban nyilvánul meg, hogy az „(...) elem egyazon aktuális használatában több diskurzusszintet érintő feladatokat lát el, és ez a jellemző rá általában" (Dér & Markó, 2017, 105). A szerzők Fischert (2006: 13–14) követik abban az alapfeltevésben, hogy a diskurzusjelölő poliszém rendszert kiépítve, egymással összefüggő interpretációkkal bír. Ez a homályosság és meghatározhatatlanság, illetve az őket jellemző szélesebb körű funkcióbeli potencialitás az esetenként agresszív nyelvművelés és a közoktatás útján a civil szférába szivárogva megbélyegzést, nyelvi stigmatizációt eredményezhet, ami a *hát* diskurzusjelölő esetében különösen jelentős mértékű.

A tudományos megismeréssel egy lépést tehetünk afelé, hogy az anyanyelvi nevelés során informatívabbak lehessünk a magyar spontán beszéd (sok műfajban) leggyakoribb (Dér & Markó, 2007, 63, idézi Schirm, 2011a, 28) diskurzusszervező elemének tekintetében (l. még Schirm, 2015, 2017, 2021). A mag- és periférikus jelentés/előfordulás (Bell, 1998) diakrón/szinkrón vizsgálatán felül a pontos pragmaszemantikai jegyek meghatározása további lehetőséget teremt a hatékony összehasonlításra más nyelvek lexikai készletével és egyéb diskurzusszervező eszköztárával, amit a magyar mint idegennyelv oktatása, illetve a fordítástudomány is hasznosíthat. Két nyelv diskurzusjelölői között fellépő jelentéskomponensbeli különbség meghatározásával növelhető a pragmatikai tudatosság, csökkenthető a pragmatikai transzfer. A pragmaszemantikai jegyek összekötésének kísérlete a hanganyagokban fellelhető prozódiai variációkkal pedig nyelvtechnológiai szempontból érdekes kihívás.

Jelen munkánk előzményeként említhetjük tehát Szeteli et al. (2022) formális pragmaszemantikai elemzését a ReALIS (Alberti et al., 2019, 2021) nevű rendszerben, korábbi eredményeik szintetizálásával (Alberti, 2016; Szeteli &

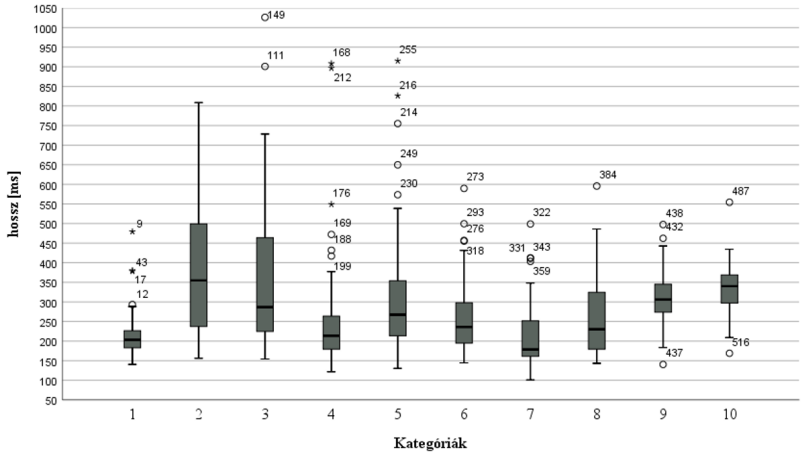
Alberti, 2018; Szeteli et al., 2019a). Ez a formális jelentés- és nyelvhasználat-modellező elmélet az egymással kommunikáló elmék kölcsönös és élethossziglani tudásának reprezentálására épül, egy formális diskurzusreprezentációs irányzat nyomdokain haladva (Kamp et al., 2011). Bölcsészeknek szánt bemutatását Alberti és Szeteli (2018), illetve Alberti (2020) kínálja. A ReALIS az említett diskurzusreprezentációs megközelítést – abból a megfontolásból kiindulva, hogy a diskurzusok reprezentálását egy tágabb kontextusban, a „diskuráló" elmék reprezentálása révén kell elvégezni – egy sok tekintetben ellentétes irányzat, a funkcionális/kognitív nyelvészet (pl. Langacker et al., 2017; Nuyts, 2017; Kugler, 2012) szempontjainak és eredményeinek az érdemi integrálásával kívánja meghaladni (Alberti et al., 2019; Szeteli et al., 2019a). A ReALIS alapállása az, hogy egy mondat interpretációjához inherens módon hozzátartozik annak figyelembevétele, hogy a feladó mit tud, mire vágyik és mire képes. Ez a címzettnek szóló (major) mondattípusok esetében (felszólító, kijelentő, kérdő; Sadock & Zwicky, 1985; König & Siemund, 2007) kiegészül azzal, hogy a feladó mit feltételez ugyanezen modális dimenziókban a címzettjéről – az mit tud, mire vágyik és mire képes –, azaz hogyan *mentalizálja* a címzettet (Wimmer & Perner, 1983), potenciálisan beleértve azt is, hogy a címzett feltételezhetően hogyan mentalizálja őt. Valamint a major mondattípusok esetében kiegészül valamilyen céllal: ami egy olyan szándék, amely arra irányul, hogy a címzettben váltson ki egy bizonyos szándékot.

Jelen munkánk célja, hogy – összhangban Schirm (2011a; 2011b) törekvésével – példákkal és a rendszerük feltárásával igazoljuk, hogy a magyar szóbeliségben leggyakrabban előforduló szó, a *hát* nem holmi „töltelékszó" – ami fölösleges, szószaporító dagályosság, ebből adódóan tehát kerülendő –, hanem a hallgatóval közvetlen szóbeli kontaktusban álló beszélő jelzése, hogy az mennyire könnyen – vagy éppen nehezen – tudja majd megemészteni a hallott információt (Alberti, 2016). A *hát* tehát elősegíti a kommunikáló elmék működésének az összehangolását, jelezve, hogy a hallgató számára (is) kézenfekvő információ érkezik, vagy éppen ellenkezőleg: nehezebben elfogadható, meglepő, és/vagy a hallgatót esetleg kifejezetten kellemetlenül érintő információra kell felkészülni. A *hát* tehát az

értelmi és/vagy az érzelmi területen nyújt előrejelzést a hallgató számára, hogy milyen módon feldolgozandó információt kap éppen.

Felmerül a kérdés, hogy hogyan lehet képes egy szó, a *h-á-t* hangsor jól láthatóan ellentétes tartalmak kifejezésére. A válaszunk az, hogy úgy működik ez a különleges nyelvi elem, mint a vasúti szemafor (Alberti, 2016): különféle prozódiai megvalósításai jelzik az éppen közvetítendő tartalmat, mint ahogy a szemafor különböző állásai a vasúti pálya állapotait. Ebből adódik, hogy a *hát* valóban a szóbeliséghez kötődik; a szóbeliségtől „távol álló" írásbeli stílusokban – például a tudományos stílusban – való használatát ennek alapján nem indokolatlan tiltani, vagy legalábbis olyan jól körülhatárolt esetre korlátozni, mint a következtetés kötőszói kifejezése. Megjegyzendő azonban, hogy a szóbeliséghez közel álló írásbeli stílusokban gyakori a *hát* használata (Kocsis, 2021); az internetes kommunikáció résztvevői nem tartják a maguk számára mérvadónak a helyesírás és a nyelvművelés hagyományos szabályait, sokszor fontosabbnak tartják a prozódiai megvalósulás (és az alaphangulat) valamiféle jelzését (pl. *Hááát öööö erre csak egy szóval tudnék válaszolni...Birspálinka....* bejegyzés a http://vadaszforum.net/forum/index.php?topic=101.1785 oldalon, megtekintve 2023. 04. 28-án).

A prozódiai megvalósulásokban jelentkező különbségeket jól mutatják a Szeteli et al. (2022) kísérletében nyert adatok. A szerzők a különféle pragmatikai funkciókban megjelenő *hát*-okat tíz kategóriába sorolták. Ezzel bővítették a korábbi kutatásaikban vizsgált típusokat (Alberti, 2016; Szeteli & Alberti, 2018; Szeteli, 2019; Szeteli et al., 2019b) a hazai beszédkutatásban feltárt további típusokkal (Dér & Markó, 2017; Dér, 2017) – ugyanakkor az egységesség kedvéért egyetlen párbeszédsémára szorítkoztak. E kísérleti séma keretében megvizsgálták az egyes kategóriákba sorolt *hát*-ok millisecundumban mért hosszait (1. ábra). A dobozdiagramról nemcsak a mediánok nagy változatossága állapítható meg, hanem a szélsőséges eseteké is. A 7. típus legrövidebb ejtése (kb. 100 ms) és a 3. típus leghosszabb, 1000 ms-nál is hosszabb ejtése között mintegy tízszeres a különbség.



1. ábra. A *hát* ejtésének a hossza – 100 ezredmásodperctől több mint 1000-ig.

Jelen munkánkban bemutatjuk Szeteli et al. (2022) kísérleti sémáját, valamint további három prozódiai grafikonját, különféle példákkal illusztrálva jellemezzük a tíz *hát*-típust (2–3. szakasz), majd összevetjük a nyert eredményeket és megfigyeléseket a korpuszbeli vizsgálatokból nyertekkel (4. szakasz). A prozódiai tényezők kísérletes alapú statisztikai elemzésével új eredményekkel szolgálunk a *hát* poliszém rendszeréről, illetve megerősítjük korábbi vizsgálatok eredményét.

2. Anyag és módszer

A kísérletben 53 adatközlő vett részt (28 nő és 25 férfi, pécsi egyetemi hallgatók). Az adatközlők a kísérletvezető kérésére spontán beszédet imitálva diktafonba olvastak szövegeket. A rögzítés 44,1 kHz-es mintavételezéssel, 16 bites felbontással, wav formátumban történt. A felolvasás minden esetben ugyanazon szimulált élethelyzethez kötődött: egy elképzelt szerelmespár moziba készül. Egyikük – az adatközlő – kiválasztja az egyik filmet, majd a kísérletvezető – mint a szerelmespár másik tagja – felteszi a kérdést: Na, melyik filmet választottad? Ezt követően az adatközlő közli a választását. Válaszában szerepel a *hát* valamely funkciójában. A résztvevőknek adott, részletes instrukciókat

tartalmazó lapot az 1. függelékben mutatjuk be. A tekintett funkciók az alábbiak (megtartva Szeteli et al., 2022 jelölését): h1 határozott, h2 bizonytalan, h3 aggodalmas, h4 incselkedő, h5 lemondó, h6 mondandóindító, h7 összegző, h8 értékelő, h9 összegző-záró, hf nyomatékosító-záró. Az 53 adatközlőtől így összesen 530 bemondást rögzítettünk és dolgoztunk fel.

A tegezés kapcsán felmerülhet a kísérletvezető és az adatközlők közti viszony kérdése; például az, hogy a feladatszituáción kívül, a hétköznapi életben tegeződtek-e vagy magázódtak. Ezzel kapcsolatban azt jelenthetjük ki, hogy nem ismerték egymást korábbról. Ugyanakkor az a tény, hogy a kísérletvezető éppúgy egyetemista volt, mint az adatközlők, természetessé tette a kísérleti séma által megkövetelt tegeződést. A kísérlethez kötődően nem készültek retrospektív interjúk.

A felvételeket a Praat 6.0.24. programban elemeztük (Boersma & Weenik, 2017), ahol kinyertük a kvantitatív feldolgozáshoz szükséges időtartamokat. Megmértük az egyes *hát*-ok hosszait, majd a nem mondatvégi esetekben a *hát*-okat követő szünetek hosszát. A matematikai statisztikai számításokat az SPSS 23 programban végeztük, a szükséges esetekben a program által felajánlott eljárásokat kiegészítve a Loftus & Masson (2014) által javasolt korrekciós technikával (az ehhez kötődő számításokat ugyancsak az SPSS 23 programra bízva). A fonetikai adatokat Varga (2016) kontúr alapú intonációs elméletében interpretáltuk. Az irreguláris zönge jelenségével (McGlone, 1967) mi is szembesültünk (Markó, 2013, 19); sőt kimutattuk, hogy az egyik *hát*-típus (h_f) karakterisztikus jegyének tekintendő. Az a módszer hátránya a társalgáselemzés területén irányadó elvek viszonylatában (Schegloff, 1996; Mondada, 2013; Németh, 2020), hogy az adatközlők nem természetes körülmények között, spontán módon produkálták a *hát*-okat. Előnye ugyanakkor – kihasználva a *hát* diskurzusjelölő válaszjelölő funkcióját (l. pl. Kiefer, 1988; Németh, 1998; Schirm, 2011a, 28-45; Markó & Dér, 2011; Kondacs, 2016; Németh, 2020) – a végsőkig egységesített kísérleti séma, amelynek az alkalmazásával a különböző típusokhoz tartozó *hát*-ok statisztikai alapon is összehasonlíthatóvá válnak, miközben az adatközlők törekednek a spontán ejtésre. Úgy véljük, kutatói instrukciók nélkül hasonló

elemzés nem lenne lehetséges, vagy legalábbis igen nagy nehézségekbe ütközne az egyes *hát*-típusok esetleges előfordulása miatt

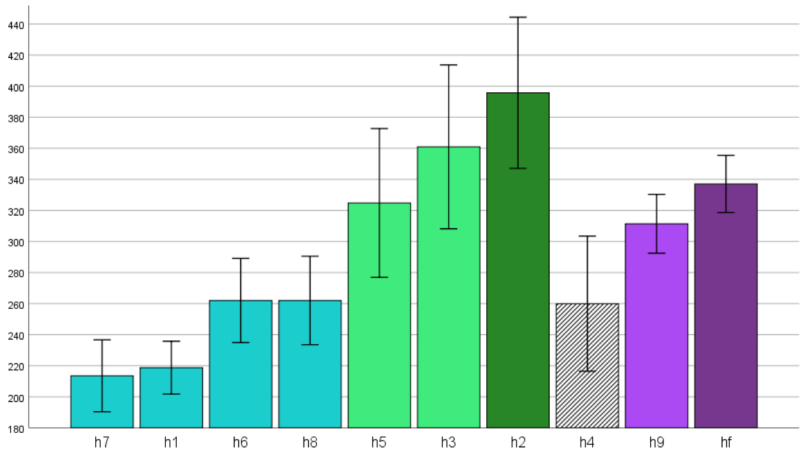
3. Eredmények

3.1. Időtartamok

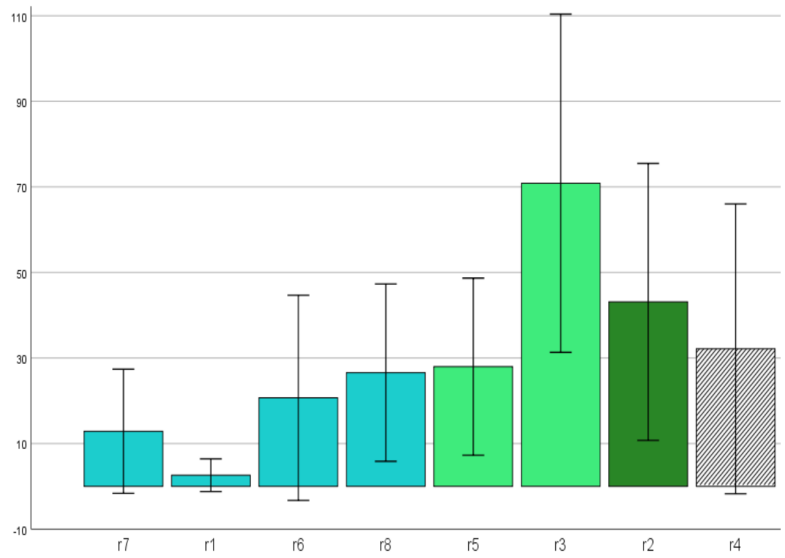
Az egyes *hát*-típusokon mért hosszadatok alakulását a hozzájuk tartozó 95%-os konfidenciaintervallumokkal a 2. ábra mutatja. Az egyes oszlopok nem az egyes típusokhoz tartozó jelölés sorrendjében láthatók, hanem a *hát*-okat funkciók szerint még külön csoportosítottuk, majd az egyes csoportokon belül a hosszuk szerint növekvő sorrendbe állítottuk őket. Az első négy *hát* (h7, h1, h6, h8) határozottságot fejez ki. A következő három inkább negatív attitűdöt vagy bizonytalanságot tükröz (h5, h3, h2), külön kategóriát jelent az incselkedés (h4), majd a mondatzáró *hát*-típusok (h9, hf) következnek.

Jól látható, hogy a konfidenciaintervallumok között több esetben nincs átfedés, azaz az adott *hát*-típusok hosszértékei között szignifikáns különbségek találhatók. Ilyen például a h7 és a h5, mivel a h7-hez tartozó konfidenciaintervallum legmagasabb pontja 240 ms alatt van, míg a h5-höz tartozóé 270 ms fölött kezdődik; nem lehet tehát közös pontjuk. A h7 és a h1 hosszadatok viszont nem térnek el egymástól szignifikánsan, mert például a 220 ms mindkét konfidenciaintervallumban ott van.

A továbbiakban megmértük a *hát*-ok és az őket követő szavak közötti szünetek hosszát. Az így kapott adatokat a 3. ábra mutatja, a *hát*-típusok megfelelő adatát ugyanabba a sorrendbe rendezve, mint a 2. ábrán. Az ábrán r jelöli az adott *hát*-típushoz tartozó szünethosszokat. A záró helyzetben megjelenő két *hát*-típus itt értelemszerűen nem szolgáltat adatokat. Ha itt is a 7. és az 5. *hát*-típust vetjük össze – amelyek hosszadata szignifikáns különbséget mutat –, akkor most az r7 és az r5 követőszünethosszokhoz tartozó konfidenciaintervallumok összevetése alapján azt mondhatjuk, hogy e két *hát*-típus ebben a paraméterben nem különbözik szignifikánsan. A 3. *hát*-típus különbözik szignifikánsan a 7.-től és az 1.-től.



2. ábra. A 10 vizsgált *hát*-típus ejtésének a hossza, típusonként rendezven.



3. ábra. A 8 vizsgált *hát*-típust követő (kitöltött vagy kitöltetlen) szünet hossza.

3.2. A hát-típusok kategorizálása

A 2. ábra magyarázatánál már előrebocsátottuk, hogy a tárgyalt tíz *hát*-típus az általuk ellátott funkciók, illetve helyzetük szerint kategóriákba sorolhatók. Az alábbiakban ezeket a kategóriákat tárgyaljuk részletesebben, mégpedig

a 2. ábrán látható sorrend szerint. Az egyes kategóriákon belül pedig a *hát*-hossz növekedése szerint vesszük sorra a *hát*-típusokat.

3.2.1. Határozott hát-ok: h7 összegző, h1 határozott, h6 mondandóindító, h8 értékelő

Ebben a szakaszban négy olyan *hát*-típust ismertetünk, amelyek mondatéliek, és közös vonásuk az alapvetően határozottnak mondható karakter.

Az alábbi (1) példában szemléltetett 7. *hát*-típus (összegző) az említett rendező elv alapján a legrövidebb átlaghosszal jellemezhető. A beszélgetésben a fordulózáró mondat élén elhelyezve a beszélő azt jelzi vele, hogy összegzi az addigiakat, eljutva arra a végkövetkeztetésre, aminek alátámasztása folyt az adott beszédfordulóban. Mindezekből az említett határozott jelleg úgy adódik, hogy a beszélő céltudatos érvelést valósít meg a fordulóban, amit mondatról mondatra a hallgató is befogad – amit el is vár tőle a beszélő: mire a beszélő eljut a végkövetkeztetés kimondásáig, addigra a hallgatótól is elvárhatónak gondolja, hogy eljusson ugyanarra a végköveztetésre.

(1) h7: összegző

B először szeretné elmondani a filmekről kialakított véleményét, majd az egészet összefoglalva hirdetne eredményt.

Az ízléstelen és felszínes vígjátékokból a múltkor végképp elegünk lett, a nyomasztó északi drámákért pedig még én sem rajongok, bár én alapjában véve kedvelem a komolyabb műfajokat. A krimi viszont általában mindkettőnknek be szokott jönni. ***Hát a sekélyes vígjátékkal és a lehangoló drámával szemben a krimi tűnik nézhetőnek.***

A következő, (2) példában szemléltetett „határozott" elnevezésű típus szintén azt sugallja, hogy a beszélő – ezúttal magában – végiggondol többféle alternatívát és érvet, majd egyértelmű döntésre jut. A kísérlet részvevőinek ezt tudtára is adtuk: *biztos … a jó döntés[ben]* a beszélő; és ebbe beleértendő az is, hogy úgy gondolja, hogy a hallgató is ugyanerre a döntésre jutna (hiszen ugyanazokat az alternatívákat és érveket kell fontolóra vennie). A két elme tevékenységének

a legteljesebb összehangolásához hozzátartozik még az a harmadik tényező is, hogy a beszélő azt gondolja, hogy a hallgató „ismeri őt annyira", hogy ehhez a döntéshez fog eljutni. A kísérlet résztvevőivel azért olvastattuk el előre a teljes beszélgetést, hogy aki felolvassa a megmutatott szöveget, az valóban el is tudja játszani, előre tudva, hogy szerepe szerint (majd) meg is indokolja a döntését a beszélő.

(2) h1: határozott

Ebben a teszthelyzetben B biztos benne, hogy az egyetlen jó döntést hozta meg. Ez az indoklásból kiderül, ami a színtiszta igazságot tartalmazza, minden csúsztatás, udvariaskodás vagy képmutatás nélkül.

Hát a krimit, az mindkettőnknek be szokott jönni. *Az ízléstelen vígjátékokból a múltkor végképp elegünk lett, a nyomasztó északi drámákért pedig még én sem rajongok, bár én alapjában véve kedvelem a komolyabb műfajokat.*

Közel szignifikáns a (3) példában alább bemutatott 6. *hát*-típus – a mondandóindító – hosszadatának eltérése az eddig tárgyalt, két legrövidebb hosszadatától. A *hát* hatása ezúttal nem egy információra, hanem az információval való valamilyen beszédcselekvésre irányul. A jelenséget Asher és Lascarides (2003) metatalk néven említi. A határozottság annak szól, hogy az „eredményhirdetés" mint beszédaktus-altípus (Austin, 1962/1975; Searle, 1979) egyáltalán nem váratlan az adott helyzetben, tekintve a mozilátogató párocska alapsztoriját. Legfeljebb az a váratlan, hogy a *hát*-ot nem maga a nyertes mozifilm megnevezése követi. Feltételezhető, hogy ebből a „szintkülönbségből" adódik a közel szignifikáns különbség a h6 és a h1/h7 hosszátlagok között, ahogy a 2. ábrán megfigyelhető.

(3) h6: mondandóindító

B alig várja, hogy beszámolhasson a döntéséről, és kielemezhesse a két kieső filmet.

Hát most eredményt hirdetek. *A krimi nyert, mivel mindketten szeretjük, ellenben a másik kettő egyáltalán nem győzött meg. A dráma túlértékeltnek tűnik, a vígjáték pedig sekélyesebbnek, mint valaha bármelyik ebben a műfajban.*

A (4) példában bemutatott 8. *hát*-típus esetében is megjelenik az (1-2) tesztszövegekben foglalt döntéstámogató érvelés, de ezúttal az érvek sorának elindításához, a beszédforduló elején elinduló értékelési folyamathoz – és nem a fordulózáró konklúzióhoz – társul a *hát*. Egyebekben viszont a beszélő azt feltételezi, hogy az értékelés mozzanatai kézenfekvőek lesznek a hallgató számára (is).

(4) h8: értékelő

B fontosnak tartja kifejteni a véleményét az elutasított filmekről, mielőtt eredményt hirdetne.

Hát a vígjátékok szörnyűek, a drámák meg nyomasztóak. *Maradt a krimi. Szerintem azt mindketten szívesen nézzük majd.*

A 2. ábra alapján – szignifikáns hosszkülönbség híján – nyitva kell hagynunk a kérdést, hogy a szakaszban tárgyalt négyféle szerkesztési módja a *hát* odaillesztésének a döntéstámogató érveléshez egyáltalán hány eltérő *hát*-típusnak tekintendő. Akár az is felvethető, hogy egyetlen típusról van szó (és a *meta-talk*-választás függetlenül rendelkezésre álló opció). E hipotézis mellett a hosszkülönbségek – verziótól függetlenül – azzal korrelálnak, hogy a beszélőnek a kísérleti alanyok az adott kísérőszöveg alapján milyen mértékű határozottságot tulajdonítottak.

3.2.2. Hááát most elbizonytalanodunk…: h5 lemondó, h3 aggodalmas, h2 bizonytalan

Ebben a szakaszban három olyan *hát*-típust tárgyalunk, amelynek (a szakaszon belül növekvő sorrendbe rendezett) ejtési hossza a 2. ábra tanúsága szerint szignifikánsan hosszabb (vagy majdnem az), mint az előző szakaszban

tárgyalt négy határozott jellegű *hát*-típusé. A követőszünetek hosszáról is majdnem ugyanez mondható el (3. ábra). A hosszabb ejtés, illetve a követőszünettel való további „időhúzás" amúgy ikonszerű jelölésnek is tekinthető: a hosszabb gondolkodásra, „tépelődésre" hívja fel a figyelmet (Németh, 2020).

Az (5) példában a lemondó *hát*-típust mutatjuk be, ez az 5. típus Szeteli et al. (2022) munkájában. Nem a döntésében bizonytalan a beszélő, sőt úgy gondolja, hogy várhatóan A sem fog meglepődni az érveken és a végkövetkeztetésen –, hanem egyszerűen nem örül annak, ami „kijött". Erre a nem igazán jó hírre készíti fel a hallgatót a *hát* hosszabb ejtésével.

(5) h5: lemondó

B nincsen róla meggyőződve, hogy élvezni fogja döntése gyümölcsét. Egy pillanatig azt is fontolgatja, hogy úgy nyilatkozzon: inkább maradjunk itthon! Végül azonban egykedvűen bár, de mégiscsak kiválasztja az egyik filmet, mivel tudja, hogy A semmit nem utál jobban az otthon való ücsörgésnél.

Hát a vígjátékot. *Nem igazán repesek, de unom már a krimiket, ennek a rendezőnek a drámái meg iszonyatosan nyomasztanak. Talán jobb lesz, mint otthon ülni.*

A következő, (6) példa az aggodalmas típust illusztrálja. Itt a még hosszabb ejtés (az 1. ábrán az egyetlen 1000 ms-ot meghaladó ejtés itt történt, nyilván az „eljátszást" legkomolyabban vevő kísérleti alanytól) a dupla negatív érzésnek tulajdonítható: nemcsak hogy nem fog örülni a hallgató a kiválasztott filmnek, de még a háttérben álló fő indok is féltékenységet gerjeszthet. A kiemelkedő követőszünetet (3. ábra) ennek a másodlagos negatív faktornak tulajdonítjuk: a beszélő a hosszabb követőszünettel azt sugallja a hallgatónak, hogy „te is hagyj egy kis időt magadnak a döntés feldolgozásához, meggondolatlan azonnali reakció helyett!"

(6) h3: aggodalmas

B úgy érzi, hogy a krimiket egyre jobban unják, a vígjátékokról pedig a múltkor már A is kénytelen volt elismerni, hogy már nem is humorosak, csak gusztustalanul alpáriak. A viszont ki nem állhatja a nyomasztó északi drámákat. B viszont nagyon jót hallott erről az izlandi filmről egy olyan csoporttársától, akinek nagyon tiszteli az ízlését, A viszont kifejezetten féltékeny pont erre a „jófej csoporttársra" – talán nem is teljesen alaptalanul. B végül úgy dönt, hogy él a választás jogával, de baljós előérzetei vannak... Kínos lesz, ha A nagyon firtatja, hogy ki is ajánlotta ezt a filmet...

Hát a drámát. *Tudom, hogy nem nagyon rajongsz ezért a műfajért, de mintha ezt az izlandi filmet valahol nagyon dicsérték volna. Asszem valami díjat is nyert. A krimiket már kissé unom, a mostanában futó vígjátékok alpáriságából pedig a múltkor már neked is eleged lett, úgy emlékszem.*

A kutató által megadott instrukciók alapján a (7) példában bemutatott 2. teszthelyzetben tekintendő a beszélő a legbizonytalanabbnak a döntésében.

