5. 2. (2025)

BUDAPEST

BESZÉDTUDOMÁNY
SPEECH SCIENCE

ELTE Nyelvtudományi Kutatóközpont
ELTE Research Centre for Linguistics

BESZÉDTUDOMÁNY – SPEECH SCIENCE

2025 (5) 2

Szerkesztők/Editors:

Gráczi, Tekla Etelka

Mády, Katalin

ELTE Nyelvtudományi Kutatóközpont

ELTE Research Centre for Linguistics

Budapest

Szerkesztők/Editors:

Gráczi, Tekla Etelka

Mády, Katalin

Szerkesztőbizottság/Editorial board:

Bučar Shigemori, Lia Saki (Ludwig-Maximilians-Universität München)

Bunta, Ferenc (University of Houston)

Hámori, Ágnes (ELTE Research Centre for Linguistics)

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

Huntley-Bahr, Ruth (University of South Florida)

Kohári, Anna (ELTE Research Centre for Linguistics)

Markó, Alexandra (Special Service for National Security, Institute for Expert Services & ELTE Research Centre for Linguistics)

Mildner, Vesna (University of Zagreb)

Olaszy, Gábor (Budapest University of Technology and Economics)

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

Sztahó, Dávid (Budapest University of Technology and Economics)

Trouvain, Jürgen (Saarland University)

White, Laurence (Newcastle University)

Technikai szerkesztés/Typesetting: Garai, Luca

Borítóterv/Cover design: Iuga, Laura ©

Korrektúra/Proofreading: Jankovics, Julianna & Garai, Luca

A folyóiratszám kiadását az ELTE Nyelvtudományi Kutatóközpont támogatta. / This volume was supported by ELTE Research Centre for Linguistics.

©ELTE Nyelvtudományi Kutatóközpont / ELTE Research Centre for Linguistics, H-1068 Budapest, Benczúr u. 33.

Szerkesztői előszó

Kedves Kollégák!

A kétnyelvű *Beszédtudomány–Speech Science* folyóirat a 2019-ig 27 éven át megjelenő *Beszédkutatás* folyóirat utódja. Célunk a beszédtudomány különböző területeiről érkező kutatások ismertetése magyar és angol nyelven.

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

A kéziratokat a beszédtudomány bármely területéről várjuk, például artikuláció, akusztikum és percepció; fonológiai folyamatok érvényesülése a beszédben; prozódia, szintaxis; pragmatikai vonatkozások; beszédtechnológia, beszédfelismerés, beszédszintézis; kriminalisztikai kutatások és alkalmazások; az anyanyelv és idegen nyelvek elsajátítása; két- és többnyelvűség; klinikai kutatások, beszéd- és nyelvi zavarok; korpuszok, adatbázisok fejlesztése, diszharmóniás jelenségek a beszédben, valamint további, a beszéd jellemzőivel, feldolgozásával, létrehozásával kapcsolatos kérdések.

A folyóirat évi két számmal jelenik meg. A kéziratok beküldésének határideje január 15. és május 15. A beküldés és a lektorálási folyamat a folyóirat honlapján keresztül történik. Minden kéziratot két független bíráló véleményez kettős-vak eljárással. A részletek megtalálhatók a honlapon közzétett szerzői útmutatóban.

Ajánljuk figyelmükbe az e számban megjelent, a beszéd különböző vetületeit elemző írásokat inspirálódásra, legújabb eredmények felfedezésére.

Üdvözlettel a szerkesztők,

Gráczi Tekla Etelka és *Mády Katalin*

Editorial foreword

Dear Colleagues,

The bilingual journal *Beszédtudomány – Speech Science* is the successor of *Beszédkutatás* (Speech Research), published until 2019 for 27 years. Our goal is to present research from various fields of speech science and to increase the number of international publications in English and Hungarian.