(7) h2: bizonytalan

B nem biztos benne, hogy a legjobb döntést hozta meg, illetve úgy érzi, hogy a pillanatnyi kínálat alapján nincs olyan döntés, amit igazán lelkesen tudna képviselni. Közös tapasztalatnak gondolja a következőket. Korábban néhány krimi kicsit unalmasnak bizonyult, de azért nézhetőnek. A vígjátékot szokott választani, de a múlt héten általa választott vígjáték ízléstelenségén már ő is kiakadt. A drámákat pedig A ki nem állhatja.

Hát a krimit*. Az azért többé-kevésbé mindkettőnknek be szokott jönni. Az ízléstelen vígjátékokból a múltkor már neked is eleged lett, nyomasztó északi drámákért pedig még én sem rajongok. Vagy nagyon unod már a krimiket?*

3.2.3. Incselkedés hát-tal: h4 incselkedő

A Szeteli et al. (2022) tanulmányában 4. sorszámú tesztben a kutató incselkedésre kérte a kísérlet résztvevőit. Az ötletet Csukás István Pom Pom-meséi

szolgáltatták, ahol Picur Pom Pommal a *Megvárhatlak?* kérdésre így incselkedik: *Hát...*

(8) h4: incselkedő

B úgy érzi, hogy a krimiket egyre jobban unják, az északi drámákat pedig A ki nem állhatja. Nyomott hangulatba kerül tőle, és egész este azzal szekálná, hogy biztosan az a vonzó csoporttársa ajánlgatja neki az ilyen baromságokat. Úgy dönt hát B, hogy A kedvéért a vígjátékot választja (annyira alpári csak nem lehet, mint a múlt heti, mert arról a mélypontról már csak felfelé vihet az út), de azért kicsit incselkedik vele, megenged magának némi játékot, ha már ilyen nagylelkű volt. Az B célja, hogy A először el se higgye, hogy a múltkori szörnyűség után tényleg a vígjátékot választotta – aztán viszont hogy fog majd örülni (és milyen jó lesz majd a film után az este)...

Hát a vígjátékot. [A: *A vígjátékot?!*] *A vígjátékot hát! Tudom, hogy mennyire élvezed az amcsi vígjátékokat, és úgy szeretem, ha vidám vagy este.*

A 2. és a 3. ábrán azt látjuk, hogy az incselkedő *hát*-típus hosszadatai (h4, r4) a hosszabb határozott *hát*-típusokéval mutat párhuzamot (h6, h8, r6, r8). Ez némileg váratlan eredmény; kézenfekvőnek tűnhet, hogy az incselkedés részét képező információkésleltetést a *hát* megnyújtásával valósítjuk meg. Hogy valójában mi történt, azt ezúttal az 1. ábra világítja meg. Miközben h4, h6 és h8 mediánja egyaránt 200 és 250 ms közé esik, addig a maximális érték h6 és h8 esetében 600 ms alatt marad, h4 esetében viszont eléri a 900 ms-ot. Ennek magyarázata az lehet, hogy az incselkedést a kísérlet résztvevőinek egy része valóban a *hát* megnyújtásával próbálta jelezni, mások viszont a *hát* kimondásán gyorsan túlestek, és a névelőn vagy a főnéven hajtottak végre szótagnyújtást.

A *hát* akár el is hagyható: [*aaa ... VÍGjátékot!*] – akár ilyen prozódiai mintázat alkalmazásával is kifejezheti valaki az incselkedést. Ez érvként szolgál a mellett, hogy az incselkedést ne a *hát* diskurzusjelölőhöz társuló

pragmatikai kontribúciónak tekintsük. Ez a megközelítés meg tudja magyarázni azoknak a tesztalanyoknak a viselkedését is, akik a *hát*-ot olyan röviden ejtették, mint a határozott típusokat (3.2.1), és azokét is, akik olyan hosszan ejtették, mint a határozatlan típust (3.2.2), figyelembe véve az incselkedés összetett karakterét, a magában hordozott ellentétet. Az utóbbi esetben a hosszan ejtett *hát* az értelmi szinten indokolható: B nem a saját ízléséből fakadó, A számára (is) várható döntést hozta. Az előbbi esetben a *hát* rövid ejtése az érzelmi szinten indokolható: a B kedvére való döntés nem teszi szükségessé a 3.2.2. pontban tárgyalt hezitációs faktort. Ráadásul mindkét forgatókönyv azzal a csavarral is „megfordítható" – sőt talán megfordítandó –, hogy a beszélő időlegesen éppen ellentétes jelzést akar adni a hallgatónak, egy tizedmásodpercre becsapva őt, legalábbis a végső döntés felől nézve.

3.2.4. Mondatzáró hát-típusok: h9 összegző-záró, h_f nyomatékosító-záró

A (9) példában bemutatott 9. *hát*-típus a mondat végén (vagy legalábbis az ige mögött, arra csüggeszkedve), éppúgy hangsúlytalanul van jelen a mondatban (*non-accented* Varga (2016) rendszerében), mint a 3–4. szakaszban bemutatott *hát*-típusok. Funkciója megegyezik a mondatéli összegző 7. *hát*-típusával, arra ki is lehetne cserélni a (9) szövegben. Ejtési hossza azonban szignifikánsan nagyobb annál, ahogyan azt a 2. ábrán láthatjuk, összevetve h9 és h7 konfidenciaintervallumait. Ezt a különbséget inkább a dallamegység-kezdő és -záró helyzetnek tulajdonítjuk (kezdőpozícióban gyorsabban ejthető), mint a beszélői határozottság különbségének – mindamellett e kérdés eldöntését jövőbeli kutatásokra bízzuk.

(8) h9: összegző-záró

B nem igazán tud szívből választani, mivel nincs kedve egy újabb krimihez, és a dráma előzetese sem győzte meg igazán. Ezért arra a következtetésre jut, hogy A kedvében járni a legkézenfekvőbb.

Krimiből sokat néztünk mostanság, ez a dráma meg nyomasztónak tűnik. ***Nem volt jobb ötletem, a vígjátékot választottam hát.*** *Gondoltam, ezzel legalább neked kedvezek. Talán még nekem is tetszeni fog, ha kellően ráhangolódok.*

A következő, (10) példában bemutatott *hát*-típus annyiban kivételes, hogy hangsúlyt visel, amivel új dallamegységet indít (karakterdallamot Varga (2016) rendszerében) – ami persze megmarad egyetlen szótagosnak. Mint az 1 és 2. ábra mutatja, a leghosszabban ejtett *hát*-típusok közé tartozik; nyilván abból adódóan, hogy a hangsúlyos helyzetet meg kell mutatni a szegmentumon. Maga a hangsúlyos helyzet pedig a nyomatékosító funkció kézenfekvő kifejezése: „győzködni" kell a hitetlenkedő beszédpartnert.

(9) h_F: nyomatékosító-záró
B nem igazán tud szívből választani, mivel nincs kedve egy újabb krimihez, és a dráma előzetese sem győzte meg igazán. Ezért arra a következtetésre jut, hogy A kedvében járni a legkézenfekvőbb...
Hát a vígjátékot. [A: *A vígjátékot?!*] ***A vígjátékot hát!*** *Tudom, hogy mennyire élvezed az amcsi vígjátékokat, és úgy szeretem, ha vidám vagy este.*

4. A kísérletes megközelítés hozadékai a korpuszbeli mintavételen alapuló korábbi vizsgálatokhoz képest

Mint leszögeztük a Bevezetés végén, a 2–3. szakaszban ismertetett kísérletet úgy terveztük meg, hogy a korábbi kutatásainkban vizsgált *hát*-típusokat (Alberti, 2016; Szeteli & Alberti, 2018; Szeteli, 2019; Szeteli et al., 2019b) kiegészítettük a hazai beszédkutatásban feltárt további típusokkal (Dér & Markó, 2017; Dér, 2017) – ugyanakkor a kísérleti séma messzemenő egységessége kedvéért egyetlen párbeszédsémára szorítkoztunk. Az új – kísérletes – nézőpont nyilván új eredményekkel és megfigyelésekkel kecsegtet,

ugyanakkor ezek integrálása a korábbi – korpuszbeli mintavétel alapú – leírások rendszerébe korántsem triviális feladat. E szakaszban ennek látunk neki, elsődlegesen a *Beszédkutatás* hasábjain megjelent munkákra lapozva.

4.1. A figyelmet a követő egységre irányító diskurzusjelölői funkció

Dér (2010: 170) négy diskurzusjelölő összevetésében tárgyalja a *hát*-ot, ugyanakkor az egyik konklúzió tekintetében nem nyilatkozik róla, nem terjeszti ki rá a vonzó általánosítást: „az így és az ilyen esetében ... igen új, de egyre erősebben terjedő jelenség a figyelmet a követő egységre irányító diskurzusjelölői funkció."

A 3.2.1. és 3.2.2. pontokban a *hát*-típusok bemutatott „szemaforként való működése" azonban éppen „a figyelmet a követő egységre irányító diskurzusjelölői funkció" gyanánt fogható fel. Mi több, a *hát* ebben a funkcióban nem pusztán ráirányítja a figyelmet a követő egységre, hanem a prozódiai megvalósulása azt is előre jelzi, hogy hogyan kell a hallgatónak feldolgoznia az ott nyújtott információt, legalábbis a beszélő saját információállapota alapján kialakult meggyőződése szerint.

Érdekes kutatói kérdésként vetjük fel, hogy az így és az ilyen prozódiai megvalósulása (beleértve az esetleges ismételgetésüket, illetve a kitöltött vagy kitöltetlen követőszünetet) szintén szemaforként viselkedik-e a „követő egység" feldolgozásának a miként jét tekintve.

4.2. Nem pusztán a szó átvétele...

A Dér (2012: 131) által elvégzett korpuszvizsgálat azt mutatta meg, hogy „számos magyar diskurzusjelölő tipikusan az általa bevezetett diskurzusszegmens elején fordul elő, a kötőszói eredetű jelölők (pl. *hát, tehát, és, de*) szinte mind ilyenek; kivételt azok képeznek, amelyek kötőszóként sem vagy nem mindig tagmondatkezdő helyzetűek (pl. *meg, pedig, bár*). Ebből következően a társalgások beszédlépéseit bevezető elemként is gyakorinak kell lenniük."

A *hát* még a diskurzusszegmens-eleji diskurzusjelölők közül is kiemelkedett a vizsgált korpuszokban (Dér, 2012: 135–136): „az előfordulások több mint felét (581 db, 52,7%) mindössze háromféle egyszavas diskurzusjelölő adta ki, a *hát*, a *de* és az *és*. ... [A] szóátvételek során leggyakrabban (9 előfordulás és afelett) használt diskurzusjelölőket mutat[ó ábráján az látható, hogy] a *hát* kétszer gyakrabban fordult elő a második leggyakoribb *de* jelölőhöz képest, amely az *és*-sel körülbelül azonos arányban tűnt fel." Majd leszögezi (Dér, 2012: 140), hogy „a vizsgált jelölők egyike sem szolgált pusztán a szó átvételére, a két fő diskurzusjelölői szerepkörnek megfelelően az előzményekhez, illetve a kontextus valamely eleméhez való kapcsolást fejezték ki, illetve attitűdjelölést. A kapcsoló funkció természetszerűleg adódik a szóátvétel mint funkció mellé, hiszen a kezdő pozícióban lévő diskurzusjelölő célszerűen számos információt hordoz: hogyan integrálják a hallgatók a következő megnyilatkozás jelentését a diskurzus során eddig elhangzott információk sorába; hogyan viszonyul a követő megnyilatkozás az előzőhöz; milyen a beszélői viszonyulás az elhangzó megnyilatkozásokkal kapcsolatban."

A 3.2.1-3. pontokban bemutatott kísérleti sémában rögzítettük a *hát* szóátvételi funkcióját – igazodva ezzel a korpuszokban megfigyelhető alapvető szerephez –, majd ebből a rögzített nézőpontból összpontosítani tudtunk azokra a különbségekre, amelyek az imént számba vett funkciók közül elsődlegesen „az elhangzó megnyilatkozásokkal kapcsolatos beszélői viszonyulásra" vonatkoztak. Jelesül, a beszélő tekinthette a döntését a hallgatóval közösnek tekinthető evidenciák (pl. jelenlegi és korábbi filmválaszték) és következtetési szabályok (pl. melyik félnek milyen film fog várhatóan tetszeni) alapján kézenfekvőnek, vagy éppen ellenkezőleg, egyáltalán nem kézenfekvőnek – tekintve akár a racionalitást (korábbi döntések objektív mérlegelését), akár az érzelmi oldalt (kinek az akarata érvényesül, akár harmadik fél esetleges befolyását is figyelembe véve). A prozódia két hosszadatnak – magának a *hát*-nak, illetve a követőszünetnek – a skálaszerű opcióival bizonyult reagálni az előző mondatban felvázolt funkcionális

különbségekre, (óvatosan fogalmazva) felvethető hipotézissé téve azt az elkülönítést, hogy a *hát*-hossz a racionális dimenzióra érzékeny indikátor, míg a követőszünethossz az érzelmire. Megjegyzendő, hogy Szeteli et al. (2022) más fonetikai kísérletek eredményével összevetve arra jut, hogy a frekvencia a *hát* diskurzusbeli pozíciójának az indikátora; a hangerőről pedig az a sejtésünk, hogy – bár korrelál a két másik fonetikai jellemzővel – alapvetően a beszélői egyéniség indikátora.

4.3. Multifunkcionalitás a társalgásban

Dér (2017: 172) a *hát*-tal kapcsolatos vizsgálatait a BEA (Beszélt Nyelvi Adatbázis) korpuszán (Gósy, 2012) végezte el négy műfaj összevető elemzése révén. Redeker (1991), valamint González (2004: 78) felosztását követve (Dér, 2017: 169) az alábbi három pragmatikai diskurzusszinttel/-struktúrával számol (a szemantikai struktúra mellett):

„A pragmatikai relációk szintjén háromféle struktúra létezik, amelyek létrehozásához és fenntartásához (egyebek mellett) a pragmatikai diskurzusjelölők járulnak hozzá:

a) Retorikai struktúra: ide tartoznak a beszélői szándékokhoz, gondolatokhoz, cselekvésekhez kapcsolódó funkciók, például a témaváltás vagy a személyes értékelés.

b) Szekvenciális struktúra: két nagyobb funkciókör található itt, egyrészt a szegmenshatárok körülhatárolása (ezen belüli funkciók pl. a nyitó és záró szegmenshatár jelzése), másrészt a diskurzushálózat fenntartása.

c) Inferenciális komponens: itt szerepelnek azon kognitív kontextushoz kapcsolódó funkciók, amelyeken a beszélő és a hallgató osztozik, elsősorban az inferenciákat segítik elő vagy korlátozzák (funkciók pl. az arcfenyegetés csökkentése vagy az előfeltevések jelölése)."

Rámutat (Dér, 2017: 182), hogy az általa vizsgált négy műfaj közül „a társalgásokban jelenik meg a legtöbbféle funkció. (...)[A] legerősebb fokú multifunkcionalitás (...) a társalgásokra jellemző, és okát annak dialogikus

voltában találhatjuk meg: a szekvenciális struktúra kezelése több feladat ellátását igényli."

Az alábbiakban táblázatba foglaltuk, hogy a jelen tanulmányban vizsgált 10 *hát*-típus hogyan feleltethető meg a Dér (2017: 170–171) által felállított 12 összetett *hát*-típusnak, az utóbbiak bemutatási sorrendjét alapul véve (1. táblázat).

Mielőtt kommentálnánk az egyes típusokat, általános megjegyzéseket teszünk. Dér a három pragmatikai struktúrát aránytalanul veszi figyelembe: míg a retorikainak nevezett funkciót mind a 12 típusában specifikálja, addig az inferenciális és a szekvenciális funkció csupán egy-egy típusban kerül specifikálásra. A jelen tanulmányban kimutatott, funkcióval társuló prozódiai különbségek azt mutatják, hogy az inferenciális dimenziót is cizellálni kellene, bizonyos pragmatikai komponenseket ott – vagy ott is – megjelenítve a retorikai dimenzió helyett vagy mellett. A másik általános megjegyzésünk arra vonatkozik, hogy az. 1. táblázat végére helyezett Dér-féle KÉSL és TÉMV típusoknak egyik általunk vizsgált típust sem tudtuk megfeleltetni – ami a kísérletes sémánk természetes velejárója: a tesztalanyok egy előre megszerkesztett szöveget követtek, így elakadásra vagy témaváltásra „nem volt lehetőségük", szemben a spontán szövegalkotást bemutató beszélt nyelvi korpusszal. A további 10 Dér-típust meg tudtuk feleltetni a mi 10 típusunknak az első két oszlop tanúsága szerint; előre bocsátva, hogy a megfeleltetés nem mindenütt egyértelmű.

Kezdve a kommentálást a 3.2. szakasz 3.2.1. pontjában bemutatott határozott *hát*-típusainkkal, a KONK leírása leginkább az általunk összegzőként megnevezett h7 *hát*-típusra illik, ahol B megadott szövege a három választható film kommentálását követően összegzi egy *hát*-tal indított mondat formájában a filmválasztásra irányuló döntést. Ugyanakkor megjegyzendő, hogy a kísérleti sémánk szerint végső soron mind a tíz forgatókönyvünk arról szól, hogy a B szereplő milyen következtetési lépésekkel jutott el a

1. táblázat. A pragmatikai relációk szintjén definiált háromféle struktúra, amit Dér (2017: 170–171) a *hát*-típusok 12 összetett funkcióba való besorolására javasol.

DÉR	h_i	RETORIKAI	INF.	SZEKVENCIÁLIS
KONK	h7 összegző	következtetés, konklúzió		
ÉRT	h1 határozott (h8)	értékelés, véleménykifejezés		
IND	h6 mondandóindító	mondandó/téma indítása		szegmenshatárnyitás (h1, h2, h3, h4, h5, h6, h8)
EVID	h8 értékelő	evidencia jelölése		
ELL	h5 lemondó (h2, h3, h4, h5)	ellentét visszaadása		
GYENG	h3 ahhodalmas (h2)	tompítás/gyengítés (udvariasság kifejezése)		
MAGY	h2 bizonytalan (h3)	magyarázkodás, ... részletezés		
ÉRZ	h4 incselkedő (h2, h3, h5)	érzelmi többlettartalom (beleegyezés, csodálkozás, felindultság) kifejezése		
ÖSSZ	h9 összegző-záró	összegzés, történet (le)zárása		
NYOM	h_f nyomatékosító	nyomósítás/erősítés/élénkítés/figyelemfelkeltés		
KÉSL		elakadás, késleltetés jelzése		
TÉMV		mondandótól való eltérés (új téma, témaváltás)		

választásáig – és a következtetési szempontok megválasztásában vannak különbségek. Ahogy fent írtuk, az inferenciális pragmatikai komponensben kellene szisztematikus különbségeket tenni; nyilván úgy ítéljük meg, hogy az e tanulmányban megkezdett úton tovább haladva, de ennek kifejtése már egy önálló tanulmányt igényelne.

A határozott h1 típust társítottuk Dér értékelő/véleménykifejező ÉRT típusával; ezúttal is megjegyzendő, hogy az értékelő/véleménykifejező karakter a kísérleti sémánkból adódóan a többi típusra is jellemző, de a h1 típusunkhoz kevésbé tudtunk más Dér-típust rendelni, mint a többi *hát*-típushoz.

A mondandóindítás (IND) is hét típusunkat jellemzi, de a h6-hoz társítható leginkább, ahol B explicitté teszi, hogy mi is fog történni a beszédfordulóban: eredményt hirdet.

Az értékelőnek nevezett h8 típusunk nyilván az ÉRT típushoz is kapcsolandó (ahogy végső soron a többihez is), de az 1. táblázatban az evidenciajelölő EVID Dér-típushoz társítottuk, mivel a *hát*-tal indított mondat a filmválasztás hátterében figyelembe vett korábbi tapasztalatokat fogalmazza meg.

Áttérve a határozatlan típusainkra (3.2.2), nyilván mindháromhoz hozzárendelhető lenne az ellentét mint karaktervonás (ELL), továbbá az incselkedéshez is, és az ellentétek diszkutálása nyilván magyarázkodást (MAGY), illetve érzelmekre való hivatkozást is maga után von (ÉRZ). Mivel most az a célunk, hogy a tíz *hát*-típusunk vizsgálatát olyan módon mutassuk be, hogy tíz Dér-típus vizsgálataként is fel lehessen fogni, a táblázatban mutatott elsődleges társítást javasoljuk. E szerint a lemondó típusunkat (h5) társítottuk a Dér-féle ELL típushoz (úgy választ B, hogy ő maga sem örül annak, amit választott), az aggodalmast (h3) a „gyengítő" udvariassághoz (GYENG), és a bizonytalant (h2) a „magyaráz(kod)óhoz" (MAGY). Az aggodalmas (h3) társítása szorul kommentálásra: azt állítjuk, hogy az aggodalmaskodás explicit kifejezése (a leírás alapján sokan így is mondták ki: *hááát öö ... a drámát*) udvariassági tényező. A beszélő jelzi, hogy a hallgató számára kedvezőtlen döntést hosszadalmas vívódás után hozta meg, elismerve, hogy kényes a helyzet. A határozott prozódia lenne udvariatlan az adott helyzetben, hiszen arra utalna, hogy a beszélő ignorálja a hallgató szempontjait és/vagy „csak azért is" a hallgatói szempontok ellenében dönt.

Az incselkedés (h4) esetében (3.2.3) éreztük legjellemzőbbnek az érzelmi többlettartalom (ÉRZ) megjelenítését, bár nyilván valamennyi határozatlan típusban eleve ott van a határozatlanságból fakadó – kellemetlen – érzés; különös tekintettel arra, hogy a kísérleti sémában a határozatlanság egy választási helyzethez kötődik, ami a beszélő számára így vagy úgy

kellemetlen következményekkel járhat (pocsék filmet kell végignéznie, vagy a partnere lesz elégedetlen, vagy mindkettő). Visszatérve az incselkedésre, a beszélő „érzelmi hullámvasútra" ülteti a hallgatót, ahogy azt a 3.2.3. pontban taglaltuk.

A két zárótípus (3.2.4) társítása nem szorul magyarázatra.

A tanulmányban bemutatott elemzés tehát Dér (2017) korpuszbeli vizsgálatának kísérletes rekonstrukciójaként is felfogható. Ebben a felfogásban azt mondhatjuk, hogy új megfigyelésekkel szolgál a *hát*-típusok prozódiai megvalósulásának relevanciájára vonatkozóan, és kijelöli az inferenciális pragmatikai struktúra szükségszerű finomításának útjait.

4.4. Párbeszédek felidézése (hát hogy...)

Dér (2022: 29) érdekes adalékkal szolgál a párbeszédes kísérleti sémánkra vonatkozóan. Azt mutatja be ebben a tanulmányában, hogy „a *hogy* önállóan is képes lenne a beszédpartner megnyilatkozását felvezetni korábbi dialógusok felidézése során (11), de együttes használat[a a *hát*-tal: *hát hogy...*] még egyértelműbbé teszi, hogy beszélőváltás történt [ama korábbi dialógusban]." „A korábbi dialógusok felidézése valós vagy fiktív párbeszédekben tehát a legjellemzőbb a *hát + hogy* együttes használatára."

(10) *Számvevőszék már megkeresett, tett föl kérdéseket.* ***Hát hogy hány ellenjegyzője van egy számlának, meg ilyesmi.*** (MNSz2, doc#2618, beszélt nyelvi)

Azt a sejtést szeretnénk kimondani a jövőbeli kutatásra hagyva az igazolást, hogy a határozott *hát*-típusok közé tartozik a *hát* ezen használata; mivel a beszélő arra utal, hogy szinte kitalálható, hogy milyen kérdésekről esett szó a *hát* előtti mondatban.

4.5. A követőszünet kitöltése (hááát öööö...)

Mint a 4.3. pontban céloztunk rá, az aggodalmas (h3) *hát*-típusra – és éppen erre a típusra – jellemző volt, hogy a kísérleti alanyok hosszú köve-

tőszünetet alkalmaztak, és sokan így mondták ki, részben kitöltve a szünetet: *hááát öö*. A felolvasandó szövegben semmilyen módon nem utaltunk arra, hogy a *hát* hangsort hogyan mondják ki. Ez egyértelmű kísérletes megerősítése Németh (2020: 44) korpuszbeli vizsgálatainak, aki kimondja: „a [hát + öö] mintázatot a beszélők akkor használhatják, amikor választ kezdenek egy nekik feltett kérdésre (*hát*), de nem tudják a választ azonnal produkálni (*öö*). A beszélők tekintetviselkedésének vizsgálata azt sugallja, hogy a *hát*-hoz csatlakozó *öö* halasztó funkciót tölt be: produkciójának idejére a szókeresés eseteihez hasonlóan a beszélők elkerülik a szemkontaktust partnereikkel."

Így fogalmaz Németh (2020: 31, 39) a számunkra releváns kérdésben: „az angol *uh(m)*-hoz hasonlóan az *öö* esetében is előfordul a magyar korpuszban olyan példa, amikor a beszélő egy kínos / tabu kifejezés produkciója előtt használja az *öö*-t, vagy akkor, amikor negatív kritikát készül közölni. Ekkor az *öö* lehetséges funkcióját érdemes percepciós oldalról megközelíteni. Lerner (2013) szerint az ilyen fordulókonstrukciós hezitációs technikák kifejezhetik a beszélőnek a soron következő kifejezéshez való viszonyát a hallgató(k) felé, miszerint kínosnak tartja az adott szót vagy témát", illetve „Lerner (2013) megmutatja, hogy a forduló megalkotása során alkalmazott halasztó, hezitációs jelenségek alkalmasak annak kifejezésére, hogy a beszélő a közeledő beszédrész tartalmát érzékenynek, kínosnak tartja."

5. Összegzés

Ebben a munkában elsődlegesen Szeteli et al. (2022) tanulmányára építve tíz *hát*-típust mutattunk be szórendi és prozódiai jellemzéssel, amely a releváns ejtési hosszadatok statisztikai értékelésén alapult. Pontosabban kilenc *hát*-típust különítettünk el; egy esetben pedig úgy foglaltunk állást, hogy a vizsgált funkció, az incselkedés valójában nem a *hát* funkciója (3.2.3. szakasz). A mondatéli *hát*-típusokra kimondtunk egy olyan általánosítást, hogy „szemaforként" működnek: a hallgató az első szó hallatán

felkészülhet, hogy könnyen vagy nehezebben „emészthető” információt kap (3.2.1, 3.2.2). Azzal a meggyőződéssel bocsátjuk útjára e tanulmányt, hogy kísérletes megközelítésünk hozadékai a korpuszbeli mintavételen alapuló vizsgálatokhoz képest számot tarthatnak a *Beszédtudomány* olvasóinak érdeklődésére, miután a releváns eredményeket és megfigyeléseket kiemeltük az eredeti tudományos kontextusból és összevetettük a korpusz alapú megközelítésekkel (Dér, 2010, 2012, 2017, 2022; Németh, 2020) – amire önálló szakaszt szenteltünk (44. szakasz) a bevezetést (1. szakasz), valamint a kísérleteink módszertanának (2. szakasz) és eredményeinek (3. szakasz) a bemutatását követően.

Eredményeink hasznosítására több területen is látunk lehetőséget. Ismeretesek például olyan beszédtechnológiai fejlesztések, amelyek célja az érzelmek automatikus felismerésére képes algoritmusok kidolgozása (pl. Vicsi & Sztahó, 2009). Amennyiben egy ilyen algoritmus fel tudja ismerni az elhangzó *hát*-okat és azokat prozódiai sajátosságaik alapján megfelelő *hát*-kategóriákba tudja sorolni, az algoritmus felhasználója máris pontosabb képet kaphat a beszélő beszéd közbeni érzelmi állapotáról, attitűdjeiről. További alkalmazási lehetőséget az oktatás jelenthet. A magyart idegen nyelvként elsajátítani szándékozó tanulók számára lejátszhatók anyanyelvi beszélők *hát*-okat tartalmazó felvételei, így vizsgálható, hogy képesek-e „visszakódolni” a *hát*-ok funkcióit, ezáltal pontosabban megérteni az anyanyelvi beszélő szándékát, állapotát. Egyúttal ilyen gyakorlatokkal a *hát* megértésének képessége is fejleszthető. Mindezek mellett azonban globális célunk – csatlakozva a Schirm és Dér idézett munkáiban testet öltő törekvéshez – eloszlatni azt a tévhitet, hogy a *hát* fölösleges, így *hát* kerülendő töltelékszó. Ma már világosan látjuk, hogy a *hát* az élő beszéd természetes eleme, és önmagában vagy mondatban különböző pozíciókban sajátos módon jelentésárnyalatokat tud kifejezni, sőt, akár értelmezési keretet nyújt a beszéd hallgatója számára, amikor egy-egy közlést értelmez.

Hivatkozások

Alberti, G. (2016). Hát a meg meg a hát. In B. Kas (Ed.), *„Szavadd ne feledd!" Tanulmányok Bánréti Zoltán tiszteletére* (pp. 17–27). Budapest: MTA Nyelvtudományi Intézet.

Alberti, G. (2020). A világ nyelvei, a nyelv világai. In G. Böhm, D. Czeferner, & T. Fedeles (Eds.), *Bölcsész Akadémia 4* (pp. 126–150). Pécs: PTE BTK KTDT.

Alberti, G., Dóla, M., Kárpáti, E., Kleiber, J., Szeteli, A., & Viszket, A. (2019). Towards a cognitively viable linguistic representation. *Argumentum*, *15*, 62–80.

Alberti, G., Dóla, M., Kárpáti, E., Kleiber, J., Viszket, A., & Szeteli, A. (2021). Lehetséges lehetséges világaink. *Jelentés és Nyelvhasználat*, *8*, 105–145. URL: `http://www.jeny.szte.hu/jeny-2021-albertig-etal`.

Alberti, G., & Laczkó, T. (2018). *Syntax of Hungarian: Nouns and noun phrases. In series: Comprehensive grammar resources: Hungarian*. Amsterdam: University Press.

Alberti, G., & Szeteli, A. (2018). Realis: a mi kölcsönös és élethossziglani tudásunk. *Iskolakultúra*, *28*, 3–14. URL: `https://www.iskolakultura.hu/index.php/iskolakultura/article/view/31580/31267`.

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge: University Press.

Austin, J. (1962/1975). *How to Do Things with Words*. Oxford: Clarendon Press.

Bell, D. (1998). Cancellative discourse markers: A core/periphery approach. *Pragmatics*, *8*, 515–541.

Boersma, P., & Weenik, D. (2017). Praat: Doing phonetics by computer. URL: `http://www.praat.org`.

Dér, C. I. (2010). „töltelékelem' vagy új nyelvi változó? a hát, úgyhogy, így és ilyen újabb funkciójáról a spontán beszédben. *Beszédkutatás*, *18*, 159–170. URL: http://real.mtak.hu/26142/1/Beszedkutatas_2010.pdf.

Dér, C. I. (2012). Beszélőváltások során használt diskurzusjelölők a magyar spontán beszédben. *Beszédkutatás*, *20*, 130–141. URL: http://real.mtak.hu/26150/1/Beszedkutatas_2012.pdf.

Dér, C. I. (2017). A *hát* multifunkcionalitása a beszédműfajok és a diskurzusjelölő-társulások függvényében. *Beszédkutatás*, *25*, 169–184. URL: https://ojs3.mtak.hu/index.php/beszkut/article/view/387/196.

Dér, C. I. (2022). Háthogy esetleg netalántán azt gondolnátok rólam... a *szóval* és a *hát* diskurzusjelölővel álló hogy kötőszós mellékmondatokról. *Jelentés és Nyelvhasználat*, *9*, 43–57. doi:http://www.jeny.szte.hu/jeny-2022-dercsi.

Dér, C. I., & Markó, A. (2007). A magyar diskurzusjelölők szupraszegmentális jelöltsége. In T. Gecső, & C. Sárdi (Eds.), *Nyelvelmélet* (pp. 61–67). Székesfehérvár–Budapest: Kodolányi János Főiskola – Tinta Kiadó.

Dér, C. I., & Markó, A. (2017). A hát funkciói a prozódiai megvalósulás függvényében. *Beszédkutatás*, *24*, 105–117. URL: https://ojs.mtak.hu/index.php/beszkut/article/view/385/197.

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* (pp. 1–20). Amsterdam: Elsevier.

González, M. (2004). *Pragmatic markers in oral narrative*. Amsterdam/Philadelphia: John Benjamins.

Gósy, M. (2012). Multifunkcionális beszélt nyelvi adatbázis – bea. *Általános Nyelvészeti Tanulmányok*, *24*, 329–349.

Kamp, H., Genabith, J., & Reyle, U. (2011). Discourse representation theory. *Handbook of Philosophical Logic*, *15*.

Kiefer, F. (1988). Modal particles as discourse markers in questions. *Acta Linguistica Hungaricam*, *38*, 107–125.

Kocsis, E. (2021). *Diskurzusjelölők az internetes nyelvhasználatban. Kézirat*. Pécs: PTE BTK Nyelvtudományi Tanszék.

Kondacs, F. (2016). A hát diskurzusjelölőről az óvodások diskurzusaiban. In T. Váradi (Ed.), *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2016* (pp. 45–58). Budapest: MTA Nyelvtudományi Intézet.

Kugler, N. (2012). *Az evidencialitás jelölői a magyarban, különös tekintettel az inferenciális evidenciatípusra*. Budapest: ELTE BTK Mai Magyar Nyelvi Tanszék.

König, E., & Siemund, P. (2007). Speech act distinctions in grammar. In T. Shopen (Ed.), *Language Typology and Syntactic Description* (pp. 276–324). Cambridge: Cambridge University Press volume 1.

Langacker, R., Arrese, J., Haßler, G., & Carretero, M. (2017). Evidentiality in cognitive grammar. In J. Marín-Arrese, G. Haßler, & M. Carretero (Eds.), *Evidentiality Revisited: Cognitive Grammar, Functional and Discourse-Pragmatic Perspectives* (pp. 13–56). Amsterdam: John Benjamins.

Lerner, G. (2013). On the place of hesitating in delicate formulations: A turn-constructional infrastructure for collaborative indiscretion. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational Repair and Human Understanding* (pp. 95–134). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511757464.004.

Loftus, G., & Masson, M. (2014). Using confidence intervals in within-subject design. *Psychonomic Bulletin & Review*, *1*, 476–490.

Markó, A. (2013). *Az irreguláris zönge funkciói a magyar beszédben*. Budapest: ELTE Eötvös Kiadó. URL: https://www.eltereader.hu/media/2014/04/Marko_Az_irregularis_READER.pdf.

Markó, A., & Dér, C. I. (2011). A diskurzusjelölők használatának életkori sajátosságai. In J. Navracsics, & Z. Lengyel (Eds.), *Lexikai folyamatok egy- és kétnyelvű közegben. Pszicholingvisztikai tanulmányok II. (Segédkönyvek a nyelvészet tanulmányozásához 121* (pp. 49–61). Budapest: Tinta Könyvkiadó.

McGlone, R. (1967). Air flow during vocal fry phonation. *Journal of Speech, Language and Hearing Research*, *10*, 299–304.

Mondada, L. (2013). The conversation analytic approach to data collection. In J. Sidnell, & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 32–56). Oxford: Wiley-Blackwell. doi:10.1002/9781118325001.ch3.

Nuyts, J. (2017). Evidentiality reconsidered. In J. Arrese, G. Haßler, & M. Carretero (Eds.), *Evidentiality Revisited: Cognitive Grammar, Functional and Discourse-Pragmatic Perspectives* (pp. 57–83). Amsterdam: John Benjamins.

Németh, T. E. (1998). A hát, így, tehát, mert kötőszók pragmatikai funkciójának vizsgálata. *Magyar Nyelv*, *94*, 324–331.

Németh, Z. (2020). A nemlexikális öö hang interakciós szerepének elemzése magyar nyelvű társalgásokban. *Jelentés és Nyelvhasználat*, *7*, 23–50.

Redeker, G. (1991). Linguistic markers of discourse structure: review of discourse markers, by deborah schiffrin. *Linguistics*, *29*, 1139–1172.

Sadock, J., & Zwicky, A. (1985). Speech act distinctions in syntax. In T. Shopen (Ed.), *Language Typology and Syntactic Description* (pp. 155–196). Cambridge: Cambridge University Press volume I.

Schegloff, E. (1996). Turn organization: one intersection of grammar and interaction. In E. Ochs, E. Schegloff, & S. Thompson (Eds.), *Interaction and Grammar* (pp. 52–133). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620874.002.

Schirm, A. (2008). A hát diskurzusjelölő története. nyelvtudomány iii-iv. *Acta Universitatis Szegediensis Sectio Linguistica*, (pp. 185–201).

Schirm, A. (2011a). *A diskurzusjelölők funkciói: A hát, az -e és a vajon elemek története és jelenkori szinkrón státusa alapján. PhD-értekezés*. Szeged: Szegedi Tudományegyetem, Bölcsészettudományi Kar.

Schirm, A. (2011b). A diskurzusjelölők funkciói a számok tükrében. *Alkalmazott Nyelvészeti Közlemények*, *6*, 185–197.

Schirm, A. (2015). A diskurzusjelölők az osztálytermi kommunikáció szövegtípusaiban. In K. Baditzné Pálvölgyi, E. Szabó, & R. Szentgyörgyi (Eds.), *Tanóratervezés és tanórakutatás: A magyar nyelv és irodalom, az idegen nyelvek és a művészetek műveltségi területen* (pp. 49–66). Budapest: ELTE.

Schirm, A. (2017). A diskurzusjelölők és a szövegtípusok viszonyáról. *Magyar Nyelv*, *113*, 330–341.

Schirm, A. (2021). *Diskurzusjelölők szövegeken innen és túl*. Budapest: Loisir.

Searle, J. (1979). *Expression and meaning – Studies in the Theory of Speech Acts*. Cambridge: University Press.

Szeteli, A. (2019). Towards describing the extremely multifunctional hungarian discourse marker hát. In J. Emonds, M. Janebová, & L. Veselovská (Eds.), *Language Use and Linguistic Structure. Proceedings of the Olomouc Linguistics Colloquium 2018* (pp. 355–372). Olomuc: Palacký University.

Szeteli, A., & Alberti, G. (2018). Hát igen, más hát! In D. Czeferner, G. Böhm, & T. Fedeles (Eds.), *Mesterek és Tanítványok 2: Tanulmányok a bölcsészet- és társadalomtudományok területéről*. Pécs: PTE BTK Kari Tudományos Diákköri Tanács.

Szeteli, A., Dóla, M., & Alberti, G. (2019a). Pragmasemantic analysis of the hungarian inferential – evidential expression 'szerint'. *Studies in Polish Linguistics*, *14*, 207–225.

Szeteli, A., Gocsál, A., & Alberti, G. (2019b). Szemafor hát! *Jelentés és Nyelvhasználat*, *6*, 33–63.

Szeteli, A., Gocsál, A., Szente, G., & Alberti, G. (2020). A hát diskurzusjelölő prozódiai megvalósulásának vizsgálata felolvasásokban. In A. Gocsál, M. Gósy, T. Gráczi, D. Gyarmathy, V. Horváth, A. Huszár, A. Kohári, V. Krepsz, & K. Mády (Eds.), *Speech Research Conference*. Budapest: Hungarian Research Institute for Linguistics. URL: http://real.mtak.hu/118352/1/beszkut_speechresearch_2020_proceedings.pdf.

Szeteli, A., Gocsál, A., Szente, G., & Alberti, G. (2022). Differentiation of segmentally identical expressions occurring in the same or different sentence zones in hungarian by duration, pitch, intensity and irregular voicing. *Acta Linguistica Academica*, *69*, 163–187.

Varga, L. (2016). The intonation of topic and comment in the hungarian declarative sentence. *Finno-Ugric Languages and Linguistics*, *5*, 46–77.

Vicsi, K., & Sztahó, D. (2009). Ügyfél érzelmi állapotának detektálása telefonos ügyfélszolgálati dialógusban. In *VI. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 217–225). Szeged. URL: https://acta.bibl.u-szeged.hu/58711/1/msznykonf_006_217-225.pdf.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

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

Függelék

Kérjük, hogy először is gondosan olvasd el a következő tesztek értelmezéséhez az alábbi szöveget!

Azt fogjuk kérni, hogy A és B alábbi párbeszédében olvasd fel, pontosabban játszd el B szerepét! Azért említjük az eljátszást, mert fontos a kísérlet szempontjából, hogy ne monoton felolvasási hangsúlyt alkalmazz, hanem éld át a felvázolt szituációt a megadott háttértudás alapján, és úgy add elő a szóban forgó pár mondatot. Arra kérünk, hogy minden leírt szót a megadott szórend szerint mondj el, mert nagyon fontosak számunkra a finom részletek. Az írásjeleket viszont nem tettük ki B szövegében, mivel az a célunk, hogy mindent úgy hangsúlyozz, ahogyan azt természetesnek érzed. A mondatrészeket emiatt külön sorokba szedtük, hogy az értelmezést megkönnyítsük számodra.

Az eljátszás előtt gondosan olvasd át mind a dialógust, mind a háttérbeli gondolatokat! A és B húsz év körüli egyetemista szerelmespár. Egy éve minden hétvégén elmennek moziba, és felváltva választanak filmet, így meglehetősen jól ismerik már egymás ízlését ezen a téren. Ezen a héten B választhat. A különböző szituációk abban a tekintetben nem függetlenek egymástól, hogy a szereplők ízlését végig egységesen kezeljük, azonban a mozi által felkínált filmek leírása alapján a különböző helyzetekben B más-más módon dönt. Abban a pillanatban halljuk őket, amikor B már végignézte a mozi kínálatát és meghozta a döntését, de még nem mondta el A-nak, aki kíváncsian várja az „eredményhirdetést". Három film verseng:

- egy klasszikusnak ígérkező angol krimi,
- egy malackodó poénokat sejtető amerikai vígjáték,
- egy nyomasztónak tűnő izlandi dráma.

Tudni kell még, hogy mind A, mind B hétről hétre azon a kézenfekvő módon hoz döntést, hogy némileg szeretné kihasználni a választási esélyt, de semmiképpen nem akar rosszat partnerének sem.

Distinguishing between dysarthria types based on acoustic parameters

Bernadett Dam[1], Lívia Ivaskó[2]

[1] *University of Szeged, Doctoral School in Linguistics*
[2] *University of Szeged, Department of General Linguistics*

Abstract

Dysarthria is a motor speech disorder resulting from neurological impairments. Because of the variability of impairments and disordered speech characteristics, it is useful to categorize it into types. The current study gives an overview of the main types of dysarthria, describing the different underlying causes, some disordered speech characteristics arising from those impairments, as well as the corresponding acoustic parameters, and some possible methods to measure the most relevant acoustic features. Six main groups of acoustic parameters were identified that could help distinguish between the types of dysarthria. Since the properties of the acoustic signal are connected to the manner of articulation, which is dependent on the neuromuscular system, the precise description of acoustic features of dysarthric speech could provide valuable information that could aid localization and differential diagnosis.

1. Introduction

"Motor speech disorders can be defined as speech disorders resulting from neurologic impairments affecting the planning, programming, control, or execution of speech" (Duffy, 2013). Dysarthria is a collective name for a group of motor speech disorders that reflect abnormalities in the movements required for speech production. Depending on the localization and severity of the impairment, neuromuscular deficits may affect any or all of the respiratory, phonatory, resonatory, and articulatory components of speech, or they may affect a single component only (Ackermann et al., 2010). Due to the diversity of the possible underlying deficits, perceived speech abnormalities are heterogeneous, so in order to describe, understand and manage dysarthria successfully, it is

Email addresses: dam.bernadett@stud.u-szeged.hu (Bernadett Dam), ivasko@hung.u-szeged.hu (Lívia Ivaskó)

Beszédtudomány – Speech Science 2024/1. 118–135. DOI: 10.15775/Besztud.2024.1.118-135

helpful to categorize it into types. The most widely used categorization was first established by Darley et al. (1969a; 1969b), who delineated five types of dysarthria: flaccid, spastic, ataxic, hypokinetic, and hyperkinetic, as well as a sixth, mixed type. Their categorization was based on 38 dimensions or speech features, using perceptual methods. This grouping based on the combination of functional speech deficits was adopted by Duffy (2013) who added a new type, called unilateral upper motor neuron (UUMN) dysarthria. An advantage of this categorization is that the described speech features (e.g., hypernasality) can be directly tied to neuroanatomical deficits, so the precise description of speech characteristics can provide information about the localization of impairment. Successful categorization of the dysarthrias can therefore have implications for the localization and diagnosis of the underlying neurological disorder and can aid the clinical management process (Duffy, 2013).

Dysarthria assessment can be done in many ways (using perceptual, acoustic or physiological methods), the most widely used method is the perceptual, as it has several advantages. First, it costs significantly less than instrumental methods; second, a speaker's intelligibility can be easily assessed perceptually, third, different dysarthria types can be distinguished with a high success rate. However, there are some disadvantages to this method, such as its high subjectivity, being difficult to standardize, and providing limited information about the pathophysiological background of the perceived speech characteristics (Cummings, 2008).

In contrast, acoustic methods can provide objective, quantifiable measurements that may confirm perceptual judgements on one hand or highlight aspects of dysarthric speech that could not be measured perceptually on the other. Acoustic analyses of the different types of dysarthria have generally two main goals: first, identifying the exact aspects of the acoustic signal related to intelligibility deficits in dysarthria, and second, providing a precise description of the different types' acoustic profiles (Kim et al., 2011). Despite these advantages, there are some drawbacks to acoustic analyses as well. One such disadvantage is the possible difference between the most salient speech characteristics iden-

tified using perceptual and acoustic methods, for example, loudness or pitch abnormalities measured instrumentally may not be perceived by listeners (Kent et al., 1999).

This paper aims to give a general description of the acoustic profiles of the different types of dysarthria with special attention to the most relevant acoustic features in terms of distinguishing between the types. Furthermore, we attempt to give an overview of the neurological impairments underlying speech deficits in dysarthria, and how these speech abnormalities can be described with acoustic analysis. The paper is structured as follows. The next section introduces the six main types of dysarthria, from the underlying impairments to the speech characteristics. Section 3 describes the main methods of acoustic analysis that could be used to differentiate the types of dysarthria, and summarizes the methods and results of some recent empirical studies examining the acoustic differences between the types. Finally, Section 4 gives a brief summary and mentions possible future work.

2. The types of dysarthria

2.1. Flaccid dysarthria

The main distinguishing speech characteristics of flaccid dysarthria are due to muscle weakness and reduced muscle tone. Speech abnormalities can be present in any or all of the components of speech. The condition is the result of the impairment of one or more cranial or spinal nerves caused most commonly by trauma. Other possible causes include congenital, infectious or inflammatory, degenerative, and vascular diseases. The affected muscle groups depend on the lesion loci, sometimes involving only a single muscle group, which can aid the localization. The most noticeable speech deviations in this type are caused by vagus nerve (cranial nerve X) lesions, which supplies most of the muscles of the pharynx, the soft palate, and the larynx. Vagus nerve lesions can cause weakness in the soft palate, diminishment of the gag reflex, and nasal regurgitation among others. These changes can manifest in speech as aphonia, reduced loud-

ness, reduced pitch, hypernasality, nasal emission, hoarseness, stridor (audible inhalation), and diplophonia (double pitch phonation). Other speech features of this type include short phrases, monotonous pitch and loudness, and imprecise consonants. The latter cannot be tied to the lesion of a single cranial nerve (Duffy, 2013).

2.2. Spastic dysarthria

The hallmark symptom of this type is the combination of weakness and spasticity, caused by bilateral damage to the direct and indirect activation pathways of the central nervous system. The indirect activation pathways are responsible for reflexes, maintaining posture, regulating muscle tone, and give a framework for skilled movements. Their activation can have an inhibitory role. Damage to these pathways mostly affects their inhibitory role, the results are overactivity, manifesting as increased muscle tone, spasticity, and hyperactive reflexes. Direct activation pathways serve a facilitatory role, they are related to skilled, fine movements. Their damage causes loss or impairment of said fine movements. Underlying conditions causing damage to the activation pathways are generally vascular (e.g., stroke), degenerative or traumatic. As opposed to flaccid dysarthria, where individual muscle groups are affected, flaccid dysarthria can be characterized as the impairment of movement patterns, as the affected areas are tied to motor control. As a result, deficits arise in all components of speech (Duffy, 2013). Darley et al. (1969b) grouped the most notable disordered speech characteristics of spastic dysarthria into four clusters. The first cluster, prosodic excess includes slow rate and excess and equal stress. The second cluster is called articulatory-resonatory incompetence, and it consists of imprecise consonants, distorted vowels, and hypernasality. The third cluster, prosodic insufficiency, includes features such as monopitch, monoloudness, reduced stress, and short phrases. Finally, the fourth cluster, phonatory stenosis, covers low pitch, harshness, strained-strangled voice, pitch breaks, short phrases, and slow rate. Although imprecise consonants are the most salient feature of spastic

dysarthria, they can be found in all main types of dysarthria and can not be used as a distinguishing speech characteristic of this type (Duffy, 2013).

2.3. Ataxic dysarthria

Ataxic dysarthria is characterized by incoordination resulting from damage to the cerebellar control circuit. Speech abnormalities can affect all levels of speech, but are most notable in articulation and prosody (Duffy, 2013). The cerebellum influences the motor system in multiple ways, for example it plays a role in the timing of movement components, regulating the scale of movements and muscle contractions for fine movements (Laforce & Doyon, 2001). Damage to the area is caused most often by degenerative disease, but demyelinating, vascular, traumatic or toxic diseases are not uncommon either. Failure to coordinate or control movement patterns have an effect on speech too, that is why a distinguishing characteristic of ataxic dysarthria is the irregularity of alternating motion rates (AMRs, that is, the repetition of one syllable as steadily as possible). Darley et al. (1969a) identified three clusters of disordered speech characteristics in this type. The first cluster is called articulatory inaccuracy, it can be characterized by imprecise consonants, irregular articulatory breakdowns, and vowel distortions. The second cluster, prosodic excess, includes excess and equal stress, prolonged phonemes, prolonged intervals, and slow rate. Lastly, the third cluster is phonatory-prosodic insufficiency, and it consists of harshness, monopitch, and monoloudness.

2.4. Hypokinetic dysarthria

The most prominent characteristics of hypokinetic dysarthria are rigidity, reduced force and range of movement, and slow individual but fast repetitive movements, which can affect any or all levels of speech. It is caused by damage to the basal ganglia control circuit, and is most often, but not always associated with Parkinson's disease (PD). The functions of the basal ganglia control circuits include regulating muscle tone, stabilizing posture during fine

movements, regulating movements supporting goal-oriented activities, regulating force, amplitude and duration of movements, and adjusting movements to the environment. Damage to these circuits can lead to reduction of movement or the inability to inhibit involuntary movement. Resulting speech abnormalities include weak voice, hoarseness or breathiness, fast rate, syllable repetition, rapid and blurred AMRs (Duffy, 2013). Darley et al. (1969b) named only one cluster of speech abnormalities associated with this type. The cluster of disordered speech characteristics is called prosodic insufficiency and is characterized by monopitch, monoloudness, reduced stress, short phrases, variable rate, short rushes of speech, and imprecise consonants.

2.5. Hyperkinetic dysarthria

Speech abnormalities in hyperkinetic dysarthria are due to rhythmic or irregular, slow or fast involuntary movements. It is also caused by damage to the basal ganglia control circuits, resulting in deviations in any or all components of speech, which are most notable in prosody and rate. Hyperkinetic speech can give the impression that speech production starts out normally, but is distorted, slowed or interrupted by involuntary movements. As mentioned above, lesions of the basal ganglia control circuits can lead to the failure of inhibition of involuntary movements, as well as voluntary movements being slowed down. The groups of caused involuntary movements are heterogeneous, e.g., dyskinesia (a broad category of abnormal involuntary movements), myoclonus (quick contraction of muscle groups), tics (quick, stereotypical, patterned movements), chorea (quick, irregular, random movements), tremor (rhythmic movement of a body part), and dystonia (excessive contraction of muscles). Resulting speech abnormalities depend on the type of involuntary movements, and therefore they can be diverse as well. To name a few, hyperkinetic speech characteristics include prolonged intervals, strained voice quality, hypernasality, tremor, and slow and irregular AMRs (Duffy, 2013).

2.6. Unilateral upper motor neuron (UUMN) dysarthria

This type shows effects of weakness, sometimes spasticity and incoordination. Disordered speech characteristics can manifest in any or all levels of speech, most often notable in articulation, phonation, and prosody. In contrast to all other types, this type is characterized based on anatomy, the underlying cause is always damage to the upper motor pathways. This type has received limited attention, since its symptoms can be mild, recovery can be quick, and as a result, it is difficult to conduct research on it (Ackermann et al., 2010). The upper motor neuron system is bilateral, its pathways pass signals to cranial and spinal nerves which are related to muscles that play a role in speech production. Several nerves (such as the trigeminal or the vagus) receive both contralateral and ipsilateral innervation, which allows to preserve breathing, feeding, and speech functions in the case of unilateral lesions. In some cases, however, unilateral damage can result in unilateral facial weakness, weakness of the jaw, palate, vocal fold, and most noticeably the tongue. The most common possible etiology is stroke, but tumor and trauma are also frequent (Duffy, 2013). The most apparent speech deficits in this type are imprecise consonants, irregular articulatory breakdowns, and irregular, slow or imprecise AMRs. Phonatory abnormalities, such as hoarseness and decreased loudness are also described (Duffy & Folger, 1996).

3. Acoustic features relevant to distinguishing between dysarthria types

This section aims to describe how the most relevant distinguishing features among the types of dysarthria can be measured with acoustic analysis. Drawing conclusions from the essential literature (Darley et al., 1969a,b; Duffy, 2013) and the results of some recent empirical studies we can name six main clusters of disordered speech characteristics based on the acoustic parameters that are involved. These clusters and some specific characteristics covered by them are the following. (1) Temporal characteristics (slow or fast rate, prolonged

intervals, silences), (2) Changes in pitch (monopitch, pitch break, stress irregularities), (3) Changes in intensity (reduced loudness, monoloudness, loudness variability), (4) Changes in articulation (imprecise vowels and/or consonants), (5) Nasal resonance (hypernasality, nasal emission), (6) Changes in voice quality (harshness and breathiness). This categorization partially follows the three main acoustic domains (frequency, duration, and intensity), and can partially be described as a combination of them. The descriptions of these characteristics are followed by brief overviews of some recent empirical studies which relied on said features. For the sake of brevity, we only mention studies where the discussed acoustic features ranked as the most important or highly relevant features when distinguishing dysarthria types.

3.1. Temporal characteristics

Measuring speech properties pertaining to the time domain seems to be the most straightforward. The necessary procedures include segmenting the appropriate speech units (e.g., phonemes or syllables) with care based on the waveform and spectrogram, the duration of these intervals can be measured automatically using an acoustic analysis software (e.g. Praat, Boersma & Weenink, 2023). Automatic segmentation of dysarthric speech may have limitations, and therefore it is advised to manually correct the outcome. It could be fruitful to carry out intraspeaker, as well as interspeaker comparisons between different phoneme durations, as the duration of different phonemes can be affected diversely. Examining differences in such durations can aid identify the factors underlying intelligibility deficits, and the localization of neuroanatomical impairment (Kent et al., 1999). Another frequently measured property is the duration of syllables, which should be compared to the duration of silences to obtain the speech rate and the possible irregularities thereof. The preferred method for this is called the diadochokinesis (DDK) test, which is a method used to detect irregularities in rapid alternating movements, testing speech motor ability (Juste et al., 2012).