Three types of submissions are accepted: traditional research papers, short papers (pilot and case studies, negative or contradictory results), and methods papers (descriptions of methodologies, tutorials). Authors' guide is available via the journal's website: `https://ojs.mtak.hu/index.php/besztud/information/authors`

Papers in all areas of speech science are welcome, such as: articulation, acoustics and perception; realisation/manifestation of phonological processes in speech; prosody, syntax; pragmatic aspects; speech technology, speech recognition, speech synthesis; forensic research and applications; first and second language acquisition; bi- and multilingualism; 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 journal is published in two issues per year. The deadline for manuscript submission is the 15^{th} of January and the 15^{th} of May. The submission and review process are managed through the journal website. All manuscripts are reviewed by independent researchers in a double-blind peer-review process. More information is provided in the authors' guidelines on the website.

We recommend the articles published in this issue, which analyze various aspects of speech, for inspiration and to discover the latest findings.

Sincerely, the editors,

Tekla Etelka Gráczi and *Katalin Mády*

Tartalomjegyzék/Table of contents

Az izolált mandarin kínai lexikai tónusok akusztikai elemzése kínaiul tanuló magyar anyanyelvűek ejtésében

Juhász Kornélia

Nyelvtudományi Kutatóközpont
ELTE Eötvös Loránd Tudományegyetem
MTA–HUN-REN NYTK Neurofonetikai Kutatócsoport

Abstract

The experiment aims to provide an acoustic phonetic investigation into how Hungarian learners of Mandarin produce isolated Chinese lexical tones. In Mandarin Chinese (MC), four lexical tones are contrasted: high level Tone 1 (T1), rising Tone 2 (T2), low falling-rising Tone 3 (T3), and falling Tone 4 (T4). These four tones do not exclusively differ by their F0 curve, but their duration also serves as an acoustic cue for differentiation. The primary focus of the study is to acoustically compare L2 learners and MC natives' production by two acoustic characteristics: the duration, as well as the shape of the f0 curves. The speech of two L2 learner groups (beginners, advanced learners) was compared to a native MC control group (8 speakers per group, all women). Speakers were asked to read CV-structured meaningful words (*ma* syllables), characterized by the four lexical tones. The analysis included the comparison of the duration of the syllable, as well as contrasting the individual tonal realizations' f0 contours among the three speaker groups. The results show that both L2 learner groups produced Mandarin lexical tones with the same temporal characteristics as native speakers, both for absolute and relative durations. Concerning the shape and register of the f0 curves, both L2 learner groups produced the isolated tonal patterns similarly to native MC speakers, i.e., the production of the four lexical tones did not pose problems for Hungarian learners. Some minor differences were observed primarily in the case of beginners, whose production differed more from the native patterns compared to advanced learners: the concavity and the f0 range of the T2, T3, and T4 were distinct compared to the MC native realizations. The significance of the study is that, to the author's knowledge, it is the first analysis that provides statistically validated results on the acoustic comparison of isolated lexical tone production in the speech of Hungarian learners of MC.

1. Bevezetés

A jelen kísérlet célja az izolált ejtésű mandarin kínai lexikai tónusok produkciójának átfogó akusztikai fonetikai vizsgálata kínaiul tanuló magyar anyanyelvű

Email address: juhasz.kornelia@nytud.elte.hu (Juhász Kornélia)

Beszédtudomány – Speech Science 2025/2. 6–43. DOI: 10.15775/Besztud.2025.2.6-43

beszélők ejtésében. A tanulmányban a négy kínai lexikai tónust két akusztikai szempont szerint, vagyis a dallamívük menete, valamint a megvalósulásuk időtartama szerint hasonlítom össze két, eltérő mértékű célnyelvi tapasztalattal rendelkező kínaiul tanuló csoport ejtésében. A tanulmány jelentősége az, hogy a szerző tudomása szerint először nyújt statisztikai módszerekkel megerősített eredményeket a kínaiul tanuló magyar anyanyelvűek produkciójáról olyan tekintetben, hogy az elemzés mind a négy lexikai tónus izolált ejtésére is kiterjed. A következőkben bemutatom, hogy hogyan valósul meg a beszéddallam kihasználása a szóban forgó, a nyelvtanulók anyanyelveként (L1) megjelenő magyar, illetve az elsajátítandó célnyelvként (L2) megjelenő mandarin kínai esetében.

Az atonális, avagy monoton (tehát lexikai tónusokkal nem rendelkező) nyelvek esetében, mint amilyen a magyar is a beszéd hangmagasságának változása elsősorban mondat- vagy közlésszintű egységek jelentését változtatja meg (Markó, 2017). Tonális (politon) nyelvek esetében (mint amilyen a mandarin kínai) azonban a hangmagasságváltozás elsősorban lexikai szinten határozza meg a jelentést: a lexikai tónus a szó(tag) argumentuma. Ebből következően ugyanahhoz a szó(tag)hoz eltérő hangmagasságváltozás-mintázatok (dallamkontúrok) társulhatnak, amelyek jelentésmegkülönböztető szereppel bírnak (Chao, 1948/1963). Habár a kínai nyelv tonális, ez nem jelenti azt, hogy a kínai anyanyelvűek beszédében a mondatok/közlések szintjén ne jelenne meg az intonáció: a lexikai tónusok akusztikai megvalósulását az intonáció formálja és határozza meg (Shen, 1989). Akár atonális, akár tonális nyelvről van szó, a beszédképzésben mind az intonáció, mind a lexikai tónus létrehozása ugyanazzal a fiziológiai jelenséggel, azaz a hangszalagműködés változásával áll összefüggésben. A hangszalagok nyitódásából és záródásából előálló kváziperiodikus rezgést zöngének nevezünk. A zönge egyik alapvető jellemzője az alapfrekvenciája (f0), amely az egy időegység alatt lezajlott periódusok számát jelenti. Az intonáció, illetve a tónusok produkciójához szükséges változás az alapfrekvencia változásából fakad, ugyanis a beszédhang észlelt hangmagassága az alapfrekvenciával függ össze úgy, hogy a magasabb alapfrekvencia magasabb hangmagasság érzetét kelti ('t Hart et al., 1990; Gósy, 2004). Abból fakadóan, hogy a beszédhangmagasság és az f0-értékek

közötti összefüggés nem lineáris, hanem inkább logaritmikusnak tekinthető, a beszélők közötti eltérések normalizálása végett az f0-értékeket félhangokká szokás konvertálni (vö. Nolan, 2003, illetve e jelenség részletes bemutatását lásd: Juhász, 2023). A tanulmányban az alapfrekvencia-értékeket – a más szerzőktől hivatkozott ábrákat kivéve – minden esetben félhangokként jelenítem meg. A félhangok használatával kapcsolatban érdemes azt is megemlíteni, hogy a félhang-távolságokat lehetőség van zenei hangközökben kifejezni és számszerűsíteni, megkönnyítve a lexikai tónusok f0-terjedelmének bemutatását (vö. pl. Bolla, 1995).