Fougeron et al. (2022) aimed to differentiate between flaccid, hypokinetic, ataxic and two mixed (amyotrophic lateral sclerosis [ALS] and Wilson-syndrome)

dysarthric French speech using a complex feature set. The differentiation was based on seven dimensions describing intelligibility, articulation, maximum phonation time, voice, prosodic contrast, speech rate, and diadochokinetic rate. Out of all features, DDK rate proved to be the most informative one, most successfully distinguishing the ataxic and one mixed (ALS) group from the rest. Kim et al. (2011) classified English speakers based on eight acoustic features. The study represented all types identified by Duffy (2013). They found that one of the two main contributors to type classification was articulation rate. Lowit and Kuschmann (2012) focused on intonation combined with temporal measures in hypokinetic and ataxic English speakers. Among others, their results showed faster speech rate for the hypokinetic group compared to the ataxic group. Liss et al. (2009) focused on differences in speech rhythm between four English speaking dysarthric groups (ataxic, hypokinetic, hyperkinetic, and mixed flaccid-spastic). The recordings were analyzed along the lines of eleven parameters, such as articulation rate or the standard deviation of vocalic intervals. The results revealed that hypokinetic speakers had normal or fastened speech rate, the mixed group showed slow and prolonged speech, and in the case of hyperkinetic dysarthria, vocalic intervals showed high variability. As for ataxic speech, rather than finding one or two prominent parameters, it is the combination of features that leads to successful differentiation. Nishio and Niimi (2001) examined speech rate and its components in Japanese speakers, comparing all dysarthria types described by Duffy (2013). The analyzed features were speaking rate, articulation rate, and speech/pause ratio. According to the results, the flaccid and hypokinetic group had similar articulation rate to the control group, however, these two groups had the highest speech/pause ratio. The slowest speaking rate was observed in the spastic and mixed groups, the slowest articulation rate belonged to the mixed group. Finally, Kis et al. (2020) analyzed the speech rate of Hungarian dysarthric speakers, grouping the subjects based on etiology (Parkinson's disease, stroke, and sclerosis multiplex). The results of diadochokinesis tests showed significant differences between all

three groups: the PD group's speech rate was the highest, while the stroke group's speech rate was the lowest in every task.

3.2. Changes in pitch

Several deviations in speech can be described with the examination of the pitch, which is the human perception of the fundamental frequency (F0). It influences whether a voice is perceived as high or low, and its alteration plays a role in suprasegmental features such as stress and intonation. Its deviations can be measured as follows. Monopitch, that is, reduced stress or intonation manifests as a flat F0 contour, irregularities in stress or intonation mean abnormal F0 patterns, perceived low pitch is related to a low frequency F0, and pitch break is a silent interval within the pitch contour. Voice tremor manifests in the rhythmic oscillation of F0, the frequency of these oscillations depends on the underlying condition causing the tremor. Due to the diversity of its alterations, F0 is analyzed along the lines of numerous parameters, such as statistical properties (mean, mode, standard deviation), F0 contour, jitter, and tremor (Kent et al., 1999; Ball & Lowry, 2001; Ball, 2021). When analyzing the fundamental frequency, it is necessary to keep in mind the demographic data of the speaker (age, sex), as these have a high influence on the pitch, consequently, comparison should be made only between members of the same demographic group.

F0 was one of the main contributors to type classification when comparing all dysarthria types in the study of Kim et al. (2011). Thoppil et al. (2017) analyzed vowel formants in three types of dysarthria: ataxic, spastic, and extrapyramidal (the latter could be hypokinetic, hyperkinetic or UUMN dysarthria) using speech samples of Malayalam speakers. The examined the values of the fundamental frequency, the first two formants, and pitch break. The authors report that F0 jitter and flat F0 were mostly found in the extrapyramidal group, and pitch break is most common among ataxic speakers. Lowit and Kuschmann (2012) conducted a variety of intonation measures, such as the mean length of intonation phrases or the syllable–pitch-accent ratio. They

found significant between-group differences, such as a higher number of rising pitch accents for ataxic speakers.

3.3. Changes in intensity

Intensity is proportional to amplitude and is related to perceived loudness. We can describe general characteristics of loudness with the attributes of intensity (highest, lowest or mean value), while its monotonicity or excess variability manifest as abnormalities in the intensity contour. Analyzing intensity could be important not only because it describes loudness but also because examining it along with temporal and pitch-related features gives us valuable information about prosody. Analyzing these three parameters is especially advantageous when studying dysarthric speech, since prosodic alterations are common, yet different in nature in the different types. It is worth noting that stress is produced differently in different languages by changing either one or a combination of the three parameters of duration, pitch and intensity (Gósy, 2004), so when analyzing dysarthric speech we should be aware of the stress patterns of the language spoken by the person. By looking at stress patterns we will also be able to examine speech rhythm, which is the pattern of alternating stressed and unstressed syllables (Kent & Read, 2002).

3.4. Changes in articulation

Measuring the quality of articulation is a complex task, as it depends on the type of analyzed phonemes (e.g., vowels or consonants), as well as on the analyzed attributes (e.g., the alteration of the manner or place of articulation). The most common measures for vowel analysis include the values of the first three formant (F1, F2, F3) frequencies, F1–F0 difference value, F2–F1 difference value, and formant frequency fluctuation (Kent et al., 1999). These values are especially informative in the case of dysarthria, because they are related to the horizontal and vertical movements of the tongue. By measuring the frequencies of vowel formants, we can describe centralization, vowel space reduction, and abnormal formant frequencies. Since consonants form a heterogenous

group, we should distinguish sonorants from obstruents, studies on the latter being more common in the case of dysarthria (Kent et al., 1999). From the point of view of dysarthria research, spectral moment analysis can be a useful approach, the value of the first spectral moment being the most informative (Kim et al., 2011). Another promising metric is the slope of F2 transition in consonant-vowel sequences, which can be tied to overall intelligibility (Kent et al., 1999). The precision of the articulation of stop consonants can be described partially by the acoustic energy present during the occlusive phase. The occlusive phase is normally almost perfectly silent, but in the case of some dysarthric speakers (especially those with Parkinson's disease) produce energy during this phase. This can happen in two ways: incomplete closure can cause turbulence noise (spirantization), and laryngeal dysfunction can cause voicing (Kent et al., 1999). Lansford and Liss (2014) focused on vowel acoustics, comparing four groups of English speakers (ataxic, hypokinetic, hyperkinetic, and a mixed flaccid-spastic). Measurements were made using the first two formant frequencies of ten different vowels. F2 slope metrics (average F2 slope and F2 slope of the most dynamic vowels) showed significant differences between the groups. The hypokinetic group had greater average F2 slopes than the ataxic or the mixed group, as did the hyperkinetic group compared to the mixed group. Fougeron et al. (2022) found that the quality of articulation was one of the most relevant features when differentiating between flaccid, hypokinetic, ataxic and two mixed (ALS and Wilson-syndrome) groups, as it distinguished the flaccid and the two mixed groups from the rest.

3.5. Nasal resonance

Nasal resonance abnormalities include hypernasality and nasal emission. Nasality is caused by the dysfunction of the velopharyngeal valve, resulting in unwanted resonance in the nasal cavity. Its presence can complicate acoustic analysis, as it has several complex effects on the acoustic signal. It can be best described with some combination of five characteristics affecting vowels: (1) increase in formant bandwidth, (2), decrease in overall energy of the vowel,

(3) presence of a low-frequency nasal formant (250–500 Hz for adult males), (4) slight increase of the F1 frequency and lowering of the F2 and F3 frequencies, (5) the presence of one or more antiformants (Kent & Read, 2002). Nasal emission is caused by airflow escaping through the nasal cavity, noise arising. It is most apparent during the production of voiceless consonants. It manifests acoustically as broadband noise (Rollins & Oren, 2020) and as quasiperiodic noise (Zajac, 2021).

Castillo-Guerra (2009) compared six types of dysarthria (flaccid, spastic, ataxic, hyperkinetic, hypokinetic, mixed; all English speakers) based on twelve acoustic dimensions. One of the features that proved to be the most useful for classification was hypernasality.

3.6. Changes in voice quality

Voice quality or phonation is tied to glottal function. Its impairment can have diverse effects on speech, which can make its analysis challenging (Kent et al., 1999). Here, we discuss two types of voice abnormality for the sake of brevity: harshness and breathiness. In the case of harshness, the intensity of the fundamental frequency is prominent compared to the harmonics. Additionally, cepstral peak prominence values are lower in harsh voice than in normal voice (Heman-Ackah et al., 2014). Breathy voice is caused by insufficient glottal closure, excess airflow escaping through it. It has complex effects on the acoustic signal: higher amplitude of the first harmonic, high frequency noise, higher proportion of high-frequency energy (Hillenbrand et al., 1994).

In Fougeron et al. (2022), the voice quality score is the main contributor in the classification of the mixed types (ALS and Wilson-syndrome), as well as in many of the two-class classifications, such as hypokinetic (in PD) vs. mixed (ALS) dysarthria. Castillo-Guerra (2009) found that breathiness was one of the most relevant features when comparing five types of dysarthria. Interestingly, the results assigned no relevance to harshness, which is one of the most important features in the traditional approach.

Table 1 summarizes the most relevant distinctive features of the six types of dysarthria along their main acoustic manifestations.

4. Summary and future work

The current paper intended to give an overview of the main types of dysarthria, including the different underlying causes, some disordered speech characteristics arising from those impairments, as well as the corresponding acoustic parameters, and some possible methods to measure the most relevant acoustic features. Since the properties of the acoustic signal are connected to the manner of articulation, which is dependent on the neuromuscular system, the precise description of acoustic features of dysarthric speech could provide valuable information that could aid localization and differential diagnosis.

Six main groups of acoustic parameters were identified that could help distinguish between the types of dysarthria. The acoustic profiles of the types are not based on individual features, but rather on patterns described by the combination of several features. That is why the simultaneous analysis of multiple dimensions is needed in order to describe the types of dysarthria and identify the combination of features most relevant in terms of distinguishment. We see that temporal characteristics are examined the most extensively, however, we argue that other characteristics deserve attention as well, as they might provide important clues to the description and classification of each type, as well as the localization of impairment. Future work is needed to test the validity of said groups of acoustic parameters and to find the most informative features.

Acknowledgements

Supported by the ÚNKP-23-3-SZTE-52 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.

Dysarthria type	Speech characteristics	Acoustic manifestation
Flaccid	breathy voice	high amplitude H1, high frequency noise
	nasal emission	broadband noise, quasiperiodic noise
	short phrases	short duration
	hypernasality	nasal formant, antiformant
	irregular AMR	variable syllable duration
Spastic	slow speech rate	longer phoneme and pause intervals
	pitch break	silent interval within the F0 contour
	slow and regular AMR	long syllable duration, low standard deviation between them
Ataxic	excess and equal stress	increased and monotonous F0 or increased intensity
	imprecise vowels	abnormal formant structures
	loudness variability	changes in intensity
	irregular AMR	variable syllable durations
Hypokinetic	monopitch	flat F0 contour
	monoloudness	monotonous intensity
	reduced loudness and stress	low intensity and F0
	fast speech rate	short intervals
	unnecessary pauses	silence
Hyperkinetic	prolonged intervals	intervals longer than normal
	breathy voice	high amplitude H1, high frequency noise
	hypernasality	nasal formant, antiformant
	tremor	F0 tremor
	slow and irregular AMR	long syllable duration, high standard deviation between them
UUMN	slow speech rate	longer phoneme and pause intervals
	imprecise articulation	varied (e.g. abnormal formant structures, irregular F2 slopes)
	harsh voice	low CPP
	reduced loudness	low intensity

Table 1: Distinguishing speech characteristics of the main types of dysarthria and their acoustic manifestations (H1: first harmonic, AMR: alternating motion rate, F0: fundamental frequency, UUMN: unilateral upper motor neuron, F2: second formant, CPP: cepstral peak prominence).

References

Ackermann, H., Hertrich, I., & Ziegler, W. (2010). Dysarthria. In J. S. Damico, N. Müller, & M. J. Ball (Eds.), *The Handbook of Language and Speech Disorders* (pp. 362–390). Chichester, West Sussex: Wiley-Blackwell.

Ball, M. (Ed.) (2021). *Manual of Clinical Phonetics*. London: Routledge.

Ball, M., & Lowry, O. (2001). *Methods in Clinical Phonetics*. London: Whurr Publishers.

Boersma, P., & Weenink, D. (2023). Praat: doing phonetics by computer [computer program. *Version*, *6*. URL: `http://www.praat.org/`. Retrieved 3 June 2023.

Castillo-Guerra, E. (2009). Acoustic study of dysarthria. *International Journal of Biomedical Engineering and Technology*, *2*, 352–369.

Cummings, L. (2008). *Clinical Linguistics*. Edinburgh: Edinburgh University Press.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research*, *12*, 462–496.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, *12*, 246–69.

Duffy, J. R. (2013). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. St. Louis, MO: Elsevier.

Duffy, J. R., & Folger, W. N. (1996). Dysarthria associated with unilateral central nervous system lesions: A retrospective study. *Journal of Medical Speech-Language Pathology*, *4*, 57–70.

Fougeron, C., Kodrasi, I., & Laganaro, M. (2022). Differentiation of motor speech disorders through the seven deviance scores from MonPaGe-2.0.s. *Brain Sciences*, *12*, 1471–1487.

Gósy, M. (2004). *Fonetika, a beszéd tudománya [Phonetics, The Science of Speech]*. Budapest: Osiris.

Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., Rubin, A., Eller, R., Chandran, S., Abaza, M., Lyons, K., Divi, V., Lott, J., Johnson, J., & Hillenbrand, J. (2014). Quantifying the cepstral peak prominence, a measure of dysphonia. *Journal of Voice*, *28*, 783–788.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, *37*, 769–778.

Juste, F. S., Rondon, S., Sassi, F. C., Ritto, A. P., Colalto, C. A., & Andrade, C. R. (2012). Acoustic analyses of diadochokinesis in fluent and stuttering children. *Clinics*, *67*, 409–414.

Kent, R. D., & Read, C. (2002). *The Acoustic Analysis of Speech*. Albany, NY: Thomson Learning.

Kent, R. D., Weismer, G., Kent, J. F., Vorperian, H. K., & Duffy, J. R. (1999). Acoustic studies of dysarthric speech: Methods, progress, and potential. *Journal of Communication Disorders*, *32*, 141–180.

Kim, Y., Kent, R. D., & Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type, and severity of dysarthria. *Journal of Speech, Language, and Hearing Research*, *54*, 417–429.

Kis, O., Tóth, A., Jakab, K., & Klivényi, P. (2020). A beszédsebesség vizsgálata Parkinson-kór-, sclerosis multiplex, valamint stroke-eredetű dysarthriák esetében [examining speech rate in dysarthria caused by parkinson's disease, sclerosis multiplex, and stroke]. *Rehabilitáció*, *30*, 3–10.

Laforce, R., Jr., & Doyon, J. (2001). Distinct contribution of the striatum and cerebellum to motor learning. *Brain and Cognition*, *45*, 189–211.

Lansford, K. L., & Liss, J. M. (2014). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language, and Hearing Research*, *57*, 57–67.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language, and Hearing Research*, *52*, 1334–1352.

Lowit, A., & Kuschmann, A. (2012). Characterizing intonation deficit in motor speech disorders: An autosegmental-metrical analysis of spontaneous speech in hypokinetic dysarthria, ataxic dysarthria, and foreign accent syndrome. *Journal of Speech, Language, and Hearing Research*, *55*, 1472–1484.

Nishio, M., & Niimi, S. (2001). Speaking rate and its components in dysarthric speakers. *Clinical Linguistics and Phonetics*, *15*, 309–317.

Rollins, M., & Oren, L. (2020). Effects of nasal emission and microphone placement on nasalance score during /s/. *Proceedings of meetings on acoustics Acoustical Society of America*, *42*, 060001.

Thoppil, M., Kumar, C., Kumar, A., & Amose, J. (2017). Speech signal analysis and pattern recognition in diagnosis of dysarthria. *Annals of Indian Academy of Neurology*, *20*, 352– 357.

Zajac, D. J. (2021). Speech aerometry. In M. J. Ball (Ed.), *Manual of Clinical Phonetics* (pp. 264–281). London: Routledge.

Tanulásban akadályozott (enyhe értelmi fogyatékos) fiatalok alaphangjellemzői a spontán beszédben

Jankovics Julianna[1,2]

[1] *BGéSzC Öveges József Technikum és Szakképző Iskola*
[2] *Eötvös Loránd Tudományegyetem*

Abstract

Disorder of intellectual development (intellectual disability) is a collective term that is defined by three factors: reduced intelligence, deficits in adaptive skills, and the appearance of symptoms before the age of 18. Individuals with intellectual disability often experience impairments in general cognitive functions such as thinking and spatial orientation, which significantly impact their language production and perception. This study examines the prosodic structure in spontaneous speech of young adults with mild intellectual disabilities. The main hypotheses are: (1) in all types of spontaneous speech, fundamental frequency is higher in people with mild intellectual disabilities; (2) there are differences between the two genders in prosodic characteristics, the average fundamental frequency of women is higher, and their vocal range and interval are wider compared to men; (3) in four types of spontaneous speech, there is a difference in prosodic characteristics.

The study involved 16 participants with mild intellectual disabilities (8 women and 8 men), with an average age of 19.5 years, and 16 mentally healthy control subjects (8 women and 8 men) of similar ages. The classification of mild intellectual disabilities was determined based on the BNO code and IQ values obtained from expert committee documents.

Four types of audio recordings were created for the study, including a two-part interview, picture description, and narrative recall. The recordings were annotated using Praat software, and scripts were utilized during the analysis to ensure accuracy. The scripts facilitated the determination of average fundamental frequency (f_0), f_0-minimum, and f_0-maximum values for each speech segment. Additionally, the vocal range and interval were calculated for each speech type and segment, representing the distance between the highest and lowest fundamental frequency values.

According to the results, the average fundamental frequency was higher in the speech of people with mild intellectual disabilities in four types of recordings, and in terms of gender, the average f_0 was higher for women, as expected. Furthermore, there was a difference between the prosodic characteristics of each speech type.

Email address: jankovics.julianna@btk.elte.hu (Jankovics Julianna)

Beszédtudomány – Speech Science 2024/1. 136–157. DOI: 10.15775/Besztud.2024.1.136-157

1. Bevezetés

1.1. Az értelmi fogyatékosság terminológiája, nyelvi- és beszédbeli jellemzők

Az értelmi fogyatékosság komplex jelenség, terminológiája időről időre változik a kor kihívásainak megfelelően. Diagnosztizálásához számos körülményt figyelembe kell venni, de az egyértelmű, hogy az értelmi fogyatékos emberek csoportja három fő tulajdonsággal jellemezhető: az első a normál övezet alá eső intellektuális működés; a második az adaptív működés vagy az önellátáshoz szükséges képességek deficitje; a harmadik megállapítás pedig a korai kezdetet hangsúlyozza, vagyis ennek az állapotnak már 18 éves kor előtt jelen kell lennie (Hodapp & Dyckens, 2003, magyar nyelven Csákvári & Mészáros, 2012).

A különböző tudományterületek eltérő fogalmakat alkalmaznak az állapot leírására. A *mentális retardáció* és az *értelmi fogyatékosság* elsősorban az orvosi terminológia körébe tartozik. Az Egészségügyi Világszervezet (WHO) által kiadott orvosi klasszifikációs rendszer, az International Classification of Diseases (ICD – nálunk Betegségek Nemzetközi Osztályozása, BNO) a *mentális retardáció* terminust alkalmazza. A mentális retardáció „abbamaradt vagy nem teljes szellemi fejlődés, amelyre jellemző a különböző készségek romlása, olyan készségeké, melyek a fejlődés során jelennek meg, és készségeké, amelyek az intellektus minden szintjét érintik, például a kognitív, nyelvi, mozgásbeli, szociális készségek, képességek" (BNO-10, 1995, 334). Az előbbi meghatározás az intellektus és a készségek, képességek közötti kapcsolatot hangsúlyozza. A hazánkban alkalmazott klasszifikációs rendszer a BNO 10. revíziója alapján került kidolgozásra. Így Magyarországon az IQ-pontszámok alapján hat csoportját különböztetjük meg az értelmi fogyatékosságnak (enyhe, közepes, súlyos, igen súlyos, mentális retardáció, nem osztályozott mentális retardáció). Az Egészségügyi Világszervezet 2018. június 18-án kiadta a BNO újabb, 11. kiadását. Az osztályozást elfogadták, és ez a kiadás 2022. január 1-jén lépett hatályba. Ebben több jelentős változás történt az előző verzióhoz képest. Míg a 10. változatban a *Mentális és viselkedészavarok* főcsoportban jelent meg a mentális retardáció F70–F79 kóddal, addig a 11. kiadásban ebben a főcsoportban a *Disszociatív*

*zavarok*on belül jelenik meg, a kódolás pedig 6B60–6B6Z közé esik. A legfőbb változás pedig, hogy a mentális retardáció az *értelmi fejlődés zavará*ra módosul (BNO-10, 1995; Szekeres, 2018; ICD-11, 2019; BNO-11, 2022). Fontos azonban megjegyezni, hogy az új kódolási rendszerrel még csak ismerkednek a különböző intézmények, többek között a szakértői bizottságok még a BNO 10. revíziója alapján készítik el a kódolást.

Az *intellektuális képességzavar* pszichológiai szempontú terminus *(intellectual disability)*, amely azokra a személyekre alkalmazható, akik „az intellektuális-kognitív működések, valamint a kortárs csoportokhoz viszonyított adaptív magatartás jelentős akadályozottságával jellemezhetőek" (Lányiné Engelmayer, 2017, 15).

A (gyógy)pedagógia is saját terminust alkalmaz a különböző állapotok leírására: *akadályozottság.* Az akadályozottság nem egy rögzített állapot, hanem ez dinamikusan változik a környezettel együtt (Hatos, 2015). A különböző ellátórendszerekben (pl. köznevelési intézményekben) az akadályozott élethelyzetbe került személyek érdekében a gyógypedagógusnak feladata egyrészt a személyes képességek speciális fejlesztése, másrészt a környezet állította akadályok (szociális hátrányok, előítéletek vagy hiányzó szolgáltatások) csökkentése (Mesterházi, 2001, 156-157). Az akadályozottságon belül két csoportot különítünk el. A *tanulási akadályozottság* „gyógypedagógiai pszichológia komplex vizsgáló eljárásaival megállapított intellektuális képességzavar (ezen belül az enyhe fokú értelmi fogyatékosság) következtében kialakuló átfogó és tartós iskolai tanulási nehézség. [...] A tanulási akadályozottságra jellemző tanulási nehézségek egyes formái azonban ennél szélesebb körben is előfordulhatnak a tanulási szempontból nem diagnosztizált, tanköteles korú népesség gyenge tanulási eredményt mutató tanulóinak körében" (Mesterházi & Szekeres, 2021, 114). Ahogy tehát a definíció mutatja, ebbe a csoportba sorolhatók egyrészt az enyhén értelmi fogyatékosnak minősített tanulók, valamint a gyenge tanulmányi eredménnyel rendelkező diákok. Az *értelmi akadályozott* személyek pedig elsősorban középsúlyosan vagy súlyosan értelmi fogyatékosok, akiknek teljes életen át tartó speciális támogatásra van szükségük (Hatos, 2001; Barthel, 2022).

A jelen kutatásban a fenti definíciókból kiindulva az általam vizsgált populációnál az **enyhe értelmi fogyatékos** kifejezést használom. A terminus alatt a fentebb említett hármas tényező (csökkent intelligencia, az adaptív funkciók gyengesége, korai kezdet) teljesülését veszem figyelembe az intelligencia kitüntetett szerepével. Erre azért van szükség, mivel a kísérleti személyeket a szakértői véleményekben található diagnózis és BNO-kód alapján soroltam az enyhe értelmi fogyatékos férfiak és nők csoportjába.

Az értelmi fogyatékos személyek beszédprodukciója és -percepciója eltér az ép fejlődésű személyekétől, és ez azzal magyarázható, hogy az általános információfeldolgozó funkciók, valamint a gondolkodás, a téri tájékozódás gyakran sérülnek értelmi fogyatékosság esetén (Lukács & Kas, 2014). Ebben a csoportban szinte mindenkinél jellemző a megkésett beszédfejlődés, és nem ritka, hogy bizonyos beszédfejlődési szakaszon hosszabb ideig megrekednek (Lányiné Engelmayer, 2017). Radványi (2005) szerint az alábbi területeken mutathatók ki elmaradások: lelassult fejlődés, gyenge kommunikációs szándék, a beszédészlelés és a beszédmegértés zavara, hangok képzési torzítása, a tartalmi kifejezés szegénysége, szűkös szókincs, diszgrammatizmus, megrekedés az alacsony közlési formáknál, továbbá a kommunikációs nehézségek közé sorolható a metakommunikáció nehezítettsége is (Mesterházi & Szekeres, 2021). A gyenge kommunikációs szándék leginkább abban mutatkozik meg, hogy kevésszer kezdeményeznek beszélgetést, valamint ritkán fogalmaznak meg spontán módon kérdéseket (Hatos, 2000).

1.2. Prozódia és alaphangjellemzők

A prozódia/szupraszegmentum fogalmát sokan sokféleképpen próbálták meg definiálni. A (Markó, 2015, 18) által megfogalmazott definíció az egymáshoz tartozást erősíti, ezért a prozódia fogalmán az alábbi definíciót értem: „a beszédprodukciós folyamat által létrehozott komplex beszédjelnek az a vetülete, amely az idő, a frekvencia és az intenzitás folyamatváltozásaiként írható le, és amelynek az észlelése kizárólag állandó viszonyításban, nagyobb egységeken (minimálisan szótagok viszonylatában) lehetséges."

A beszéddallam (más szóval hanglejtés vagy intonáció) „a zönge alaphangja modulációjának az észlelésünkre tett hatása" (Markó, 2015, 21). A beszéd dallama a hangszalagok rezgésén alapul, akusztikai szempontból a zönge legkisebb frekvenciájú és legnagyobb amplitúdójú összetevője, az alaphang (f0) folyamatos és célzott változtatásának az eredménye (Gósy, 2004, 187). A dallamészlelet és az alapfrekvencia között pszichofizikai kapcsolat van (Beckman & Venditti, 2010). A hanglejtés változásának mértékét, terjedelmét több tényező befolyásolja. A beszéddallam összefügg az idő paraméterével, hiszen az adott időegység alatti f0-változás mértéke határozza meg. A beszéddallamot nyelvi-kommunikációs tényezők (például a beszédhelyzet, a beszédtípus) befolyásolják (Markó, 2017). A hanglejtés továbbá a hangsúllyal is összefügg, hiszen az alapfrekvencia változása mindkét szupraszegmentális tényezőben alapvető fontosságú, emiatt legtöbbször sem akusztikai eszközökkel, sem perceptuálisan nincs lehetőség az alapfrekvencia és a beszéddallam szétválasztására (Markó, 2015).

Mind a nemzetközi, mind a magyar fonetikai kutatásokban gyakori az eltérő korú beszélők beszédbeli jellemzőinek vizsgálata (vö. például Markó & Bóna, 2012; Markó, 2015; Markó et al., 2021), valamint a különböző beszédtípusok (leggyakrabban a spontán beszéd és az olvasás) alaphangjellemzőinek (vö. például Abu-Al-Makarem & Petrosino, 2007; Skarnitzl & Vaňková, 2017; Tóth, 2017; Gráczi et al., 2019) összehasonlítása egészséges beszélők esetében. A nemek között meglévő különbségek a férfiaknál és a nőknél is különböző hormonális és pszichés hatásokra vezethetők vissza (Balázs & Bóna, 2016), továbbá a gége folyamatosan változik a reproduktív évek alatt (Raj et al., 2010).