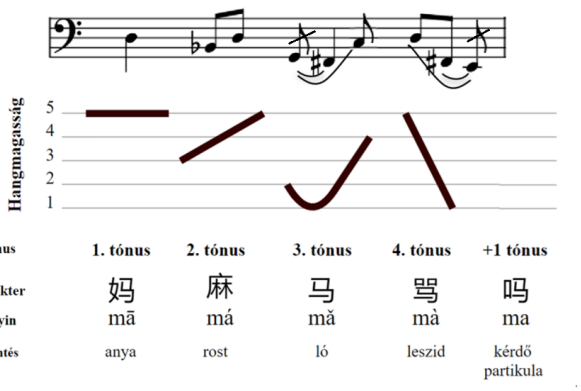
A mandarin kínai nyelv, más szóval a sztenderd kínai nyelvváltozat a hét nagy kínai dialektuscsoport közül az észak-mandarinon, elsődlegesen a pekingi változaton alapszik. A mandarin kínaiban 4(+1) lexikai tónust tudunk elkülöníteni: a magas szinttartó 1. tónust (T1), az emelkedő 2. tónust (T2), az ereszkedő-emelkedő 3. tónust (T3) és az ereszkedő 4. tónust (T4). Ezen tónusok dallammeneteit az 1. ábra mutatja be. Az 1. ábrán a négy lexikai tónus mellett megjelenő +1 tónust neutrális tónusnak szokás megnevezni, és a neutrális tónusú morfémákat tónustalan szótagnak szokás tekinteni, hiszen e morfémák nem rendelkeznek önálló dallammenettel, a megvalósulásuk a tónuskörnyezet és az intonáció függvényében változik (Shen, 1989; Juhász, 2024). A jelen tanulmány kizárólag a négy teljes értékű lexikai tónusra fókuszál, és a neutrális tónusok ejtésével – rendkívül diverz megvalósulásukból fakadóan – nem foglalkozik. Az 1. ábrán látható kottáról az olvasható le, hogy egy férfi beszélő ejtésében a szinttartó T1 relatíve magas hangmagasság-tartományban, stabilan kitartva (*D* hangként) valósul meg, Ugyanezt a hangmagasságot (a *D* hangot) jeleníti meg az emelkedő T2 záró fázisához tartozó hangjegy, valamint az ereszkedő T4 tónus ereszkedő fázisa. Míg az emelkedő T2 esetében egy négy félhangos (nagy tercnyi) emelkedést figyelhetünk meg a kottaképben (*Bé–D*), addig az ereszkedő T4 esetében egy 10 félhangnyi (kis szeptim) ereszkedés figyelhető meg (*D–E*). Az ereszkedő-emelkedő T3 esetében pedig egy kis szekundnyi (1 félhangnyi, *G–Fisz*) lelépést egy bővített kvartnyi (hat félhangnyi, *Fisz–C*) fellépés követ. Az ereszkedő-emelkedő T3 ejtése esetében mindenképpen érdemes megemlíteni azt

is, hogy a pekingi dialektusban a T3 alacsony f0-tartományban megvalósuló ejtése irreguláris zöngeképzéssel, pontosabban recsegő/rekedtes fonációval (creaky voice) társul. Ez azt jelenti, hogy a T3 esetében az alapfrekvencia lelassul és a hangszalagok feszítettsége is csökken (Markó, 2013). E képzési tulajdonság alapvetően korlátozza e lexikai tónus akusztikai vizsgálatát abból fakadóan, hogy a legtöbb akusztikai elemzésben is használt szoftver (például a jelen tanulmányban is használt Praat (Boersma & Weenink, 2022) az irreguláris zönge frekvenciatartományában már nem mér megbízhatóan, illetve egyáltalán nem mér f0-értékeket (Dorreen, 2017; Dallaston & Docherty, 2019).

Visszatérve a lexikai tónusok akusztikai tulajdonságaira és megjelenítésére, ezen kottaképekkel bizonyos szempontból ellentmond az öt fokozatú hangmagasságskála és a lexikai tónusok ebben a paradigmában való bemutatása. Az öt fokozatú hangmagasságskálán a szinttartó T1 tónust 5–5-ként szokás kódolni, az emelkedő T2 tónust 3–5-ként, az ereszkedő-emelkedő T3 tónust 2–1–4-ként és az ereszkedő T4 tónust 5–1-ként (Chao, 1948/1963). Fontos kiemelni, hogy ezen sematizált öt fokozatú hangmagasságskála szolgál a kínai lexikai tónusok elsajátításakor az elsődleges referenciának. Habár a skálán bemutatott mintázatok jól bemutatják a hangmagasság-változás irányát és relatív pozícióját, azonban például a minimális f0-érték szempontjából elfedik az ereszkedő-emelkedő T3 és az ereszkedő T4 tónus közötti kottaképen bemutatott eltérést. Ha az 1. ábrán látottakkal összevetjük a tónuskontúrok tényleges megvalósulását, vagyis az f0-kontúrjaikat (2. ábra), akkor még több inkonzisztenciát fedezhetünk fel az egyes megjelenítések között, elsősorban a maximális f0-értékek tekintetében. Az 1. ábra kottája és a sematikus hangmagasságskála egyöntetűen egyezőként mutatja be a T1, T2, illetve a T4 maximális hangmagasságát. Ehhez képest az f0-kontúrok maximuma ennél sokkal árnyaltabb: e három felsorolt lexikai tónus f0-kontúrbeli realizációja közül a T4 rendelkezik a legmagasabb, a T1 pedig a legalacsonyabb f0-maximummal, míg a T2 maximuma éppen e két tónus között helyezkedik el.

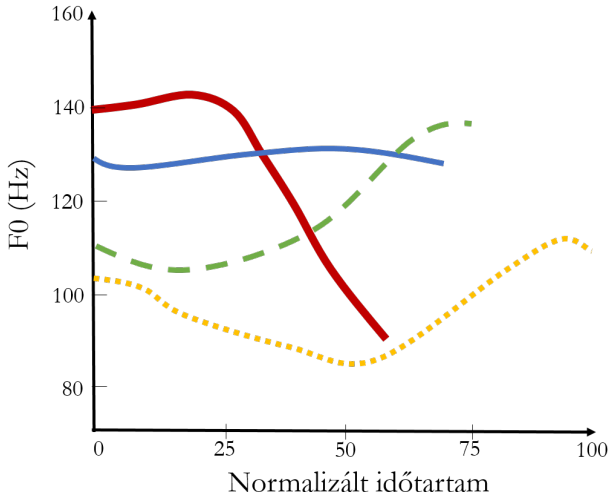
A lexikai tónusok esetében a dallamkontúr mellett egy másik fontos akusztikai tulajdonságukat is meg kell említeni: a lexikai tónusok időtartamának

relevanciáját. Zhang és munkatársai (2022) eredményei alapján a kínai suttogott beszédben az időtartam a tónus másodlagos felismerési kulcsaként szolgál, és a kísérletben a kínai anyanyelvű beszélők hajlamosak voltak kihangsúlyozni a lexikai tónusok közötti időtartambeli különbségeket suttogott beszédben azért, hogy a suttogásban neutralizálódó, a dallam által kifejezett kontrasztot fenntartsák (Zhang et al., 2022). Az időtartambeli különbségek a 2. ábrán is jól megfigyelhetők: az f0-kontúrok időtartama a leghosszabb tónushoz, a T3-hoz van normalizálva. Ehhez képest a T2 időtartamában csak ¾ olyan hosszú, mint a T3, hasonlóan a T1 megvalósulásához (amely ∼ 70%-a a T3 időtartamának). A négy tónus közül az ereszkedő T4 a legrövidebb, amely a T3-hoz képest majdnem fele olyan hosszú (∼ 55%-a a T3-nak).



1. ábra. A tónusok dallamának kottaképe férfiak ejtésében (felső sor) és sematikus dallamkontúrjai (középső sor) (Chao, 1948/1963: 85 nyomán), valamint az adott tónussal megvalósuló kínai szótagok (kínai karakterekkel, illetve a hangjelölő pinyin átirattal megjelenítve, valamint a szavak jelentése) (Juhász, 2024 nyomán, javítva).

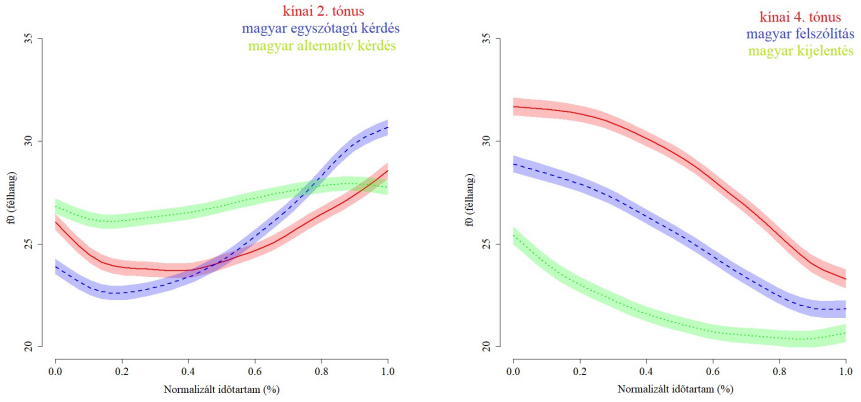
Az idegennyelv-elsajátítás (L2/célnyelv-elsajátítás), vagyis a jelen esetben a kínai lexikai tónusok ejtése és elsajátítása több szempontból is problémát okozhat az atonális magyar anyanyelvűeknek. Ha a dallamkontúrok jelentésközvetítő funkcióját vesszük középpontba, akkor az atonális magyar nyelvben is természetesen megfigyelhetők – a kínai tónusokhoz hasonlóan – monoszillabikus egységeken megvalósuló dallamok. Egy korábbi kísérletben e dallam-realizációk