Az értelmi fogyatékos beszélők alaphangjellemzőit korábban csak a nemzetközi irodalom vizsgálta Down-szindrómával élő gyermekek és felnőttek körében. Lee és munkatársai (2009) 17 és 29 év közötti Down-szindrómás (DSZ) felnőtteket (5 nő, 4 férfi) vizsgáltak, akiknek eredményeit nemben és korban illesztett nem értelmi fogyatékos kortársaik eredményével vetették össze (9 fő, átlagos életkoruk 23,5 év). Négy feladat – kitartott *ah* hangkapcsolat, emelkedő és ereszkedő glisszandó (két egymástól bizonyos távolságra fekvő hang közötti csúszás, Böhm, 1961, 101), rövid szöveg felolvasása és egyperces spontán beszéd

– alapján térképezték fel az akusztikai paramétereket. Az eredmények alapján megállapítható, hogy az átlagos f0 szignifikánsan magasabb volt a DSZ csoportokban (Lee et al., 2009). Az átlagos f0 tekintetében hasonló eredményre jutottak Seifpanahi és munkatársai (2011) is, akik huszonkét 20 és 28 év közötti (átlagos életkoruk 25,0 év, 8 nő, 14 férfi) 50 és 65 IQ-pont közötti DSZ beszélő értékeit vetették össze nemben és korban illesztett ép beszélők értékeivel. A kutatás célja az volt, hogy feltárják a legfőbb akusztikai különbségeket a DSZ és a mentálisan ép felnőttek között, valamint az irodalomban először megadják az objektív vokális paramétereket a fárszi nyelvet beszélő DSZ felnőtteknél. A kutatás eredményei alapján az átlagos f0 mindkét nemben szignifikánsan magasabb volt a Down-szindrómás beszélők esetében, a jitterértékek (a hangszalagrezgések frekvenciaingadozásának mértékét mutatja, Gósy, 2004, 31) pedig szignifikánsan alacsonyabbak voltak náluk a kontrollszemélyek értékeihez mérve. Az átlagos fonációs időt (zöngeképzési idő) tekintve a DSZ és a mentálisan ép beszélők között nem volt statisztikailag is igazolható a különbség, ám a nemek tekintetében a férfiak produkáltak statisztikailag is igazolhatóan hosszabb fonációs időt mindkét csoportban (Seifpanahi et al., 2011). Albertini et al. (2010) vizsgálatában 30 DSZ felnőtt (17 férfi, átlagos életkoruk 28,7 év és 13 nő, átlagos életkoruk 23,2 év) vett részt, a kontrollszemélyek közé pedig 60 ép halló és nem dohányzó embert (30 férfi, átlagos életkoruk 48,1 év és 30 nő, átlagos életkoruk 44,7 év) soroltak. A résztvevőknek a hallott szavakat kellett megismételniük két alkalommal, de az elemzés csak az első sorozatra terjedt ki. A szerzők többek között elemezték az átlagos alapfrekvenciát, az intenzitást, a jitter- és shimmerértéket (a hangszalagrezgések amplitúdóingadozásának mértékét mutatja, Gósy, 2004, 31), valamint az időtartamot. Az eredmények szerint a DSZ beszélőknél, különösen a férfiaknál szignifikánsan magasabb értékeket mértek az f0 tekintetében, továbbá mindkét nem esetében szignifikánsan alacsonyabb intenzitásértékek születtek a kontrollszemélyekhez viszonyítva. Továbbá csak a DSZ férfiaknál rövidebb időtartamot és alacsonyabb shimmerértékeket mértek az ép felnőttekhez képest.

Magyar nyelven mindeddig csak a tanulmány szerzője készített olyan vizsgálatokat, amelyekben az értelmi fogyatékos felnőttek beszédének szupraszegmentális szerkezetét térképezték fel. Egy 2019-es kutatásban az alaphangjellemzőket enyhe és középsúlyos értelmi fogyatékos nők (5 fő, átlagéletkor: 32,4 év) és kontrollszemélyek (5 fő, átlagéletkor: 32,0 év) spontán beszédében és felolvasásában elemezték (Jankovics, 2019). A beszédtípus és az alapfrekvencia összefüggését vizsgálva elmondható, hogy mindkét beszélői csoport esetében a felolvasásban volt a legmagasabb az átlagos alapfrekvencia, de az egyének közötti és az egyénen belüli variabilitás is jelentős volt. A csoportonként számolt átlagos hangterjedelem az értelmi fogyatékos csoportban 3,9 félhang, a kontrollszemélyeknél 4,1 félhang volt. A beszédtípusokat tekintve az értelmi fogyatékosok csoportjában a hangterjedelem a felolvasásban volt a legnagyobb (3,9 félhang), ezt követte az interjú (3,8 félhang), a képleírás (3,5 félhang), majd a tartalomösszegzés (3,3 félhang). A kontrollszemélyeknél ettől eltérően alakultak az eredmények, náluk a legszélesebb hangterjedelem az interjúban (4,1 félhang) volt, majd a felolvasás (3,9 félhang), a tartalomösszegzés (3,3 félhang) és a képleírás (3,2 félhang) következett. Az eredmények értelmezésénél azonban fontos szempont, hogy az 5. női kontrollszemélyt ki kellett zárni a vizsgálatból, mert az ő eredményei minden tekintetben eltértek a többiekétől.

A jelen kutatás célja, hogy megvizsgálja az alaphangjellemzőket (átlagos alapfrekvencia, hangterjedelem, hangköz) a tanulásban akadályozott, azon belül enyhe értelmi fogyatékos fiatalok és a nemben és korban illesztett kontrollszemélyek beszédében. A nemzetközi kutatások eredményei, valamint a saját vizsgálataim alapján az alábbi hipotéziseket fogalmaztam meg: H1: a spontán beszéd különböző típusainak az alapfrekvencia-szerkezetében különbség mutatkozik az enyhe értelmi fogyatékos személyek és a kontrollcsoport között: minden beszédtípusban az enyhe értelmi fogyatékosoknál magasabb az átlagos alapfrekvencia, tágabb a hangterjedelem és beszédszakaszonként tágabb a hangközök értéke; H2: a két nem között eltérések láthatók az alaphangjellemzőkben, a nőknek magasabb az átlagos alapfrekvenciaértéke és náluk tágabb hangterjedelem-, illetve hangközértékek mutathatók ki a férfiakhoz képest; H3: a vizsgált be-

szédtípusokban különbség mutatkozik az alaphangjellemzők tekintetében. Természetesen mindegyik feltételezés során figyelembe kell venni az enyhe értelmi fogyatékos fiatalok és a kontrollcsoport közötti eltéréseket.

2. Kísérleti személyek, módszertan

A jelen kutatásban összesen 32 fő szerepel, 8 enyhe értelmi fogyatékos férfi (EF, átlagéletkor: 19,6 év), 8 enyhe értelmi fogyatékos nő (EN, átlagéletkor: 19,4 év); a kontrollcsoportba pedig szintén 8 férfi (KF, átlagéletkor: 19,5 év) és 8 nő (KN, átlagéletkor: 20,0 év) tartozik. Az enyhe értelmi fogyatékos fiatalok mindannyian egy-egy budapesti szakiskola tanulói. Az enyhe értelmi fogyatékosság állapotát a szakértői bizottságok által kiállított dokumentumban található BNO-kód és IQ-érték alapján állapítottam meg. Magyarországon az enyhe értelmi fogyatékosság kódja az F70 (BNO-10, 1995), amely pontszámban kifejezve 50–69 IQ-pont között található. A kontrollszemélyek nemben és korban illeszkednek az enyhe értelmi fogyatékosokhoz. A kutatásban résztvevők magyar anyanyelvűek, és nincs halláskárosodásuk. A kutatás középpontjában álló személyek kiválasztása az alábbi kritériumok szerint történt: (i) mindegyikük enyhe értelmi fogyatékos legyen, (ii) semmilyen szindróma ne szerepeljen az anamnézisben, (iii) a kontrollszemélyek életkori sávjához (17–24 év) illeszkedjen az enyhe értelmi fogyatékos személyek életkora.

A hangfelvételek saját gyűjtésű anyagok, amelyek az enyhe értelmi fogyatékos fiatalok esetében az iskolában csendes körülmények között, míg a kontrollszemélyek esetében az ELTE BTK Alkalmazott Nyelvészeti és Fonetikai Tanszék stúdiójában lettek rögzítve 44,1 kHz-es mintavételezési frekvencián, legalább 16 biten az Audacity nevű programmal (Audacity Team, 2018).

Összesen négyféle hangfelvételt rögzítettem. Mindegyik feladat alapját a Gyermeknyelvi AdatBázis és Információtár (Bóna et al., 2014) képezte. Az interjú két részből épült fel, az első részben (interjú1) a kísérletvezető volt a riporter, ő tett fel a mindennapi élethez kapcsolódó kérdéseket az adatközlőknek. Az interjú második részében (interjú2) szerepcsere történt, az adatközlőknek

kellett kérdéseket feltenniük a kísérletvezetőnek. A történetmondás feladatában (tartalomösszegzés) egy hangfelvételről meghallgatott mese tartalmát kellett visszaadniuk. A képleírási feladatban (képleírás) pedig az adatközlőknek egy nyolc képből álló, fekete-fehér képsorozatot kellett elmesélniük. A GABI adatbázisban a képleírás esetében eredetileg 6 kép szerepelt, ezt egészítette ki Tóth Andrea (2017 saját készítésű képekkel, így a kutatás anyagát végül ez a 8 képből álló sorozat alkotja. A feladatok eltérő jellege miatt különböző terjedelmű beszédminták születtek. A beszédtípusok átlagos időtartamát az 1. táblázat tartalmazza.

A kutatás teljes ideje alatt az érzékeny adatközlői csoportok részvétele miatt kiemelten fontos volt a kutatásetikai szempontok betartása (vö. Bárth et al., 2021). A hangfelvételek elkészítése előtt először szóban, majd pedig írásban tájékoztattam a résztvevőket a kutatás részleteiről. Ezt követően az enyhe értelmi fogyatékos személyek esetében a szülő, a gondnok[1] vagy a gyám[2], míg a kontrollszemélyek saját maguk írták alá és töltötték ki a hozzájárulási nyilatkozatot, valamint az adatlapot. A nyilatkozatban a kutatás fókuszában álló csoportok (EF és EN) esetében arról is nyilatkozni kellett, hogy a szakértői véleményeket a kutatás vezetője megismerheti-e vagy sem. Ezek a dokumentumok tartalmazzák a budapesti és a vidéki szakértői bizottságok által kiadott szakértői véleményeket, melyekben szerepel a pontos diagnózis a BNO-kóddal együtt.

A hanganyagot szakaszszinten annotáltam a *Praat* 6.0.43. program (Boersma & Weenink, 2016) segítségével. Az elemzés során többszintű annotálást használtam, amelyet manuálisan végeztem el. Az első szint az adatközlők beszédét jelölte. A második szint a háromféle szünetet – néma, jellel kitöltött, kombinált szünet – tartalmazta. Minden percepciósan észlelhető szünetet figyelembe vettem, így nem határoztam meg minimumértéket (vö. Fletcher, 2010). Az alapfrekvenciára vonatkozó adatok kinyerése egy erre a célra készített script

[1] 2013. évi V. törvény a Polgári Törvénykönyvről; Második rész: A cselekvőképesség (2013)

[2] 149/1997. (IX. 10.) Korm. rendelet a gyámhatóságokról, valamint a gyermekvédelmi és gyámügyi eljárásról, 127. § (1997)

segítségével történt. A script beszédszakaszonként határozta meg az átlagos f0, az f0-minimum és az f0-maximum értékét. A beállításokban a várható f0-tartományt a nőknél 100–450 Hz, a férfiaknál 75–350 Hz között adtam meg. A mért adatokat kézzel ellenőriztem, a félremérésből származó adatokat töröltem. Ezek után beszédtípusonként meghatároztam a beszélők átlagos alapfrekvenciáját. A hangterjedelmet beszédtípusonként, a hangköz értékét a beszédtípusokban található beszédszakaszonként határoztam meg, amelynél a legmagasabb és a legalacsonyabb f0 távolságát vettem alapul. Az egyes frekvenciaértékek távolságait félhangban számítottam ki, amelyhez az alábbi képletet használtam: $12 \cdot \log_2 (f_{0_{\max}} / f_{0_{\min}})$, ahol az f0-maximum és az f0-minimum Hz-ben mért értékeit vettem alapul (vö. Huszár, 2019).

Az interjú első részében csak az első 1000 szótagot elemeztem, az interjú második részében pedig csak az adatközlők által feltett kérdéseket annotáltam. A tartalomösszegzést és a képleírást teljes terjedelmében elemeztem.

A statisztikai elemzést az SPSS 20. szoftver segítségével végeztem el. Az elemzésbe bevont változók ellenőrzése során kitértem a normalitás vizsgálatára annak érdekében, hogy el tudjam dönteni, hogy alkalmazhatok-e parametrikus teszteket. A kapott eredmények alapján az elemzések elvégzésekor minden esetben nemparametrikus próbát használtam, mivel a Kolmogorov–Smirnov-próba eredménye minden esetben $p < 0,05$ volt, amely azt mutatta meg, hogy a változóim nem normális eloszlást követnek. Így a továbbiakban nemparametrikus próbákat alkalmaztam: Kruska–Wallis-próbát alkalmaztam két-két csoport között a hangterjedelem és a hangköz tekintetében az esetleges eltérések megállapítására, továbbá Mann–Whitney-próbával ellenőriztem az enyhe értelmi fogyatékos nők és a kontrollnők, valamint az enyhe értelmi fogyatékos férfiak és a kontrollférfiak közötti statisztikai eltéréseket. A Spearman-korrelációval csoportonként és beszédtípusonként vizsgáltam meg az átlagos alapfrekvencia, a hangterjedelem és a hangköz értékei közötti korreláció minőségét. A tesztelés minden esetben 95%-os szignifikanciaszinten történt.

1. táblázat. A beszédtípusok átlagos időtartama (s) a beszélői csoportokban.

Beszédtípus	**Átlagos beszédidő (s)**			
	EF	**EN**	**KF**	**KN**
interjú 1. része	145,4	128,5	118,6	158,8
interjú 2. része	11,5	13,2	30,8	36,2
képleírás	48,2	56,8	48,0	55,5
tartalomösszegzés	54,6	34,3	82,4	87,2

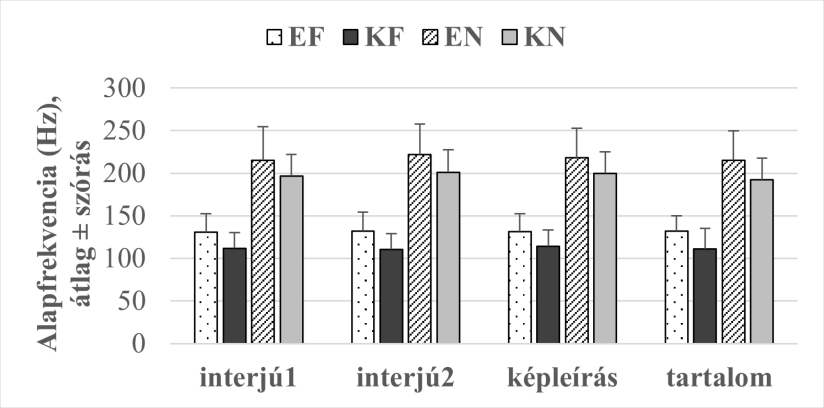
3. Eredmények

A négy beszédtípusban (interjú1, interjú2, képleírás, tartalom) az átlagos alapfrekvenciát, valamint a hangterjedelmet és a hangközt számítottam ki. Az eredmények kiértékelésénél a négy csoportot az enyhe értelmi fogyatékos férfiak (EF) és nők (EN), valamint a kontrollcsoport férfi (KF) és női (KN) tagjai jelentették.

Elsőként az átlagos f0 értékeit elemeztem. A kutatásban résztvevő adatközlők esetében a kontrollférfiaknál volt eltérés a beszédtípusok tekintetében, náluk a képleírásban volt a legmagasabb az átlagos alapfrekvencia (113,96 Hz, szórás = 19,05 Hz), míg a másik három csoportnál az interjú második részében mértem a legnagyobb értékeket. Az enyhe értelmi fogyatékos férfiaknál az interjú első részében (130,89 Hz, szórás = 21,39 Hz), míg a többi csoport esetében a tartalomösszegzésben mértem a legalacsonyabb értékeket. Minden esetben az enyhe értelmi fogyatékosoknál volt nagyobb az átlagos f0 értéke, a nemeket tekintve pedig természetszerűleg a nőknél volt magasabb az átlagos f0 a férfiakhoz képest. Az egyes csoportoknál a beszédtípusokban tapasztalható szórásértékek nagyban különböztek egymástól, az értékek szóródása minden beszédtípusban az enyhe értelmi fogyatékos női adatközlőknél volt a legnagyobb (>34 Hz) (1. ábra és 2. táblázat).

A Mann–Whitney-teszt az alapfrekvencia tekintetében az alábbi beszédtípusokban igazolt szignifikáns különbséget két-két csoport összefüggésében: az enyhe értelmi fogyatékos férfiak és a kontrollférfiak között az interjú első ré-

szében ($Z = -19,242$, $p < 0,001$), az interjú második részében ($Z = -8,285$, $p < 0,001$), a képleírásban ($Z = -9,706$, $p < 0,001$) és a tartalomösszegzésben ($Z = -14,512$, $p < 0,001$); a két női csoport között az interjú első részében ($Z = -11,555$, $p < 0,001$), az interjú második részében ($Z = -4,361$, $p < 0,001$), a képleírásban ($Z = -6,956$, $p < 0,001$) és a tartalomösszegzésben ($Z = -7,058$, $p < 0,001$).



1. ábra. Alapfrekvencia (Hz) a beszédtípusokban.

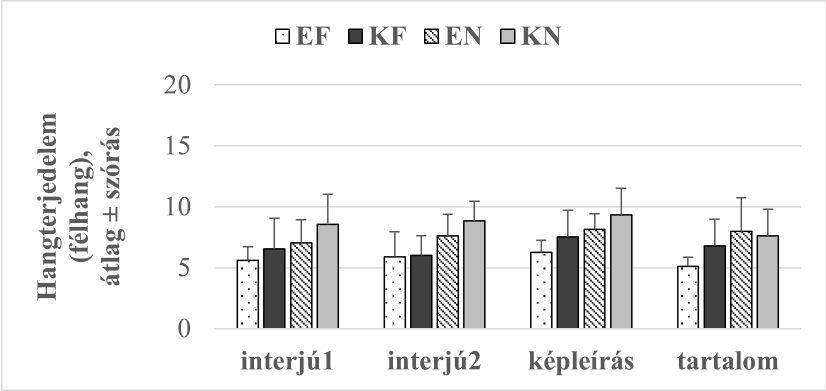
2. táblázat. Az átlagos alapfrekvencia (Hz) értékei a beszédtípusokban.

		EF	**KF**	**EN**	**KN**
interjú1	átlag	130,89	111,48	215,34	196,38
	szórás	21,39	18,80	39,02	25,29
interjú2	átlag	132,30	110,44	221,81	200,98
	szórás	21,90	18,40	35,80	26,73
képleírás	átlag	131,19	113,96	218,34	199,78
	szórás	20,92	19,05	34,12	25,04
tartalomösszegzés	átlag	132,27	111,09	215,14	192,13
	szórás	17,45	24,09	34,51	25,36

A beszélők hangterjedelmét a logaritmus függvény alapján számoltam ki, így minden beszédtípusra egy-egy érték született adatközlőnként. Mindegyik

csoportban a képleírás feladatában mértem a legnagyobb hangterjedelmet (EF: 6,27 Hz; KF: 7,53 Hz; EN: 7,63 Hz; KN: 9,34 Hz), de a legszűkebb hangterjedelem esetében már nem volt teljes egyezés a csoportok között. Az enyhe értelmi fogyatékos férfiaknál és a kontrollnőknél a tartalomösszegzésben, a kontrollférfiaknál az interjú második részében, az enyhe értelmi fogyatékos nőknél pedig az interjú első részében mértem a legszűkebb hangterjedelmet. Mind az enyhe értelmi fogyatékos csoportok, mind a kontrollcsoportok összehasonlításában a nők esetében volt nagyobb a hangterjedelem értéke a férfiak értékeihez viszonyítva (2. ábra és 3. táblázat).

A Kruskal–Wallis-próba minden beszédtípusban szignifikáns eltérést igazolt a résztvevői csoportok között a hangterjedelem tekintetében. A Mann–Whitney-próba az alábbi beszédtípusokban igazolt szignifikáns eltérést két-két résztvevői csoport összehasonlításában: a két enyhe értelmi fogyatékos csoport között a képleírásban ($Z = -2,521$, $p = 0,012$) és a tartalomösszegzésben ($Z = -2,310$, $p = 0,021$); a két férfi csoport összevetésében a tartalomösszegzésben ($Z = -2,521$, $p = 0,012$); a kontrollcsoportok között az interjú második részében ($Z = -2,415$, $p = 0,016$).



2. ábra. Hangterjedelem (félhang) a beszédtípusokban.

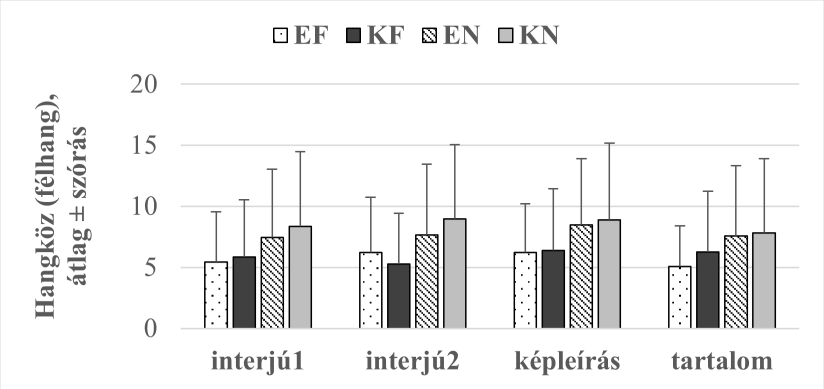
A megnyilatkozásokban mért hangközértékek az alábbiak szerint alakultak. A kontrollférfiaknál és az enyhe értelmi fogyatékos nőknél a képleírásban, a másik két csoportban az interjú második részében fordult elő a legtágabb hangköz.

3. táblázat. A hangterjedelem (félhang) értékei a beszédtípusokban.

		EF	**KF**	**EN**	**KN**
interjú1	átlag	5.59	6,54	7,02	8,54
	szórás	1,1	1,5	1,89	2,5
interjú2	átlag	5,89	6,00	7,63	8,84
	szórás	2,06	2,85	1,74	1,6
képleírás	átlag	6,27	7,53	8,16	9,34
	szórás	0,98	2,46	1,25	2,19
tartalomösszegzés	átlag	5,09	6,81	7,99	7,63
	szórás	0,76	1,74	2,74	2,16

Az enyhe értelmi fogyatékos férfiaknál és a kontrollnőknél a tartalomösszegzésben, a kontrollférfiaknál pedig az interjú második részében volt a legalacsonyabb a hangközök értéke, de az enyhe értelmi fogyatékos női adatközlőknél az interjú első részében mértem a legalacsonyabb hangközértéket. Mind a négy csoportban a vizsgált beszédtípusok összességében, az enyhe értelmi fogyatékos személyeknél voltak szűkebbek a hangközök a kontrollszemélyekhez viszonyítva a férfiaknál és a nőknél egyaránt, az egyetlen kivételt az interjú második része jelentette, amelyben az enyhe értelmi fogyatékos férfiaknál volt tágabb a hangköz a hozzájuk illesztett kontrollszemélyekhez képest (3. ábra és 4. táblázat).

A Kruskal–Wallis-próba mindegyik spontán beszédtípusban szignifikáns eltérést igazolt. A Mann–Whitney-próba a hangköz tekintetében az alábbi beszédtípusokban igazolt szignifikáns eltérést két-két csoport összehasonlításában: az enyhe értelmi fogyatékos férfiak és nők között az interjú első részében ($Z = -5,391,\ p < 0,001$), a képleírásban ($Z = -4,003,\ p < 0,001$) és a tartalomösszegzésben ($Z = -3,475,\ p < 0,001$); a két kontrollcsoport között az interjú első részében ($Z = -5,854,\ p < 0,001$), az interjú második részében ($Z = -4,893,\ p < 0,001$), a képleírásban ($Z = -3,940,\ p < 0,001$), valamint a tartalomösszegzésben ($Z = -2,668,\ p < 0,008$).



3. ábra. Hangköz (félhang) a beszédtípusokban.

4. táblázat. A hangköz (félhang) értékei a beszédtípusokban.

		EF	**KF**	**EN**	**KN**
interjú1	átlag	5,45	5,85	7,44	8,34
	szórás	4,1	4,69	5,61	6,13
interjú2	átlag	6,21	5,27	7,65	8,98
	szórás	4,51	4,15	5,78	6,06
képleírás	átlag	6,2	6,4	8,47	8,87
	szórás	4,01	5,01	5,43	6,28
tartalomösszegzés	átlag	5,05	6,25	7,57	7,83
	szórás	3,33	4,98	5,76	6,08

Az eddig megismert nemzetközi kutatások alapján nem tudtam hipotézist felállítani arra vonatkozóan, hogy milyen összefüggés van az alaphangjellemzők között az enyhe értelmi fogyatékos személyek és a kontrollszemélyek esetében, de szerettem volna erre is kitérni a vizsgálatban.

Markó munkatársaival (Markó et al., 2021) felnőtt beszélők felolvasásában elemezte az alaphangjellemzőket, a vizsgálat eredménye szerint mind az f0 átlaga, mind a hangterjedelem és a hangköz esetében is nagy volt a beszélők közötti variabilitás, amelyek hátterében különböző tényezők állhatnak. Az egyes alaphangjellemzők közötti összefüggések ezért teljesen egyénfüggők.

A Spearman-féle korrelációelemzést mind a négy csoport között elvégeztem, a tényezők közötti kapcsolat csak az átlagos alapfrekvencia és a hangköz között volt kimutatható. Az átlagos f0 és a hangköz között az enyhe értelmi fogyatékos férfiak csoportjában gyenge kapcsolat van ($\rho = 0,363$); a kontrollcsoport férfi tagjainál közepesen erős kapcsolat mutatható ki ($\rho = 0,496$); az enyhe értelmi fogyatékos nők csoportjában gyenge ($\rho = 0,244$), majdnem elhanyagolható a kapcsolat, míg a női kontrollcsoportban az átlagos f0 és a hangköz között szintén nagyon gyenge a korreláció ($\rho = 0,107$), ami azt jelenti, hogy az átlagos alapfrekvencia nagysága kevéssé befolyásolja a hangköz tágságát.

4. Összegzés, következtetések

A jelen kutatásban 16 enyhe értelmi fogyatékos adatközlő és 16 nemben és korban illesztett kontrollszemély alaphangjellemzőit vizsgáltam meg a spontán beszéd különböző típusaiban, kétrészes interjúban, képleírásban és tartalomösszegzésben. Az alaphangjellemzőkön belül az átlagos alapfrekvencia, a hangterjedelem és a hangköz értékeit elemeztem négy csoportban (EF, EN, KF, KN).

Az első feltételezésem az volt, hogy a vizsgált beszédtípusok mindegyikében az enyhe értelmi fogyatékos adatközlőknél magasabb az átlagos alapfrekvencia, tágabb a hangterjedelem és beszédszakaszonként tágabb a hangközök értéke. Ezt a hipotézist csak részben támasztották alá az eredmények, hiszen minden beszédtípusban az enyhe értelmi fogyatékos fiataloknál volt magasabb az átlagos alapfrekvencia, melynek hátterében elsősorban a mentális állapottal összefüggő eltérések állhatnak (vö. Lee et al., 2009; Albertini et al., 2010; Seifpanahi et al., 2011), de nem támasztották alá az adatok minden beszédtípusban, hogy az enyhe értelmi fogyatékos személyeknél tágabb a hangterjedelem és beszédszakaszonként tágabb a hangközök értéke.

A második hipotézisem szerint a két nem között jelentős az eltérés az alaphangjellemzőkben. Természetesen a nőknél magasabb az átlagos alapfrekvencia értéke, továbbá az enyhe értelmi fogyatékos nőknél szélesebb a hangterjedelem és a hangköz az ugyanehhez a populációhoz tartozó férfiakhoz képest, a kont-

rollcsoportokban azonban nem minden esetben teljesült ez a tendencia. Ez a nemek közötti hormonháztartás és anatómiai eltérés különbségeiből adódik (vö. Raj et al., 2010; Balázs & Bóna, 2016).

A harmadik hipotézis arra vonatkozott, hogy a vizsgált beszédtípusok között eltérés mutatkozik a vizsgált alaphangjellemzőkben. Ez a feltételezés is alátámasztást nyert, mert mind a négy beszédtípusban mért értékek között különbséget adatoltam az átlagos alapfrekvencia, valamint a hangterjedelem és a hangköz mértékének alakulásában. Ennek hátterében egyrészt az eltérő beszédtervezési stratégiák alkalmazása áll, másrészt szerepet játszhat a beszélők közötti és a beszélőn belüli variabilitás is (vö. pl. Markó et al., 2021).