2. ábra. A kínai lexikai tónusok f0-görbéje a T3 időtartamához normalizálva, ahol T1 = vékony folytonos kék vonal, T2 = szaggatott zöld vonal, T3 = pontozott sárga vonal, T4 = vastagított folytonos piros vonal (Xu, 1997: 67 nyomán).

összehasonlítása végett az emelkedő és ereszkedő kínai lexikai tónusokat egy szótagú magyar intonációs dallamokkal is összevetettem. E kísérlet alapját az a kínai tanítási gyakorlatban gyakran megjelenő módszertan adta, miszerint a magyar anyanyelvű kínai tanárok az emelkedő T2 tónust a magyar egy szótagú kérdés dallamával, míg az ereszkedő T4 tónus a megszólítás/felszólítás dallamával feltételezik egyezőnek, és a tónuselsajátítást is ilyen módon instruálják. A kínai tónusok és a magyar intonációs kontúrok megfeleltetését a következő akusztikai összevetésben vizsgáltam: az emelkedő T2 tónust a magyar egy szótagú eldöntendő kérdő dallammal (pl. *Én?*), valamint az alternatív kérdés első szótagjának emelkedő dallamával vetettem össze (lásd a kiemelt szót a következő megnyilatkozásban: ***Én*** *vagy ő?*). Az ereszkedő T4 tónust a magyar egy szótagú vokatív megszólítás és a kijelentés ereszkedésével vetettem össze. A kísérlet eredményei rámutattak arra, hogy e két kínai tónus f0-kontúrjának lefutása minden páros összehasonlításban eltér a velük összevetett magyar dallamok megvalósulásától. A kínai emelkedő tónus a magyar egy szótagú kérdő, illetve az alternatív

dallamhoz képest homorúbban valósult meg, és az egy szótagú kérdésnél kisebb, míg az alternatív kérdés emelkedésénél nagyobb f0-terjedelem jellemezte (3. ábra bal). Az ereszkedő T4 tónus mind a magyar vokatívusz, mind a kijelentésnél domborúbb dallamívvel rendelkezett (3. ábra jobb) (Juhász, 2023). A kínai oktatásban a kínai szinttartó T1 tónus és az ereszkedő-emelkedő T3 tónus esetében nem szokás a magyar intonációval, mondattípusokkal kapcsolatos párhuzamot vonni, valószínűsíthetően azért, mert nincs olyan magyar dallamkontúr, amely egyértelműen és kizárólagosan példaként szolgálhatna az említett kínai tónusok esetében.



3. ábra. A kínai T2 és a magyar egy szótagú kérdő dallam, valamint az alternatív kérdő dallam emelkedő dallammenete (balra), valamint a kínai T4 és a magyar egy szótagú vokatívusz és kijelentés f0-kontúrjai (jobbra) (Juhász, 2023: 35, 38).