A csoportok közötti jelentős eltérések statisztikailag nem minden esetben igazolódtak. A Spearman-féle korrelációelemzés csak az átlagos f0 és a hangköz között igazolt gyenge összefüggést (a kontrollférfiak esetében közepesen erős volt ez a kapcsolat).

A jelen kutatás eredményei több területen is hasznosíthatók. Elsőként, mivel az enyhe értelmi fogyatékos fiatalok beszéde eddig kevésbé kutatott terület, mint például a Down-szindrómával élők beszéde, így jó összehasonlítási alapként szolgálhatnak a jelenlegi kutatás eredményei a más súlyossági kategóriába sorolt személyek eredményével. A különböző vizsgálatok hozzájárulnak ahhoz, hogy az enyhe értelmi fogyatékos személyek beszédének szegmentális szintje mellett a szupraszegmentális sajátosságait is megismerjük. A bemutatott eredmények egyrészt a gyógypedagógiában válhatnak hasznosíthatóvá különböző terápiás módszertanok kidolgozásával. Továbbá a jelenlegi eredmények hozzájárulhatnak a fonetikai és pszicholingvisztikai, valamint a társadalom más speciális csoportjának beszédét feltérképező kutatásokhoz.

A kutatás során kialakított korpusz további lehetőségeket kínál az alaphangjellemzők vizsgálatára (pl. jitter és shimmer, intenzitás), továbbá a különböző beszédtípusok jó alapot biztosítanak a beszéddallam (kijelentő és kérdő) elemzésére az enyhe értelmi fogyatékos személyeknél.

Hivatkozások

149/1997. (IX. 10.) Korm. rendelet a gyámhatóságokról, valamint a gyermekvédelmi és gyámügyi eljárásról, 127. § (1997). URL: https://net.jogtar.hu/jogszabaly?docid=99700149.kor a letöltés ideje: 2021. július 17.

2013. évi V. törvény a Polgári Törvénykönyvről; Második rész: A cselekvőképesség (2013). URL: https://net.jogtar.hu/jogszabaly?docid=A1300005.TV&searchUrl=/gyorskereso%3Fkeyword%3Dpolg%25C3%25A1ri%2Bt%25C3%25B6rv%25C3%25A9nyk%25C3%25B6nyv a letöltés ideje: 2018. március 20.

Abu-Al-Makarem, A., & Petrosino, L. (2007). Reading and spontaneous speaking fundamental frequency of young arabic men for arabic and english languages: a comparative study. *Perceptual and Motor Skills*, *105*, 572–580.

Albertini, G., Bonassi, S., Dall'Armi, V., Giachetti, I., Giaquinto, S., & Mignano, M. (2010). Spectral analysis of the voice in down syndrome. *Research in Developmental Disabilities*, *31*, 995–1001.

Audacity Team (2018). Audacity (version 2.3.1). URL: https://www.audacityteam.org/ letöltés ideje: 2018. november 1.

Balázs, B., & Bóna, J. (2016). Életkori sajátosságok a beszédképzésben és a beszédfeldolgozásban. In J. Bóna (Ed.), *Fonetikai olvasókönyv* (pp. 7–19). Budapest: ELTE BTK Fonetikai Tanszék.

Barthel, B. (2022). Az értelmileg akadályozottak pedagógia szakirány múltja, jelene, jövője. In B. Barthel (Ed.), *Konferenciakötet Dr. Hatos Gyula 85. születésnapja tiszteletére* (pp. 15–25). Budapest: ELTE Eötvös Kiadó.

Beckman, M., & Venditti, J. (2010). Tone and intonation. In W. Hardcastle, J. Laver, & F. Gibbon (Eds.), *The handbook of phonetic sciences. Second edition* (pp. 603–652). Oxford: Wiley–Blackwell Publishing.

BNO-10 (1995). *A betegségek és az egészséggel kapcsolatos problémák nemzetközi statisztikai osztályozása. 10. revízió. 1. kötet*. Budapest: Népjóléti Minisztérium.

BNO-11 (2022). *Mentális zavarok*. Budapest: Animula Kiadó.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (version 6.0.43. URL: `http://www.fon.hum.uva.nl/praat/download_win.html` a letöltés ideje: 2016. szeptember.

Bárth, J., Bodó, C., Deme, A., Markó, A., Rónay, Z., & Szabó, G. (2021). *Kutatásetika. Online tananyag*. Budapest: ELTE.

Bóna, J., Imre, A., Markó, A., Váradi, V., & Gósy, M. (2014). Gabi – gyermeknyelvi beszédadatbázis és információtár. *Beszédkutatás*, (pp. 246–251).

Böhm, L. (1961). *Zenei műszótár: magyarázatokkal, kottapéldákkal, táblázatokkal és hangjegyírás-útmutatóval. Bővített, átdolgozott kiadás*. Budapest: Zeneműkiadó Vállalat.

Csákvári, J., & Mészáros, A. (2012). Értelmi fogyatékos (intellektuális képességzavarral élő) gyermekek, tanulók komplex vizsgálatának diagnosztikus protokollja. In A. Torda (Ed.), *Diagnosztikai kézikönyv* (pp. 1–83). Budapest: Educatio.

Fletcher, J. (2010). The prosody of speech: Timing and rythm. In W. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (pp. 521–602). Oxford: Wiley–Blackwell Publishing. (2nd ed.).

Gráczi, T., Krepsz, V., Markó, A., Huszár, A., & Száraz, B. (2019). Az f0-jellemzők felolvasásban és spontán beszédben. *Alkalmazott Nyelvtudomány*, *19*, 1–16. URL: `http://alkalmazottnyelvtudomany.hu/wordpress/wp-content/uploads/Graczi_tan.pdf`.

Gósy, M. (2004). *Fonetika, a beszéd tudománya*. Budapest: Osiris Kiadó.

Hatos, G. (2000). Az értelmileg akadályozott gyermekek az óvodában és az iskolában. In S. Illyés (Ed.), *Gyógypedagógiai alapismeretek* (pp. 409–428). Budapest: ELTE BGGYK.

Hatos, G. (2001). Értelmi akadályozottság. szócikk. In Z. Mesterházi (Ed.), *Gyógypedagógiai lexikon*. Budapest: ELTE Bárczi Gusztáv Gyógypedagógiai Főiskolai Kar.

Hatos, G. (2015). Az értelmi akadályozottság értelmezésének változásai. *Pedagógiatörténeti Szemle*, *1*, 1–11. URL: `http://www.jgypk.hu/pedtort/wp-content/uploads/2014/12/2015_01.pdf`.

Hodapp, R., & Dyckens, E. (2003). Mental retardation (intellectual disabilities. In E. Mash, & R. Barkley (Eds.), *Child Psychopatology* (pp. 486–519). New York: The Guilford Press. (2nd ed.).

Huszár, A. (2019). A karakterábrázolás alapfrekvencia-jellemzői. In Z. Ludányi, & T. Gráczi (Eds.), *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2019. XIII. Alkalmazott Nyelvészeti Doktoranduszkonferencia* (pp. 59–71). Budapest: MTA Nyelvtudományi Intézet.

ICD-11 (2019). International Classification of Diseases. 11th Revision. the global standard for diagnostic health information. URL: `https://icd.who.int/en/` a letöltés ideje: 2021. július 29.

Jankovics, J. (2019). Alaphangjellemzők vizsgálata enyhe és középsúlyos értelmi fogyatékkal élő felnőttek beszédében. *Beszédkutatás*, (pp. 314–330).

Lee, M., Thorpe, J., & Verhoeven, J. (2009). Intonation and phonation in young adults with down syndrome. *Journal of Voice*, *23*, 82–87.

Lukács, A., & Kas, B. (2014). Nyelvelsajátítás és értelmi fogyatékosság. In C. Pléh, & A. Lukács (Eds.), *Pszicholingvisztika 1–2: Magyar pszicholingvisztikai kézikönyv* (pp. 1383–1404). Budapest: Akadémiai Kiadó.

Lányiné Engelmayer, A. (2017). *Intellektuális képességzavar és pszichés fejlődés. Második, átdolgozott és bővített kiadás*. Budapest: Medicina Könyvkiadó.

Markó, A. (2015). *A spontán beszéd prozódiai szerkezete. Időzítés és beszéddallam*. Budapest: Akadémiai Kiadó.

Markó, A. (2017). Hangtan. In G. Nagy (Ed.), *Nyelvtan* (pp. 75–206). Budapest: Osiris Kiadó.

Markó, A., & Bóna, J. (2012). Eltérő beszédmódok intonációs sajátosságai fiatal és idős korban. In G. Balázs, & A. Veszelszki (Eds.), *Nyelv és kultúra, kulturális nyelvészet* (pp. 253–258). Budapest: Magyar Szemiotikai Társaság.

Markó, A., Huszár, A., Krepsz, V., & E, G. T. (2021). Az alapfrekvencia jellemzőinek longitudinális összevetése felnőtt beszélők felolvasásában. *Beszédtudomány*, *2*, 99–134.

Mesterházi, Z. (2001). Oligofrénpedagógia. tanulásban akadályozottak. tanulásban akadályozottak gyógypedagógiája. In Z. Mesterházi (Ed.), *Gyógypedagógiai lexikon* (pp. 156–157). Budapest: ELTE BGGYK.

Mesterházi, Z., & Szekeres, A. (Eds.) (2021). *A nehezen tanuló gyermekek iskolai nevelése. Egyetemi tankönyv a Gyógypedagógia szak Tanulásban akadályozottak pedagógiája szakirány számára. 2., javított kiadás*. Budapest: ELTE BGGYK.

Radványi, K. (2005). A kommunikáció és a beszéd fejlesztése az értelmileg akadályozott gyermekeknél. In I. Varga (Ed.), *Speciális didaktika I. Az értelmi akadályozottsággal élő gyermekek tanítása* (pp. 28–69). Szeged: Szegedi Tudományegyetem, Juhász Gyula Tanárképző Főiskolai Kar.

Raj, A., Gupta, B., Chowdhury, A., & Chadha, S. (2010). A study of voice changes in various phases of menstrual cycle and in postmenopausal women. *Journal of Voice*, *24*, 363–368.

Seifpanahi, S., Bakhtiar, M., & Salmalian, T. (2011). Objective vocal parameters in farsi-speaking adults with down syndrome. *Folia Phoniatrica et Logopaedica*, *63*, 72–76.

Skarnitzl, R., & Vaňková, J. (2017). Fundamental frequency statistics for male speakers of common czech. *Acta Universitatis Carolinae Philologica*, *3*, 7–17.

Szekeres, G. (2018). Pszichiátriai zavarok osztályozása és differenciáldiagnosztikája. URL: `https://semmelweis.hu/pszichiatria/files/2018/11/Pszichi%C3%A1triai-zavarok-oszt%C3%A1lyoz%C3%A1sa-%C3%A9s-differenci%C3%A1ldiagnosztik%C3%A1ja.pdf` október 17.

Tóth, A. (2017). *A spontán beszéd a nem és az életkor függvényében gyermek- és fiatal felnőttkorban. PhD-értekezés*. Budapest: ELTE BTK.

Towards decoding brain activity during passive listening of speech

Milán András Fodor[1], Tamás Gábor Csapó[2], Frigyes Viktor Arthur[2]

[1] *Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics,*
[2] *Department of Telecommunications and Artificial Intelligence Faculty of Electrical Engineering and Informatics Budapest University of Technology and Economics*

Abstract

The aim of the study is to investigate the complex mechanisms of speech perception and ultimately decode the electrical changes in the brain accruing while listening to speech. We attempt to decode heard speech from intracranial electroencephalographic (iEEG) data using deep learning methods. The goal is to aid the advancement of brain-computer interface (BCI) technology for speech synthesis, and, hopefully, to provide an additional perspective on the cognitive processes of speech perception.

This approach diverges from the conventional focus on speech production and instead chooses to investigate neural representations of perceived speech. This angle opened up a complex perspective, potentially allowing us to study more sophisticated neural patterns. Leveraging the power of deep learning models, the research aimed to establish a connection between these intricate neural activities and the corresponding speech sounds.

Despite the approach not having achieved a breakthrough yet, the research sheds light on the potential of decoding neural activity during speech perception. Our current efforts can serve as a foundation, and we are optimistic about the potential of expanding and improving upon this work to move closer towards more advanced BCIs, better understanding of processes underlying perceived speech and its relation to spoken speech.

Keywords: BCI, speech synthesis, deep learning

1. INTRODUCTION

1.1. Brain-Computer Interfaces and Deep Learning

Brain-Computer Interfaces (BCIs) offer an exciting direction for direct communication between the human brain and external devices (Birbaumer, 2006). Originally developed to assist individuals with neuro-motor disorders, BCIs have

Email addresses: milanfodor@gmail.com (Milán András Fodor), hello@victorarthur.com (Frigyes Viktor Arthur)

the potential to revolutionize a wide range of fields, including communication and rehabilitative technologies (Luo et al., 2023).

Recent advancements in deep learning have enabled considerable improvements in the interpretative power of BCIs. Deep learning, a subset of machine learning, involves artificial neural networks with multiple hidden layers, allowing for complex pattern recognition from high-dimensional data (Schirrmeister et al., 2017; Bashivan et al., 2016). As these deep learning techniques become more sophisticated, their application in BCIs is broadening, particularly in the field of communication BCIs, where one of the main goals is to reconstructing intelligible speech from neural activity (Akbari et al., 2019b). However, significant challenges remain, particularly in less explored areas such as exploring the passive side of communication by decoding perceived speech, which is the primary focus of our research.

1.2. The cognitive background of listened and spoken speech

The human speech process, both in speaking and listening, involves a multitude of complex cognitive processes. Neural signals generated during these processes hold rich information, which, if decoded successfully, could significantly enhance BCI technology for speech synthesis (Hickok et al., 2014; Pulvermüller et al., 2006).

Speech perception encompasses numerous processes such as acoustic analysis, phonetic and phonological processing, lexical access, and semantic comprehension (Pei et al., 2011; Herff et al., 2015). These processes are interconnected, often occurring in parallel, which leads to intricate neural representations of perceived speech within the brain (Brandmeyer et al., 2013; Mesgarani et al., 2014).

Research into speech perception has revealed the involvement of several key brain regions. The superior temporal gyrus (STG) and the posterior superior temporal sulcus (pSTS) are particularly integral for processing acoustic features and phonetic components of speech (Mesgarani et al., 2014; Okada et al., 2010). These areas respond to various speech sounds and their characteristics, and their

activation patterns often mirror the spectro-temporal dynamics of the incoming speech signal.

Beyond the acoustic-phonetic level, speech comprehension involves additional cognitive stages such as lexical access and semantic comprehension, which are associated with other brain regions. Wernicke's area, situated in the posterior part of the superior temporal gyrus, plays a significant role in understanding spoken language, linking the sound of speech to meaning (Price, 2012).

Interestingly, the perception of speech also engages brain regions traditionally associated with speech production. For instance, Broca's area, known for its role in speech production, also plays a part in speech perception, particularly when listeners are anticipating or predicting upcoming speech sounds (Friederici, 2011). Similarly, activity in motor-related areas like the motor cortex and the cerebellum has also been observed during speech perception, potentially reflecting the listeners' internal simulation or mirroring of the speaker's articulatory movements (Eichert et al., 2020; Lotte et al., 2018).

Key areas of the brain involved in speech perception are highlighted in Figure 1. These complex cognitive processes and their associated neural representations present both a challenge and an opportunity for BCI technology. Our research seeks to decode these intricate neural activities during speech perception to aid the advancement of BCI systems for speech synthesis, potentially enabling more naturalistic, communication-focused BCI technology.

1.3. Speech Synthesis from neural activity

Speech synthesis, the artificial production of human speech, is a rapidly evolving field that has undergone substantial advancements, particularly with the incorporation of deep learning and neural network methodologies (Shen et al., 2016; Oord et al., 2016) next to regression-based approaches. These technological advancements have not only enhanced the intelligibility, naturalness, and expressivity of synthetic speech, but also allowed for the integration of complex neural data as an input source.

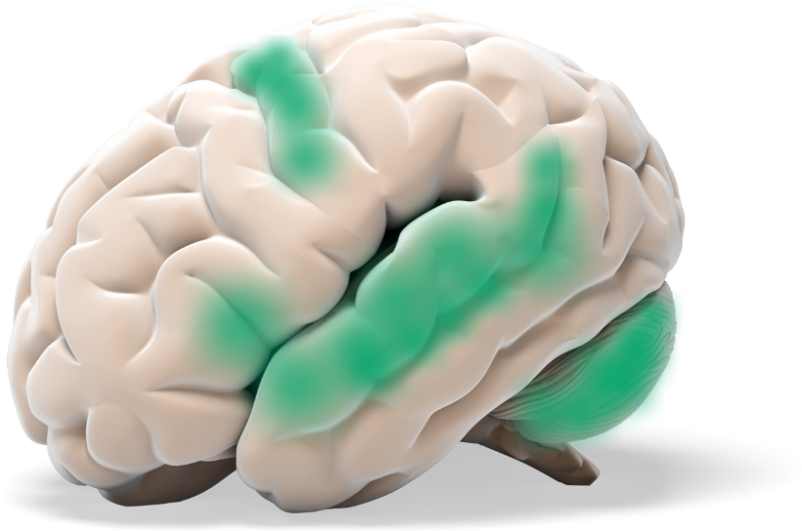


Figure 1: Broca's area, the motor cortex, the cerebellum, Wernicke's area, and the superior temporal gyrus, posterior superior temporal sulcus highlighted as important areas of the brain regarding speech. Figure based on (Guenther, 2006; Hickok & Poeppel, 2007; Von Kriegstein et al., 2010; Hein & Knight, 2008) .

Both neural network-based methods and traditional regression-based approaches, like those presented by Pasley et al. (2012), have distinct advantages and disadvantages. Neural networks, particularly deep learning models, excel in handling complex, non-linear relationships in data, which can be crucial for accurately modeling the intricate patterns in auditory signals. They often achieve higher accuracy and can generalize better to new, unseen data. However, these models require large amounts of training data and substantial computational resources. On the other hand, traditional regression-based methods, while sometimes less accurate in complex scenarios, offer greater transparency and can be more interpretable. They are typically simpler to implement and require less computational power, making them more accessible for smaller-scale studies or applications with limited resources. Additionally, traditional methods can be more robust to overfitting when dealing with small datasets. Therefore, the choice between these approaches should be guided by the specific requirements and constraints of the speech reconstruction task at hand.

A key challenge lies in the adaptation of speech synthesis systems to real-world environments. Everyday communication often takes place amidst background noise, room reverberations, or with multiple speakers, conditions that can considerably impair the performance of conventional speech synthesis systems (Godoy et al., 2018). Developing algorithms capable of effectively synthesizing speech under such challenging conditions is a critical area of ongoing research.

As the field of speech synthesis evolves, there is an emerging interest in faster, more accurate and more naturalistic approaches. One possible avenue to get closer to this goal is could leveraging BCI to decode heard speech from neural activity (Pei et al., 2011; Brandmeyer et al., 2013). This innovative approach would allow for the synthesis of speech that the user hears, rather the user's input. When further developed, this method could potentially assist in instantly storing information we consciously perceive. Additionally, it may enhance overt speech synthesis, as we hear our own speech.

2. RESEARCH OBJECTIVE

This study seeks to employ a dataset of iEEG recordings collected during passive listening of speech. Utilizing deep learning algorithms, we aim to construct a model that aspires to decode the heard speech from these neural activities. By doing so, we anticipate contributing to advancements in BCI technology and enhancing our theoretical understanding of cognitive speech processing. The adoption of these advanced computational techniques could enable us to unravel the intricate neural representations of perceived speech, and these insights pave the way for advancements in BCI systems (Schirrmeister et al., 2017; Bashivan et al., 2016). The adoption of iEEG data is particularly advantageous due to its high signal-to-noise ratio, enhanced spatial resolution, and ability to capture a broad range of frequency bands, making it highly suitable for speech decoding (Halgren et al., 2019; Crone et al., 2001).

We articulate our decision to situate our research in the context of passively listened speech. While existing BCI research often emphasizes speech production, the area of listened speech remains relatively unexplored. We propose that this angle harbors untapped potential, promising novel insights into the cognitive dimensions of speech processing and a fresh angle for speech decoding efforts (Pei et al., 2011; Brandmeyer et al., 2013). There may be certain disabilities where brain damage affects auditory processing in such a way that, although neural representations of heard sounds are present, they are not fully perceived by the individual. While recording sound is an obvious solution, this technology, when fully developed, could offer an instantaneous alternative that might only record sounds the individual focuses on. It also has implications for overt speech decoding, since we hear our own words, when speaking.

In conclusion, this study represents an effort to elucidate the complex relationship between speech perception and production, and their neural representations, while advancing the development of naturalistic, communication-focused BCI technology.

.

3. METHODS

3.1. Dataset

This research uses the 'Open multimodal iEEG-fMRI dataset' (Berezutskaya et al., 2022), a publicly available resource that combines iEEG with fMRI data. The high spatial and temporal resolution of the dataset offers detailed insights into speech and language processing.

3.1.1. Participants

The dataset contains data from fifty-one Dutch epilepsy patients undergoing diagnostic procedures at the University Medical Center Utrecht. The study was approved by the Medical Ethical Committee of the University Medical Center Utrecht, in line with the Declaration of Helsinki (2013). There were 32 female and 19 male participants. The ages varied, with a mean of 25 years and a standard deviation of 15 years. For patients under 18 years of age, consent was obtained from their parents or legal guardian. From the 51 patients, 16 provided written consent for their clinical data to be used for research. From these, we later chose the best suited ones (see 3.4) based on correlations with the speech envelopes and their electrode placements. This resulted in four "prime subjects" (s43, s46, s55, s60), one "ideal electrode placement" subject (s38) and one reference subject (s13) with not ideal electrode placements. For these six participants whose data were used in the study, the ages ranged between 14 and 42 years, with a mean age of 26 years and a standard deviation of 11.94 years. The group comprised 4 females and 2 males.

3.1.2. Experimental procedures

The patients participated in two main types of experiments: movie-watching and resting state. The movie-watching experiment, which involved the patient watching a short film, was part of the standard battery of clinical tasks for presurgical functional language mapping. The resting state experiment, which required the patients to rest for three minutes, was conducted for research purposes. For those patients who did not participate in a separate resting state

task, a 3-minute 'natural rest' period was selected from their 24/7 clinical iEEG recordings.

3.1.3. Stimuli

The stimulus for the movie-watching experiment was a 6.5-minute short movie composed of fragments from "Pippi on the Run" (Pårymmen med Pippi Långstrump, 1970). The movie was edited to form a coherent plot and consisted of 13 interleaved blocks of speech and music, each 30 seconds long. The movie was originally in Swedish but dubbed into Dutch. Detailed annotations of the audio and video content of the movie stimulus can be found in the dataset. The annotation includes the marking of 129 unique visual concepts. Importantly for our study, it also contains the onsets and offsets of several language features such as phonemes, syllables, words, clauses, and sentences.

3.1.4. Electrode Implantation

Electrode types varied based on clinical requirements. Forty-six patients had ECoG grids with 48 to 128 contact points. Six patients had high-density ECoG grids with 32 to 128 contact points. Sixteen patients had sEEG electrodes with 4 to 173 contact points. Most electrodes covered perisylvian areas and frontal and motor cortices.

3.1.5. Data Acquisition

Intracranial EEG (iEEG) data were acquired using a 128-channel recording system (Micromed, Treviso, Italy) during the experimental tasks. The majority of patients' data were sampled at 512 Hz and filtered at 0.15–134.4 Hz, while in some cases, the data were sampled at 2048 Hz and filtered at 0.3–500 Hz. An external reference electrode was used for signal referencing, typically placed on the mastoid part of the temporal bone. Besides, six patients had their HD ECoG data recorded either simultaneously with the clinical channels or in separate sessions.

3.2. Data availability

The dataset can be accessed at: `https://openneuro.org/datasets/ds003688`. To maintain confidentiality, identifiable information and individual MRI scans have been removed. The order of subjects in the dataset has been randomized to further ensure anonymity.

3.3. Data validation

Preprocessing of the iEEG data was carried out using MNE-Python (`https://mne.tools`).

To ensure data quality, the subjects' neural activity during speech and music blocks was compared by the team behind the dataset. (Berezutskaya et al., 2022).

3.4. Prime subjects

To facilitate the most effective and meaningful analysis for this study, we utilized a rigorous selection process for the subjects, oriented primarily around a key determinant: the level of correlation that each subject demonstrated with the speech envelope during the movie, as noted by the team who compiled the dataset (Berezutskaya et al., 2022).

Additionally, our selection strategy was influenced by the need to optimize our limited time and GPU resources. This subject selection methodology stemmed from the hypothesis that individuals whose neural activity closely mirrored the dynamic ebb and flow of the speech envelope would be ideal candidates for this study. From the pool of potential subjects, four individuals were eventually selected, as shown in Fig. 2.

These participants displayed notably high correlation values, likely stemming from the placement of intracranial electrodes covering key areas associated with speech perception and production, including the Broca's area, the motor cortex, the cerebellum, Wernicke's area, and the superior temporal gyrus. The selection process ensured the recruitment of subjects whose neural responses would yield

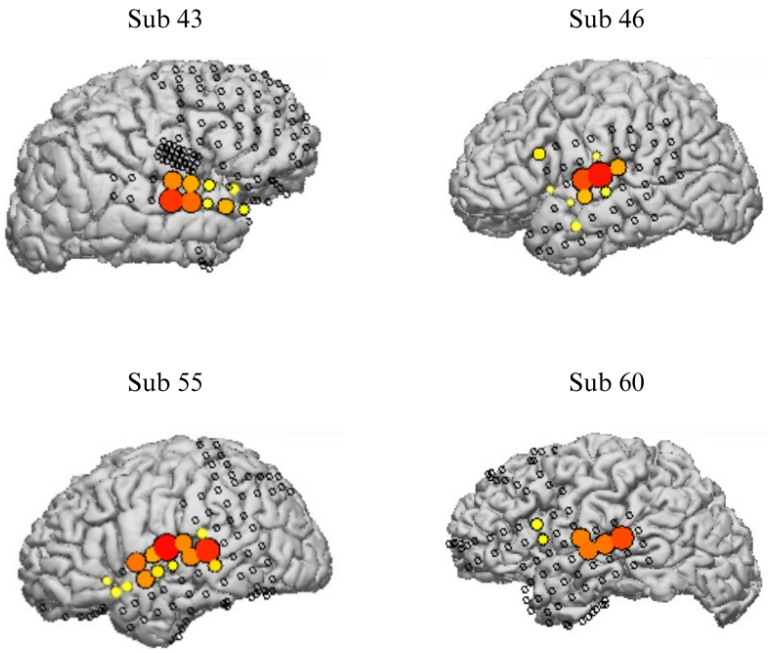


Figure 2: The four subjects with the highest correlation with the speech envelope. From Berezutskaya et al. (2022).

the richest and most insightful data for decoding and reconstructing speech from neural signals.

In addition to data-driven selection, manual selection ensured coverage of essential brain regions. In particular, Subject 38 was chosen for their exceptional coverage of electrodes over the motor cortex, the Broca's area, and the superior temporal gyrus — see Fig. 3. This unique electrode placement may offer a unique opportunity for more accurate and nuanced speech reconstructions.

Finally, we chose subject 13 as a reference because their electrode placements were less ideal, according to the literature.

By employing both quantitative and qualitative selection criteria, we identified the subjects who were most likely to contribute valuable data to the study.

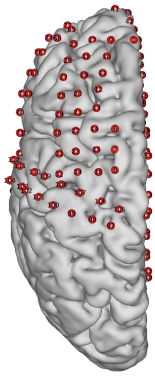


Figure 3: The electrode positions for Subject 38. Extracted from the Open multimodal iEEG-fMRI dataset (Berezutskaya et al., 2022)

3.5. Preparing data for training

3.5.1. Audio and lag correction

The audio data is first loaded using the librosa library. Figure 4 provides a visual representation of the cross-correlation between the electrode's high-frequency band signal and the sound envelope. Different colors correspond to different 30-second speech blocks. The observed average delay is approximately 150 milliseconds, which we accounted for by shifting the audio backwards by 150 milliseconds, thereby enhancing the alignment of the decoded speech with the original auditory stimulus.

For mel-spectrogram estimation from speech, 80 bins were used using librosa mel-filter defaults. Essential STFT parameters were set, including a filter length of 1024, a hop length of 10 ms, and a mel frequency range spanning from 0 to 8000 Hz, 80 frequency bins. The sampling rate was 22050 Hz.

3.5.2. Filtering and cropping only speech segments of iEEG

All subjects' brain signal data were sampled at 512 Hz. Initially, 'ECoG' and 'sEEG' type channels were selected, and defective channels were removed. A notch filter was applied to counter line noise at 50 Hz and its harmonics.

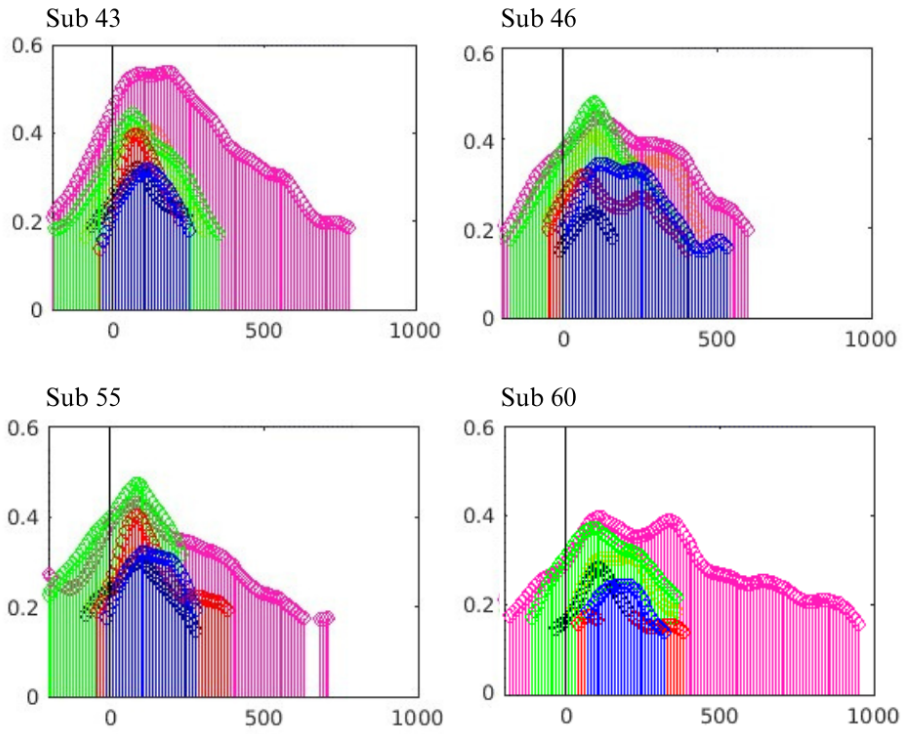


Figure 4: Lagplots of the cross-correlation of the electrode's high-frequency band signal and the sound envelope (Berezutskaya et al., 2022).

The data was then re-referenced using the technique known as common average referencing (CAR).

We extracted EEG data corresponding to the segments where speech was present in the movie. This was achieved by selectively slicing the raw_car data (the preprocessed EEG data) and the mel_data (mel-spectrogram estimated from the speech stimuli) based on provided annotations, thereby focusing the analysis on the brain's response to auditory speech stimuli. The result of this procedure was a refined set of data (raw_car_cut and mel_data_cut), encapsulating the EEG responses to speech stimuli, thus enhancing the relevance and accuracy of the subsequent deep learning model training.

3.5.3. Feature extraction

In the feature extraction process, we first apply linear detrending to the EEG data, effectively removing linear trends and reducing potential artifacts. The data is then segmented into overlapping windows, each defined by a specific length (0.05 ms) and shift (0.01 ms). Within these windows, we perform band-pass filtering, specifically targeting the 1–120 Hz frequency range, as everything from a delta to high gamma frequencies are relevant to speech, when we are also interested in speech perception (Lopez-Bernal et al., 2022). Subsequently, the Hilbert transform is applied to the filtered data to derive the analytic signal, enabling us to calculate the amplitude envelope. The final step involves computing the mean amplitude of this envelope for each window across all EEG channels. The resulting output is a 2D feature matrix, where each row represents a time window and each column corresponds to an EEG channel. This matrix encapsulates the mean amplitude of the target frequency band for each window and channel, providing a concise representation of the EEG data for further analysis.

3.6. Deep learning training

Deep learning, renowned for its efficacy in abstract pattern extraction from extensive high-dimensional data sets, is a natural fit for parsing intracranial electroencephalogram (iEEG) data. Notably, its prowess in related tasks motivated its selection.

We utilized Fully Connected Deep Neural Networks (Fc-DNNs) and 2D Convolutional Neural Networks (2D-CNNs) for this research, chosen after assessing their inherent properties and suitability for predicting mel-spectrograms from iEEG data. Figure 5 presents a simplified illustration of the transformation process: iEEG inputs being fed into the DNN architecture, and subsequently producing mel-spectrogram outputs. The selection process was iterative, involving comprehensive evaluation of multiple model architectures, training strategies, and optimization techniques. The configurations delivering optimal performance were chosen for the final models.

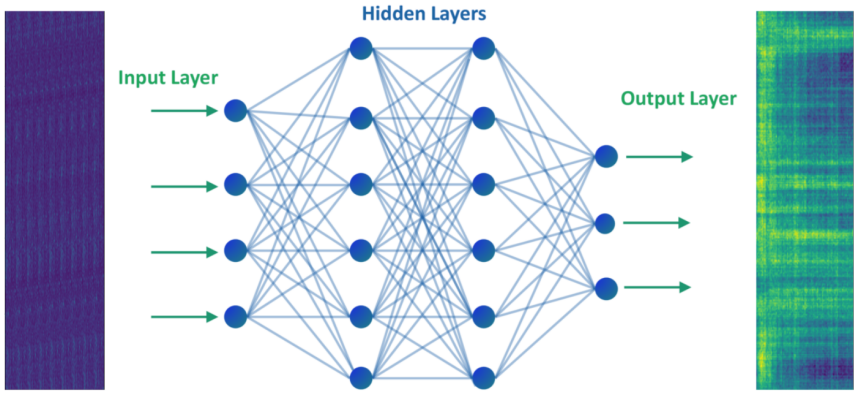


Figure 5: Visual representation of the iEEG input and mel-spectrogram output of the DNN.

To ensure transparency and repeatability, all code, files, and scripts utilized for data preprocessing, model training, and result analysis are publicly shared at: (WARNING: clicking this link might reveal author identities)

`https://github.com/MILANIUSZ/speech2brain2speech`.

The training process employed an RTX 3070 GPU and an AMD Ryzen 5 3600 processor, with the environment set up using Docker and the public image "thegeeksdiary/tensorflow-jupyter-gpu". This setup ensured efficient hardware utilization for model training and evaluation.

3.6.1. Fc-DNN

Fully Connected Deep Neural Networks, also known as Multilayer Perceptrons (MLPs), are versatile neural networks used extensively for regression tasks. This study employed an Fc-DNN model with one hidden layer of 3000 neurons. This configuration was systematically chosen after multiple iterations to ensure optimal performance while avoiding overfitting, as increasing model complexity didn't substantially improve accuracy but led to overfitting.

The Rectified Linear Unit (ReLU) was used as the activation function for the input layer due to its ability to handle the vanishing gradient problem, and

a linear activation function for the output layer, fitting for a regression task. Adam optimizer was used due to its efficiency.

Data was partitioned into training, validation, and test sets (80%, 10%, 10%, respectively). EEG and mel-spectrogram data were scaled using MinMaxScaler and StandardScaler, respectively. The Fc-DNN model was built using Keras with a hidden layer of 3000 neurons (ReLU activation) and an output layer of 80 neurons (linear activation). The model was compiled with Mean Squared Error as the loss function, trained for a maximum of 50 epochs with a batch size of 32, with early stopping for overfitting prevention.

3.6.2. 2D-CNN

Two-Dimensional Convolutional Neural Networks excel at grid-like data processing tasks. For this study, a 2D-CNN was used to process the spectrogram data obtained from EEG recordings, with an 80% allocation for the training set (similarly to Fc-DNN). Both the input and output data were normalized using the mean and standard deviation from the training set.

The 2D-CNN model architecture consisted of three convolutional layers with 'swish' activation function and dropout layers to prevent overfitting. Padding was applied to the input so that the output has the same length as the original input when the stride is 1. The model included a max pooling layer for dimensionality reduction, followed by a flatten layer and a dense layer with 'swish' activation. The output layer was a dense layer with a linear activation function, aligned with the training spectrogram shape. The model was compiled with the 'Adam' optimizer and the 'mean squared error' loss function, with training conducted over 100 epochs with a batch size of 128. Overfitting prevention was handled through early stopping and learning rate adjustment.

After training, the predicted spectrogram was inverse transformed to its original scale and saved for subsequent evaluation.

For detailed training parameters of the neural network, please refer to the supplementary materials and our GitHub repository.(WARNING: clicking this link might reveal author identities) `https://github.com/MILANIUSZ/speech2brain2speech`.

3.6.3. Evaluation methods

The performance of the Fc-DNN and the 2D-CNN was evaluated using Mean Squared Error (MSE) as the measure, with lower MSE indicating better mel-spectrogram prediction from EEG signals. Training was conducted 10 times for each model.

For qualitative assessment, the predicted, scaled mel-spectrograms were plotted in comparison to the orginial test mel-spectrograms. The discrepancies offered insights into the models' performance.

Subject 13 was selected as a baseline because the implanted electrodes primarily covered areas on the occipital lobe. The occipital lobe is hypothesized to have fewer associations with the cognitive processes involved in perceived speech. Therefore, this choice provides a meaningful reference point for our analysis.

Additionally, an informal auditory evaluation was done by the first author. The reconstructed mel-spectrograms were converted back into audio signals using using the Griffin-Lim algorithm, implemented through the librosa library in Python, allowing aural comparisons of original and synthesized signals, revealing potential model shortcomings.

4. RESULTS

4.1. Fully-connected deep neural network

The Fc-DNN was trained on the data from 6 subjects (four "prime" subjects (s43, s46, s55, s60), one "ideal" electrode placement subject (s38), and one "not ideal" electrode placement subject (s13)), and the performance of the model for each subject is summarized in Table 1. The table presents the best training loss and validation mean squared error (MSE) achieved for each subject.

The training loss values represent how well the model is able to predict the mel-spectrogram data from the EEG signals during training. Lower training loss indicates a better fit of the model to the training data. The validation MSE, on

Subject	Best Training Loss	Best Validation MSE
38	0.0210	0.6982
43	0.0336	0.7381
46	0.2643	0.7923
55	0.2015	0.7210
60	0.3900	0.6520
13	0.3256	0.8052

Table 1: Performance of the Fc-DNN for each subject.

the other hand, provides a measure of the model's performance on unseen data, with lower MSE values representing better generalization performance.

From Table 1, it can be observed that the model achieved the lowest training loss with subject 43, indicating the model was able to fit the training data most effectively for this subject. On the other hand, the model demonstrated the best generalization performance on unseen data with subject 60, as indicated by the lowest validation MSE.

4.2. Two-dimensional convolutional neural network

Just like the Fc-DNN, the 2D-CNN model was trained on the data from the six different subjects. The performance metrics for the 2D-CNN, specifically the best training loss and the validation mean squared error (MSE) for each subject, are outlined in Table 2.

The 2D-CNN model performance is evaluated using the same metrics as the Fc-DNN model: the training loss, the validation MSE and informal listening to the synthesized audio. In this case, subject 46 achieved the lowest validation MSE.

4.3. Mel-spectrogram demonstration samples

Fig. 6 shows an original speech stimuli sample (top) and those mel-spectrograms generated from iEEG input by the 2D-CNN (middle) and Fc-DNN (bottom) networks. Based on visual inspection, we can see that the result of 2D-CNN

Subject	Best Training Loss	Best Validation MSE
38	0.4121	0.7023
43	0.5321	0.7326
46	0.8043	0.6920
55	0.9605	0.7879
60	0.9039	0.7922
13	0.9039	0.8781

Table 2: Performance of the 2D-CNN for each subject.

is oversmoothed, whereas the FC-DNN was able to generate more "realistic" patterns. However, the similarity between the original audio stimuli and the predicted spectrogram is still not satisfactory.

Fig. 7 compares the iEEG-to-speech results of two subjects. Based on this, the results of subject 55 seem to be more realistic, probably because his electrodes are located at more relevant areas of the brain.

4.4. Audio synthesis

Both models' synthesized audio underwent informal human evaluation by the first author, to assess its quality and intelligibility. While the speech wasn't comprehensible, in some cases, the model captured some auditory elements, such as the silences. However, the accurate reconstruction of speech content remains a huge challenge.

5. Discussion and way forward

5.1. Speech decoding

The selected deep learning architectures, Fc-DNN, and the 2D-CNN, especially in light of the limited training data, have demonstrated the potential of the approach by finding patterns in perceived speech's neural activity indicated by the reduction of test loss in a consistent manner.

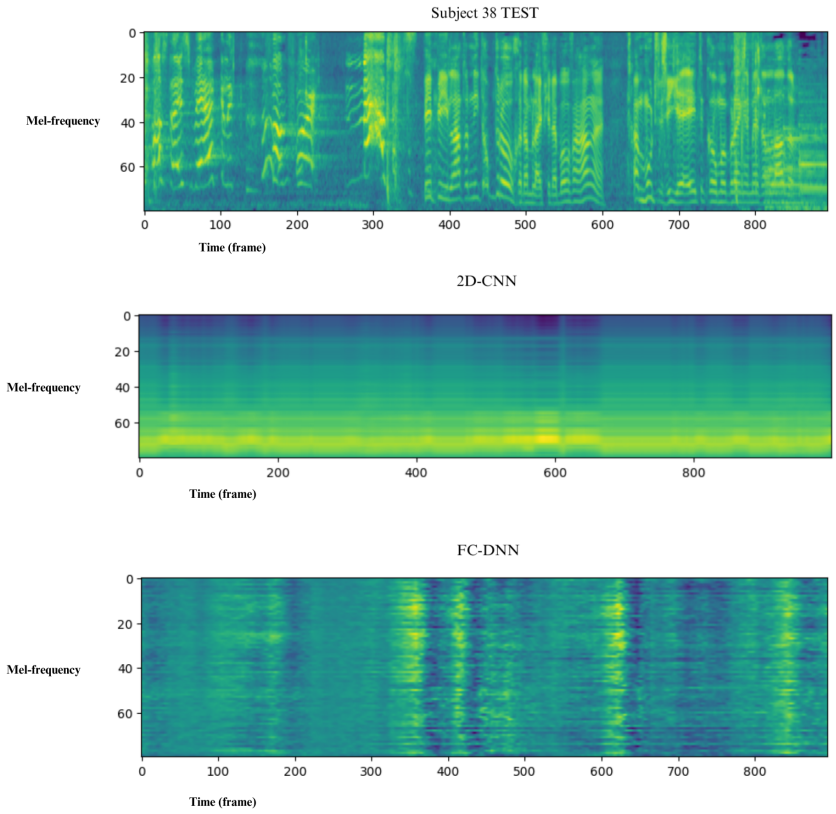


Figure 6: Mel-spectograms for subject 38.

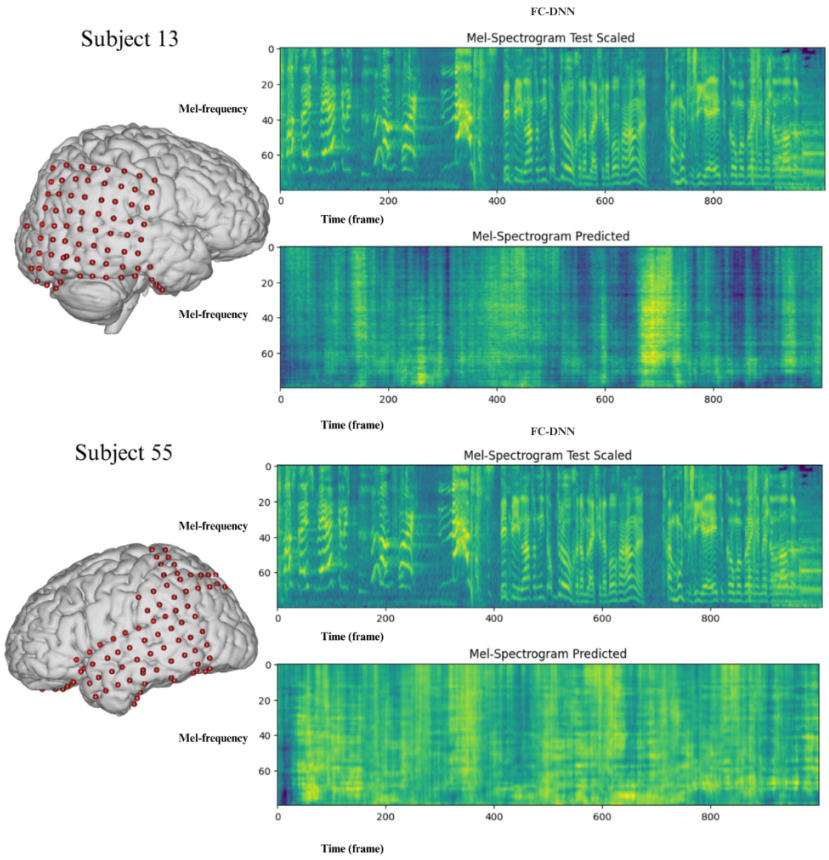


Figure 7: Electrode placement and mel-spectrograms (based on the FC-DNN) comparison for subjects 13 and 55.

Despite the demonstrated potential of this approach, significant challenges remain with the methodology. A noteworthy issue in the current study is the difficulty in achieving satisfactory accuracy levels for both validation and test sets concurrently, despite multiple training iterations. The resulting mel-spectrograms, although indicating some pattern recognition and learning within the data, failed to provide a realistic spectrogram. Consequently, the audibility and clarity of the synthesized speech generated from these mel-spectrograms were low.

Previous studies, such as Anumanchipalli et al. (2019), which synthesized intelligible speech from neural activity recorded during active reading tasks, reported successful outcomes. However, our study primarily relies on passive tasks, which don't generate robust motor and auditory brain responses like active tasks. Therefore, the differences between the results obtained in their study and ours can be attributed to the contrasting nature of the tasks involved.

At the same time, our study resonates with other research, such as Akbari et al. (2019a), which aimed to decode spectrograms from brain activity recorded during passive listening tasks. Their findings, which reported challenges in generating realistic spectrograms and clear synthesized speech, echo the issues encountered in our study.

However, we must exercise caution when comparing these studies due to methodological variations such as data collection techniques, preprocessing steps, model architectures, and evaluation metrics. For instance, some studies might employ invasive electrocorticography (ECoG) for data collection, resulting in high-resolution data, while others might utilize non-invasive methods like EEG or fMRI.

Despite these variations, the overall trend underscores the complexity of speech decoding, especially during passive listening scenarios, and highlights the need for more careful data preparation and/or significant technological advancements for reliable synthesis of clear speech from such brain activity.

5.2. Cognitive conclusions

Upon comparing the accuracy and spectrograms, it seems that the patients with electrode placements, which were hypothesized to yield better results based on the literature, shown in Fig. 2, do indeed show improved outcomes. As illustrated in Fig. 7, a disparity in performance can be observed between subject 13 (for which we achieved validation MSE of 0.805 with the FC-DNN and 0.878 with the CNN), serving as our baseline, and subject 55. The latter's electrode placement is more closely aligned with regions typically associated with speech processing, thereby reinforcing the crucial role of electrode placement in the accurate prediction of perceived speech.

These findings also hint at the possibility of shared characteristics in neural activity during passive listening and spoken speech, which might align with theories such as the 'motor theory of speech perception' (Liberman & Mattingly, 1985; Galantucci et al., 2006), the 'neural reuse' theory (Anderson, 2010) or the role of 'mirror neurons' in speech (Rizzolatti & Sinigaglia, 2008). However, these connections should be interpreted with caution, as our study does not provide definitive evidence for such theories.

5.3. Limitations and future directions

The big limiting factor of our study's success was the alignment of iEEG and audio data. It is challenging, and also amplified by the limitation of the dataset size. Future endeavors should focus on improved synchronization methods, larger, more diverse datasets, and the utilization of more advanced neural network architectures, e.g. transformer-based methods which can better handle temporal misalignment. In addition, including audible speech reproduction scenarios and interpretability techniques for neural networks could offer deeper insights into cognitive processes. While our study focused on intracranial EEG data, future research may consider other modalities like MEG or fMRI for more comprehensive data.

Moreover, an interesting avenue for future work could be the integration of multi-modal data, such as neural activity from various brain regions, and

additional data sources like facial movements, articulatory gestures or visual cues (Gosztolya et al., 2019; Arthur & Csapó, 2021; Csapó et al., 2023). This approach could help enhance the decoding performance and accuracy of speech BCIs.

5.4. Future BCI

The advancement of communication BCI continues, we try to create systems that work more accurately, faster and in a more naturalistic way. However, despite all the advancements in the field, challenges remain. Current neural recording techniques, such as invasive iEEG, offer high resolution but are impractical for widespread use. There is also a demand for even more efficient, speech-specific decoding algorithms, as existing models can require extensive datasets and substantial computational resources. Further, the field might benefit from a deeper understanding of speech processes in the brain.

This study tried to emphasize the potential role of perceived speech in the field. Our current efforts can serve as a foundation, and we are optimistic about the potential to expand and improve upon this work, moving closer to more advanced and effective BCIs.

6. Acknowledgements

We would like to thank the authors of the 'Open Multimodal IEEG-FMRI Dataset' for making the data available.

This research was funded by the National Research, Development and Innovation Office of Hungary (grant nr. NKFIH FK 142163).

References

Akbari, H., Gao, Y., Belkin, M., & Ribeiro, A. (2019a). Towards reconstructing intelligible speech from the human auditory cortex. *Scientific Reports*, *9*, 874.

Akbari, H., Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2019b). Towards reconstructing intelligible speech from the human auditory cortex. *Scientific reports*, *9*, 1–11.

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences*, *33*, 245–266.

Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, *568*, 493–498.

Arthur, F. V., & Csapó, T. G. (2021). Towards a practical lip-to-speech conversion system using deep neural networks and mobile application frontend. *CoRR*, *abs/2104.14467*. URL: https://arxiv.org/abs/2104.14467. arXiv:2104.14467.

Bashivan, P., Rish, I., Yeasin, M., & Codella, N. (2016). Learning representations from eeg with deep recurrent-convolutional neural networks. *International conference on learning representations*, .

Berezutskaya, J., Vansteensel, M. J., Aarnoutse, E. J., Freudenburg, Z. V., Piantoni, G., Branco, M. P., & Ramsey, N. F. (2022). Open multimodal iEEG-fMRI dataset from naturalistic stimulation with a short audiovisual film. *Scientific Data*, *9*. URL: https://doi.org/10.1038/s41597-022-01173-0. doi:10.1038/s41597-022-01173-0.

Birbaumer, N. (2006). Breaking the silence: brain–computer interfaces (bci) for communication and motor control. *Psychophysiology*, *43*, 517–532.

Brandmeyer, A., Farquhar, J. D., McQueen, J. M., & Desain, P. W. (2013). Decoding speech perception by native and non-native speakers using single-trial electrophysiological data. *PLoS ONE*, *8*. doi:10.1371/journal.pone.0068261.

Crone, N. E., Boatman, D., Gordon, B., & Hao, L. (2001). Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology*, *57*, 2045–2053.

Csapó, T. G., Arthur, F. V., Nagy, P., & Boncz, Á. (2023). Towards Ultrasound Tongue Image prediction from EEG during speech production. In *Proc. Interspeech*. Dublin, Ireland.

Eichert, N., Papp, D., Mars, R. B., & Watkins, K. E. (2020). Mapping human laryngeal motor cortex during vocalization. *Cerebral Cortex*, *30*, 6254–6269.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Trends in Cognitive Sciences*, *15*, 459–466. URL: `http://dx.doi.org/10.1016/j.tics.2011.06.004`. doi:`10.1016/j.tics.2011.06.004`.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*, 361–377.

Godoy, M. A., Lopes, M. S., de Freitas, D., & de Araújo, A. (2018). Robust speech recognition: Bridging the gap between human and machine performance. *Expert Systems with Applications*, *103*, 50–60.

Gosztolya, G., Pintér, Á., Tóth, L., Grósz, T., Markó, A., & Csapó, T. G. (2019). Autoencoder-based articulatory-to-acoustic mapping for ultrasound silent speech interfaces. *CoRR*, *abs/1904.05259*. URL: `http://arxiv.org/abs/1904.05259`. `arXiv:1904.05259`.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of communication disorders*, *39*, 350–365.

Halgren, M., Ulbert, I., Bastuji, H., Fabo, D., Eross, L., Rey, M., Devinsky, O., Doyle, W., Mak-McCully, R., Halgren, E., Wittner, L., Chauvel, P., Heit, G., Eskandar, E., Mandell, A., & Cash, S. (2019). The generation and propagation of the human alpha rhythm. *Proceedings of the National Academy of Sciences*, *116*, 23772–23782.

Hein, G., & Knight, R. T. (2008). The superior temporal sulcus is crucial for social communication. *The Superior Temporal Sulcus is Crucial for Social Communication*, *5*, 721–727.

Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G., & Schultz, T. (2015). Brain-to-text: Decoding spoken phrases from phone representations in the brain. *Frontiers in neuroscience*, *9*, 217.

Hickok, G., Houde, J., & Rong, F. (2014). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, *39*, 393–402.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, *8*, 393–402.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Lopez-Bernal, D., Balderas, D., Ponce, P., & Molina, A. (2022). A state-of-the-art review of eeg-based imagined speech decoding. *Frontiers in Human Neuroscience*, *16*, 867281.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for eeg-based brain–computer interfaces: a 10-year update. *Journal of neural engineering*, *15*, 031005.

Luo, S., Rabbani, Q., & Crone, N. E. (2023). Brain-computer interface: applications to speech decoding and synthesis to augment communication. *Neurotherapeutics*, *19*, 263–273.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*, 1006–1010.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J. T., & Hickok, G. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Journal of Cognitive Neuroscience*, *33*, 1–13. URL: `http://dx.doi.org/10.1162/jocn.2010.21506`. doi:`10.1162/jocn.2010.21506`.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, .

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., & Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS biology*, *10*, e1001251.

Pei, X., Leff, A. P., Woollams, A. M., Lambon Ralph, M. A., & Scott, S. K. (2011). Spoken language comprehension—an experimental approach to disordered and normal processing. *Neuropsychologia*, *49*, 811–821.

Price, C. J. (2012). A critical review of the role of the left inferior frontal gyrus in language processing. *Trends in Cognitive Sciences*, *20*, 256–267. URL: http://dx.doi.org/10.1016/j.tics.2012.02.009. doi:10.1016/j.tics.2012.02.009.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*, 7865–7870.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the Brain: How Our Minds Share Actions, Emotions*. doi:10.1093/oso/9780199217984.001.0001.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, *38*, 5391–5420.

Shen, Z., Ma, X., Gao, N., Zhang, H., Yan, C., Zhu, C., Zhang, X., Zhang, J., Zhang, Y., & Liu, Y. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature Neuroscience*, *19*, 1037–1042.

Von Kriegstein, K., Smith, D. R., Patterson, R. D., Kiebel, S. J., & Griffiths, T. D. (2010). How the human brain recognizes speech in the context of changing speakers. *Journal of Neuroscience*, *30*, 629–638.

Revised annotation conventions in Hungarian speech corpora

Katalin Mády[1], Anna Kohári[1], Tekla Etelka Gráczi[1], Péter Mihajlik[1,2]

[1] *HUN–REN Hungarian Research Centre for Linguistics*
[2] *Department of Telecommunications and Artificial Intelligence, Budapest University of Technology and Economics*

Abstract

This technical report presents the revised annotation conventions for one large and two smaller Hungarian speech corpora, the BEA Spoken Language Database, the Akaka Maptask Corpus, and the Budapest Games Corpus. Annotations relying on standard Hungarian orthography rather than actual and partly reduced phonetic realisations make it possible to run both linguistic and phonetic queries on a large amount of data. Since the vast majority of the recordings contain (semi-)spontaneous speech, non-lexical phenomena such as hesitations (filled pauses) and non-verbal events such as laughter are labelled. The frequency of the occurrences of these phenomena is demonstrated on the subset Release 1 of the BEA database on speech samples of 115 speakers. Unsurprisingly, laughter and communicative grunts were more frequent in spontaneous speech when expressed in relative numbers. Hesitations occurred more often in semi-spontaneous speech than in read and spontaneous speech showing that the task demanded a higher cognitive effort from speakers. The majority of questions were found in spontaneous speech since the reading tasks did not include interrogatives.

Keywords: speech database, annotation, read speech, spontaneous speech, discourse

1. Introduction

The development of speech databases has become essential during the last decades in order to assess data from a large number of speakers for various kinds of disciplines connected to speech research. The present study introduces the annotation conventions for one large and two smaller Hungarian speech corpora. The main emphasis lies on BEA –*BEszélt nyelvi Adatbázis*, 'Spoken Language Database' (Gósy, 2012) being the largest available speech database

Email addresses: mady.katalin@nytud.hun-ren.hu (Katalin Mády), kohari.anna@nytud.hun-ren.hu (Anna Kohári), graczi.tekla.etelka@nytud.hun-ren.hu (Tekla Etelka Gráczi), mihajlik.peter@nytud.hun-ren.hu (Péter Mihajlik)

for Hungarian at present. Two smaller task-oriented corpora, the Akaka Map-task Corpus (Molnár et al., 2023) and the Budapest Games Corpus (Mády et al., 2023) have been annotated according to the same guidelines. Audio and annotation files are available for research purposes for free via the website `phon.nytud.hu/voxbox/corpora/databases.html` that is being extended by new subsets of the corpora as annotations become available.

2. Basics on the BEA database

BEA is a large speech corpus containing Hungarian speech samples: repetition, spontaneous, semi-spontaneous, and read speech samples (Gósy, 2012). The recordings started in 2007 and lasted until 2017, purporting altogether 472 speakers' speech. The spontaneous speech tasks include an interview about the speaker's life (work/studies, education, hobbies, family, etc.), a quasi-monologue or a dialogue on their opinion about a current public topic, and a discourse with an additional discourse partner on a further current topic, see Table 1. The semi-spontaneous speech tasks include a summary of two short texts that were read aloud either by the experimenter or played to the participant from a recording. The read speech samples were collected by asking the speakers to read aloud 25 sentences of various lengths and a coherent text (title + 13 sentences). In the repetition task, the experimenter read 25 sentences separately, waiting for the speaker to repeat each sentence before moving on to the next one. The altogether eight modules were recorded with all speakers. The sequence of the speech tasks is shown in Table 1. For a detailed description of the recording protocol, see Gósy (2012).

The annotation of the speech samples started shortly after the first recordings were carried out. Annotations were prepared manually, i.e. from scratch, without the help of speech recognition tools. The time-consuming work was carried out with a large group of annotators after careful training. Neuberger et al. (2014) provide an overview on the various formats. The transcription guidelines have been described in several papers (Gósy, 2012; Neuberger et al., 2014). The

speech type	speech task	dynamics	typical order of recording
spontaneous speech	interview	quasi-monologue	2
	opinion on a topic	quasi-monologue or dialogue (depending on the speaker)	3
	discourse	trialogue	6
semi-spontaneous	summarisation of a heard text on plants	quasi-monologue	4
	summarisation of a heard historical anecdote	quasi-monologue	5
reading	reading aloud 25 sentences		7
	reading aloud a text on plants		8
repetition	repetition of 25 sentences		1

Table 1: The seven speech tasks of the BEA database

elaboration of the transcription and annotation methods, their processing and monitoring took place between 2007 and 2019 meaning the joint work of a large number of annotators and researchers.

Several tasks that originally required human resources have received substantial support by machine learning tools – transcribing speech to text was no exception (Bazillon et al., 2008) Due to the increase in computing capacities and the advances in Artificial Intelligence (AI) research, deep neural networks (DNN) have become far more efficient in automatic speech recognition (ASR) than the earlier ML (Maximum Likelihood) techniques (Hinton et al., 2012).

The technical improvement was a motivation to rearrange the transcription process for the speech material of 167 participants (35% of all speakers) for whom no manual annotations had been prepared previously. Along with the largely improved recognition rates (Mihajlik et al., 2022), word and phoneme segmentation tools became available for the team working on the BEA database. These tools required a slightly different input format of the annotated text than the MAUS segmentation system (Schiel, 2004) that had been used previously for the manually annotated recordings. Additionally, labels marking non-lexical vocalisations such as laughter or hesitation have been turned into English ones, since the corpus is meant to serve the international research community.

The current paper presents the conventions along which file names and annotation file structures are set up, and it gives a description of the annotation guidelines. Subsequently, a statistical overview is provided on a subset of the database with respect to the phenomena described in the next section.

3. Setup of sound and annotation files

Speech was recorded in mono wav files via a standing microphone for each discourse partner. File names have the following structure:
bea_461_m_24_readsent_stm
The underscores divide the following pieces of information: *bea* refers to the corpus, the 3-digits number is the speaker ID. The letters *f, m* refer to the speaker's gender (female, male), the 2-digit number of their age at the recording. The recording protocol included eight modules with spontaneous and non-spontaneous speech that are encoded in the file name accordingly, here for read sentences. The final unit refers to the fact that recordings were carried out with a standing microphone, which is the case for all BEA samples.

The current annotation files are in the textgrid format of the Praat software (Boersma & Weenink, 1999). First, signal detection was run on the wav files, by which chunks of spoken passages were marked as intervals in the previously created textgrids. Since the initial and final boundary of the utterance cannot

always be detected exactly, e.g. when starting with a voiceless stop containing an initial closure phase, utterances begin and end with a silent phase of maximal 300 ms. The attempt to separate and label speakers automatically was not successful because they were recorded on the same channel. This task was performed by a team of annotators during the manual corrections following ASR.

The target speaker is encoded as SPK on the first tier of the annotation file. EXP on the second tier refers to the experimenter, i.e. the researcher being present throughout the entire recording. In the discourse module, an additional discourse partner joined the conversation who is labelled as DP on tier 3. Finally, non-human noise, such as a creaking chair, is indicated on the last NOI tier. Apart from the discourse module of BEA, textgrids contain three tiers: SPK, EXP, and NOI. Overlapping speech is indicated by overlapping intervals on tiers devoted to different speakers. Very short overlaps, e.g. backchannelling signals from a different speaker are not always marked, since they are not recognised by ASR, and they do not sincerely affect acoustic analysis.

4. Annotation guidelines

Earlier annotation guidelines followed the principle that the text along with the wav files should follow audio events as exactly as possible. This approach was applied in order to enhance manual searches in the textgrid files, e.g. when selecting certain phone sequences in the material.

The revised annotation system has a different approach: instead of being close to the actual acoustic material, the annotated text follows standard Hungarian orthography with only few exceptions. The reason for choosing a broader transcription system is manifold. First, the orthographic forms enhance the searchability of the corpus when analysing certain lexical units for purposes other than the actual phonetic realisation. An example is the discourse particle *tehát* 'this means' that is often realised with reduction resulting in forms such as *tát, tet* etc. At the same time, a pragmatic analysis might not be interested

in the actual acoustic realisation but intends to deal with all occurrences of the lexical unit in various contexts. Another reason why the full lexical form is given is that modern tools for automatic speech segmentation can deal with varying grapheme-to-phoneme mappings, thus, it is not necessary for the annotator to remain close to the actual realisation. Besides, like in the example of the discourse particle *tehát*, a frequent realisation *tát* is identical to another lexical unit *tát* 'open/gape+3rd person singular' indicating a verb instead of a discourse particle. Third, using the lexical forms makes annotators' lives easier because they can rely on their understanding of the context, and they are not required to perform broad phonetic transcription which is in fact a different task.

In the next sections, instructions for annotators will be presented. Guidelines reflecting Hungarian orthography are followed by those deviating from standard lexical forms. Section 4.2 specifies labels for non-lexical utterances of speakers, whereas section 4.3 describes an extended function of the NOI tier, originally reserved for non-human noise.

4.1. Relation to Hungarian orthography

The following punctuation marks occur in the annotations: .,?!. Since our current ASR system does not output punctuation marks, they are inserted manually by annotators during the manual correction process. Capital letters are reserved for proper names and abbreviated forms, therefore utterances start with lower-case letters. Simple hyphens are used according to Hungarian orthography, e.g. in multiple compounds longer than 6 syllables such as *teherautó-forgalom* 'van traffic'. Digits, however, are spelled out, i.e. *tizenhárom* for '13'.

4.1.1. Irregular orthography

Words with irregular orthography are given both with their actual phoneme sequence and with their lexical form in square brackets. This includes proper names with no direct grapheme-phoneme mapping, foreign names and words,

and words including digits. Another occurrence of the actual realisation and the proper written form is with mispronounced words. Some examples:

- kosut [Kossuth] (surname of the 19^{th} century politician Lajos Kossuth),
- váo [wow],
- cépluszplusz [C++] (the programming language),
- tévékettő [TV2] (television channel),
- szerklény [szekrény] (mispronounced wardrobe).

Colloquial pronunciations of lexical and morphological forms are usually given with the typical orthography in written text style or dictionary form. Most such non-standard forms include the deletion of certain segments such as word-final /r/, sometimes leading to vowel lengthening (e.g. *amikó* 'when' as a relative pronoun instead of standard *amikor*).

The same principle applies to the colloquial merger of the suffixes *-ban/ben* and *-ba/be*. If the former, the inessive (e.g. *in the house*) is replaced by the illative (e.g. *into the house*), as is frequently the case in colloquial speech, it is still written as *-ban/ben*, corresponding to the semantic context and to the usually written form. A slightly different case is the annotation of dialectal or non-standard word forms such as *köll* for standard *kell* 'needs to'. Since it is not a reduction, but an alternative word form, here the version *köll* is given.

This is handled differently with colloquial forms that are lexicalised, i.e. people would usually write them in informal style. A short list of such units contains *nemtom* for *nem tudom* 'I don't know', *asszem* for *azt hiszem* 'I think' etc.

Unfinished word fragments due to disfluencies, e.g. interruption or replanning by the speaker are marked with a double hyphen next to the incomplete word form, e.g. *ke-- kenyér* 'bread'. The label <**interr**> is used to signalise that the utterance is interrupted by the current speaker, and they continue the utterance within the same interpausal unit (IPU) with a different sentence

structure. This helps to filter out sentences from further analysis that rely on syntactically or prosodically complete phrase units.

4.1.2. Capital letters

Capital letters are primarily used for proper names. Additionally, they signalise spelled letters such as *T mint Tamás* 'T as in the proper name Tamás'. The same rule is applied to acronyms produced with individual letters such as *MTA*, the abbreviation for 'Magyar Tudományos Akadémia', Hungarian Academy of Sciences. Abbreviations for words whose letter sequence is produced as a word rather than spelled letters are annotated differently: although the university Eötvös Loránd Tudományegyetem is abbreviated as ELTE in Hungarian orthography, it is written as Elte here since it is pronounced as the sequence of the phonemes indicated by the letters. When a letter is not used as a letter, but as a mathematical symbol such as *x tengely* 'x-axis', the form is given in square brackets, preceded by the actual phoneme sequence: *iksz tengely [x tengely]*.

4.2. Non-lexical units and non-canonical speech

Spontaneous speech contains a number of non-lexical units that are verbal utterances of the speaker, but they are not directly connected to a lexical entry or even to existing phonemes of the language. The most frequent function of these units is hesitation marking by filled pauses (e.g. English *uh, umm*). In spontaneous settings with two or more interlocutors, such units are often used as communicative signals such as backchannels, expressions of emotional state or alike (e.g. English *m-hm, hmm)*. Ward (2006) describes the relationship between the form and the communicative function of such non-lexical units by the term *conversational grunts*. Filled pauses have language-dependent realisations (Horváth, 2020). Non-lexical forms with an intentional communicative meaning, often replacing lexical forms such as *yes, what?* have been studied less frequently. The latter category often contains nasal phoneme-like sounds in Hungarian (Reichel et al., 2023), unlike filled pauses which are most often realised as a schwa (Horváth, 2020).

Following Ward's terminology, our corpus annotations include two labels for non-lexical units:

- filled pauses signalising hesitation <**hes**>,
- communicative grunts in the function of backchannelling, assertion, question etc. <**hum**>.

Unlike in the earlier versions of BEA (see Section 4.4), annotators were not supposed to find a similar sequence of letters to indicate filled pauses. Instead, these are uniformly marked as <**hes**>. Disfluencies expressed by the lengthening of a segment or syllable, but without an independent non-lexical hesitation, are only marked word-medially if the word form is interrupted and contains a pause (see Section 4.1.1).

The label <**hum**> refers to the meaningful conversational grunts realised as *m-hm, m-m, hm* that are referred to as communicative grunts. They are frequent in informal communication in Hungarian, and their intonation is closely linked to their communicative function such as assertion, agreement, disagreement, question, surprise etc. Similarly to filled pauses, this kind of signal is not turned into phoneme sequences in the annotations. Given the promising results on the distinction between manually labelled hesitations and hummings using DNN (Reichel et al., 2023), we are currently expanding our model to automatically recognise these non-lexical signals along with speech sounds (Mihajlik et al., in print).

A special case of non-lexical speech is represented by passages in which the utterance is not intelligible even after careful listening. These units are marked as <**unint**> for 'unintelligible'.

A further group of meta-linguistic markers relates to either non-lexical or paraverbal phenomena that are either produced as separate units or realised on the top of (sequences of) words.

These are the following:

- laughter, either as a separate unit: <**laugh**>, or one or more words during

which the speaker laughs (speech-laugh): <***laugh***> *hát ez még soha nem jutott eszembe* </***laugh***> 'I have never thought of this';

- whispered speech: <***whisper***> *hogy magyarázzam el?* </***whisper***> 'how should I explain it?',
- singing, e.g. when the speaker refers to a song while singing a short passage of it <***sing***> *tovább, tovább, tovább* </***sing***> 'further, further, further' – cited from a well-known farewell song.

4.3. Extensions in two task-oriented corpora

Two further speech corpora have been annotated along the same guidelines: the Akaka Maptask Corpus (AMC) (Molnár et al., 2023) and the Budapest Games Corpus (Mády et al., 2023). These two smaller task-oriented corpora contain collaborative games with two interlocutors. In AMC, one participant received a map of a cave system, while the other was supposed to guide them on earth level to the appropriate exit. Thanks to the different perspectives of the two maps, participants were involved in intensive interaction. The Budapest Games Corpus is based on an object placing task with two participants, introduced in Gravano et al. (2007). Each speaker was seated in front of a laptop, divided by a board in order to prevent visual contact. Both participants saw a set of objects on their screen, in almost identical arrangement. One object was blinking on the first speakers' screen and was placed in the bottom panel for the second speaker. The first speaker was instructed to describe the exact position of the object with reference to the other objects, while the second speaker dragged the object to the suspected position using the mouse. The overlap of the target objects' position on the two screens was measured in percentage (amount of identical pixels). Speaker pairs were organised in groups and competed with the other groups, resulting in motivated and partly emotional exchange (joy, disappointment, surprise etc.). The experimenter did not participate in the task-oriented dialogues, but in some cases it was necessary to give support. If

the experimenter's speech is audible in the recording, the annotation is given on the NOI tier, that is otherwise used for non-human noise.

4.4. Deviations in the first subset of BEA

As mentioned in Section 2, the first subset of the BEA database was annotated manually by a large group of researchers, research assistants and students over years. Given the enormous input of human resources, these annotations form a specific subset of the database called Release 1. Since the annotation guidelines were different from the present ones, an attempt was made to turn these into the labels presented in the previous paragraphs. This could be done automatically for abbreviations of non-lexical vocalisations such as <laugh> instead of *NEV* (for Hungarian 'nevet') or filled pauses, i.e. <hes> instead of the phone sequence perceived in the signal (*ÖMM, öö* etc.).

It is important to emphasise that the earlier versions of the database (sound and text files) that are in use by several research teams in and outside Hungary are set up according to the guidelines described in the earlier papers listed in Section 2. For example, earlier annotations included more non-lexical human noise such as coughing, breathing, clearing one's throat etc. However, these phenomena were not marked consistently by annotators which introduces problems when training ASR models. Therefore, these labels were deleted from the unified annotations of the BEA database.

The most important difference is that Release 1 does not contain punctuation marks because annotators were instructed not to use them. Interrogative forms are labelled as <**q**> ... </**q**>, similarly to passages spoken while laughing. Annotators were asked to set the interval boundaries exactly to the start and end of the utterance, thus the onset and offset of the speech signal is not preceded by silence, unlike in later textgrid versions. Thus, IPUs are generally shorter than in textgrids created later. The label <**interr**> was not used in Release 1 for marking incomplete syntactic structures.

Overlapping speech was often not annotated for any of the speakers but simply marked as overlap. In the current version, these units are labelled as

<**unint**> even if the speech is intelligible but not annotated for two or three speakers on the different tiers.

The second substantial difference is that the text originally contained exact representations of the actually spoken form of many words such as *miér, mér, mé* for *miért* 'why'. In some cases, both the actually spoken and the intended standard form are given as *mér [miért]*. In other cases, the text simply contains the reduced word forms. In order to find these occurrences, an automatic check of lexical forms according to Hungarian orthography was run on the texts. If words not contained in the lexicon were found, they were marked and manually checked. If the reduced form coincided with another existing word form such as *mér* 'measure 3rd person singular', these cases remained undetected and could not be turned into their standard orthographic form without further text processing. The same is true for reduced suffixes such as *-ba/be* instead of *-ban/ben*, a colloquial form for the inessive. Later revisions of Release 1 of the BEA database might contain adaptations to the annotation guidelines introduced above.

A further deviation from the aforementioned annotation principles regards words that were interrupted by the speaker. If the incomplete word form could be detected with the spell checking tool, the interruption was indicated by a double hyphen as in the current annotation conventions. When incompleteness was not marked, and the resulting form was an existing word, it remained unmarked in the final version.

Even if Release 1 does not fully rely on the annotation guidelines described above, most discrepancies could be detected and adapted to the current conventions automatically or manually.

5. BEA Release 1 in numbers

Release 1 of the BEA database contains 65 hours of recorded speech from 115 speakers, consisting of over 100,000 interpausal units. The total durations of the different speech styles are as follows: around 16 hours of read speech

and sentence repetition, 6 hours of semi-spontaneous speech and 43 hours of spontaneous speech.

label	read	semi-spontaneous	spontaneous	sum
<hum>	157	567	4188	4912
<hes>	436	2388	9035	11859
−−	721	486	3525	4732
<laugh>	238	206	4832	5276
</laugh>	2	2	62	66
<q>	108	61	947	1116

Table 2: Occurrences of various speech phenomena in the database

The numbers of occurrences of certain speech phenomena in the annotated material are shown in Table 2. As mentioned in the previous sections, we marked the occurrences of hesitations, word fragments, laughter, speech-laugh, humming and questions in the textgrids corresponding to the audio files. The occurrences of disfluencies (e.g. hesitations, word fragments) are numerous in both read and spontaneous speech, facilitating corpus-based statistical analyses.

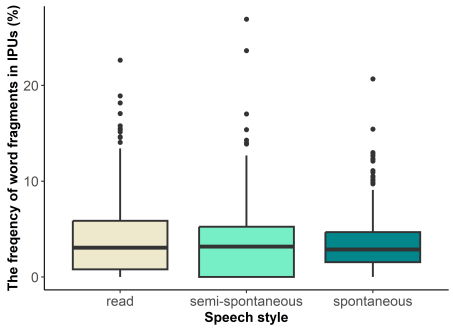


Figure 1: Relative frequency of word fragments in IPUs (%).

The number of occurrences of each phenomenon was divided by the total number of interpausal units (IPUs) and multiplied by 100 to get the percentage

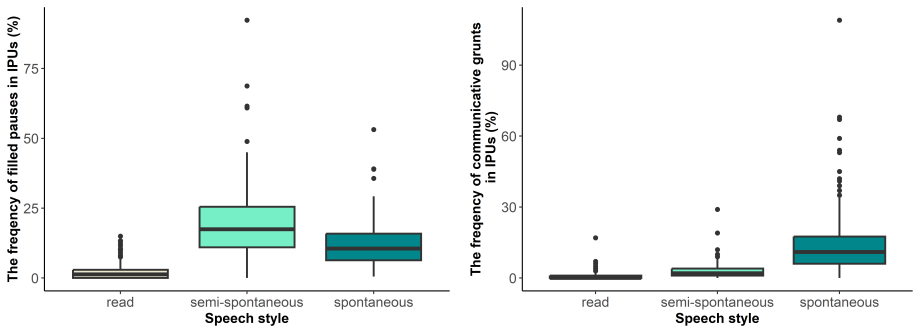


Figure 2: Relative frequency of hesitations and conversational grunts in IPUs (%).

values shown in the plots. Frequencies of various types of disfluencies are shown in Figures 1 and 2 for the different speech styles. Although drawing solid conclusions would require a more detailed analysis, the plots indicate that word fragments occur with a similar frequency in all speech styles, while hesitations can be encountered mostly in semi-spontaneous speech and are less typical in read speech. This is probably due to the fact that semi-spontaneous speech is task-oriented, i.e. speakers were supposed to retell a story instead of giving information about their lives or opinions they are more comfortable with. This example demonstrates the applicability of the database for annotated speech phenomena in a more general sense. The occurrences of tokens or phrases can be easily investigated with similar methods.

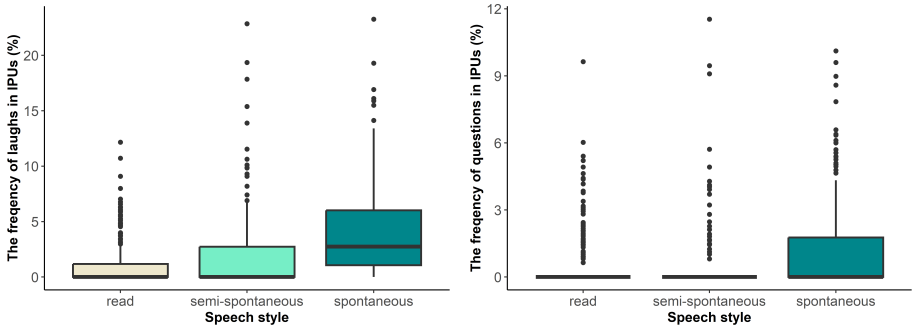


Figure 3: Relative frequency of laughters and questions in IPUs (%).

Laughters predominantly occur in spontaneous speech (Figure 3, left). However, they occur to a lesser extent in read and semi-spontaneous speech, as these tasks were also performed in the presence of a conversational partner. Laughters in these experiments seldom caused the speakers to laugh while saying one or more words. Communicative grunts are also found primarily in the spontaneous modules of the database, indicating consent in most cases (Figure 2, right). At the same time, in semi-spontaneous speech, experimenters often used communicative grunts as a feedback to encourage task continuation. Although questions only sporadically occur in the read and semi-spontaneous passages, their number is somewhat higher in spontaneous speech (Figure 3, right), which is relevant for the linguistic and phonetic investigation of various sentence types.

6. Final remarks

The revised annotation conventions will hopefully enhance linguistic and phonetic research on a large amount of annotated speech data. Both the read and (semi-) spontaneous parts of BEA are well suitable for segmental and prosodic analyses in speech research and for syntactic and semantic studies that require a large amount of data, along with the development and testing of tools in language and speech technology. The smaller Akaka Maptask Corpus and the Budapest Games Corpus are set up differently: The design was developed for both corpora in order to trigger intensive interaction between dialogue partners with varying semantic and pragmatic contents. The speech material contains various sentence types both in their canonical and non-canonical form, e.g. a large number of questions with speaker intentions other than information retrieval, e.g. surprise or uncertainty in form of self-directed questions.

The corpora (sound and annotation files) and the BEAST (BEA Speech Transcriber, Kádár et al., 2023) automatic speech recognition tool are available for research purposes for free. Researchers from countries within the EU and other countries that committed to the General Data Protection Regulation (GDPR) will receive access to the data after filling in the registration form and

signing a statement that they accept the GDPR guidelines. Research institutes from other countries need to sign an institutional contract that is in accordance with GDPR. When building your research on any of the corpora, please refer to the current paper or other relevant publications listed on our website `https://phon.nytud.hu/voxbox/bea/reg.html?lang=en`.

Acknowledgements

This work was funded by the National Research, Development and Innovation Fund (NKFIH), grants K 135038, K 143075 and FK 128814.

We would like to thank András Balog for his indispensable help with the standardisation of previous annotations and his extensive contribution by the application of ASR tools. Uwe Reichel participated in creating the set of non-canonical labels in an earlier, more detailed annotation system for the Budapest Games Corpus. We are grateful to our annotators: Lili Cziáky, Péter Csényi, Éva Alíz Ernhöffer, Gergő Zsolt Gila, Flóra Hegyi, Boglárka Kaposvári, Sára Kovács, Boglárka Mákos, Katalin Pirsel, Henrietta Pokk, Luca Pollak and Szilárd Tóth for the manual correction of the ASR-based annotations.

References

Bazillon, T., Estéve, Y., & Luzzati, D. (2008). Manual vs. assisted transcription of prepared and spontaneous speech. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA). URL: `http://www.lrec-conf.org/proceedings/lrec2008/pdf/277_paper.pdf`.

Boersma, P., & Weenink, D. (1999). *PRAAT, a system for doing phonetics by computer*. Technical Report Institute of Phonetic Sciences of the University of Amsterdam. 132–182.

Gósy, M. (2012). BEA – a multifunctional Hungarian spoken language database. *The Phonetician*, *105–106*, 51–62.

Gravano, A., Beňuš, v., Chávez, H., Hirschberg, J., & Wilcox, L. (2007). On the role of context and prosody in the interpretation of 'okay'. In *Proc. 45th Annual Meeting of Association of Computational Linguistics* (pp. 800–807). Prague.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*, 82–97. doi:10.1109/MSP.2012.2205597.

Horváth, V. (2020). Filled pauses in hungarian: their phonetic form and function. *Acta Linguistica Hungarica*, *57*, 288–306.

Kádár, M., Dobsinszky, G., Mády, K., & Mihajlik, P. (2023). "Feeding the beast" – A BEA Speech Transcriber továbbfejlesztése és integrálása neurális nyelvmodellel. In *XIX. Magyar Számítógépes Nyelvészeti Konferencia, MSZNY-2023* (pp. 135–145). Szeged.

Mády, K., Kohári, A., Reichel, U. D., Szalontai, A., & Mihajlik, P. (2023). The budapest games corpus. In T. E. Gráczi, V. Horváth, K. Juhász, A. Kohári, V. Krepsz, & K. Mády (Eds.), *Beszédkutatás – Speech Research Conference* (pp. 75–77). Budapest.

Mihajlik, P., Balog, A., Gráczi, T. E., Kohári, A., Tarján, B., & Mády, K. (2022). BEA-Base: A benchmark for ASR of spontaneous Hungarian. In *Proceedings of the Language Resources and Evaluation Conference* (pp. 1970–1977). Marseille, France: European Language Resources Association. URL: https://aclanthology.org/2022.lrec-1.211.

Mihajlik, P., Meng, Y., Kádár, M. S., Linke, J., Schuppler, B., & Mády, K. (2024). On disfluency and non-lexical sound labeling for end-to-end automatic speech recognition. In *Interspeech 2024, Kos, Greece* (pp. 1270–1274). doi:10.21437/Interspeech.2024-2157.

Molnár, C. S., Mády, K., Mihajlik, P., & Gyuris, B. (2023). The Akaka Maptask Corpus. In *Beszédkutatás – Speech Research Conference* (pp. 81–83). Budapest.

Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy, M., & Beke, A. (2014). Development of a large spontaneous speech database of agglutinative Hungarian language. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech and Dialogue. TSD 2014. Lecture Notes in Computer Science* (pp. 424–431). Springer.

Reichel, U. D., Kohári, A., & Mády, K. (2023). Acoustics and prediction of non-lexical speech in the Budapest Games Corpus. In *Beszédkutatás – Speech Research Conference*.

Schiel, F. (2004). MAUS goes iterative. In *Proceedings of the 4. International Conference on Language Resources and Evaluation* (pp. 1015–1018). Lisbon, Portugal: European Language Resources Association.

Ward, N. (2006). Non-lexical conversational sounds in American English. *Pragmatics and Cognition*, *14*, 113–184.