A fentebb bemutatott eredmények az idegennyelv-elsajátítás középponti problémáját alapozzák meg: amennyiben egy új tanulási szituációval kell megküzdenünk, hajlamosak vagyunk a már elsajátított és ismert megoldási mintázatainkat alkalmazni a problémamegoldás során. Az idegennyelv-elsajátításban ezt a problémamegoldási stratégiát transzfernek nevezzük (Odlin, 1989; Major, 2001). A transzfer-hipotézis alapján a jelen esetben az feltételezhető, hogy az idegen nyelvi kínai lexikai tónusok elsajátításakor a nyelvtanuló a már korábban elsajátított, többnyire anyanyelvi dallamkontúrjait alkalmazza a produkciójában. Azt fontos kiemelni, hogy a nyelvek közötti transzfer nem korlátozódik az anya-

nyelv célnyelvre gyakorolt hatására, hiszen a nyelvtanuló bármelyik korábban tanult idegen nyelvei is formálhatják a produkciót (Westergaard et al., 2017). Azonban a kísérletben vizsgált nyelvtanulók legdominánsabb nyelve kétségkívül a magyar, vagyis az anyanyelv, így az L1-es mintázatok tekinthetők a transzferjelenség elsődleges forrásának, (főleg abban az esetben, ha ezt a tanítási környezet is facilitálja). Így tehát az eredmények esetében potenciálisan megjelenhet az anyanyelv hatása a lexikai tónusok megvalósulásában. Fontos megemlíteni ugyanakkor, hogy az anyanyelvi hatás közvetlen vizsgálatát kizárólag úgy lehetne megerősíteni, ha a kínaiul tanuló magyar anyanyelvűek lexikaitónusprodukcióját a hozzájuk leghasonlóbb magyar monoszillabikus intonációs dallamok megvalósulásával is összehasonlítjuk, azonban erre a jelen tanulmány nem terjed ki.

A szakirodalmi forrásokban az idegennyelv-elsajátítás folyamatát, valamint az akcentusjelenségeket kiváltó változókat mind a percepció, mind a produkció aspektusából is szokás tekinteni. Az L2-elsajátítási modellek többsége a percepció elsődlegességét feltételezi a produkcióhoz képest, hiszen az emberi észlelés révén jutunk folyamatos visszajelzéshez az ejtett beszédhangok/dallamok megvalósulásáról, és ezáltal tudjuk formálni a produkciónkat (Kuhl, 1980). Ugyanakkor például Flege és Bohn (2021) Revised speech learning model-je a percepció és a produkció szimultán fejlődését feltételezi az idegennyelv-elsajátításban (az L2-elsajátítási modellek részletes bemutatását lásd: Juhász, 2024). A tanulmány szempontjából a kínai lexikai tónusok kínai natív mintától eltérő, akcentusos ejtése mind a percepciós, mind a produkciós hatásokból is bekövetkezhet (vö. Klein, 1986). A percepciós akcentus alapja az, hogy az emberi észlelésben az L1 rendszere egyfajta fonológiai szűrőként működik, és ezáltal, ha az anyanyelvi és célnyelvi mintázatok (legyenek ezek dallamkontúrok vagy éppen beszédhangok) közötti fonetikai eltérés nem jelentős, akkor az anyanyelv szempontjából nem kontrasztív akusztikai különbségek kiszűrődhetnek a percepcióból (Flege, 1995; Flege & Bohn, 2021). Például ha a fentebb bemutatott T2-es tónust és a magyar egy szótagú kérdés dallamát tekintjük, Mennen az idegen nyelvi intonáció elsajátítására fókuszáló modellje (2014) alapján azt állíthatjuk, hogy ezek

nem csak elemezhetők az autoszegmentális fonológia paradigmáján belül, hanem ugyanazon mögöttes alacsony-magas (LH) fonológiai célokkal írhatók le, amelyek emelkedő dallammenetként realizálódnak a produkcióban. Azonban ezen emelkedő dallammenetek, mint láttuk, az emelkedő dallammenet ellenére fonetikailag számos eltérő akusztikai jeggyel rendelkezhetnek (pl. eltérő f0-tartomány, maximális f0, minimális f0, a görbe meredeksége, domborúsága stb.). Az anyanyelvi perceptuális szűrő fonológiai természetéből következően funkcionális elvek szerint működik, vagyis ha az anyanyelvben például az f0-kontúr domborúsága nem rendelkezik jelentésmegkülönböztető szereppel, akkor feltételezhetően ezen akusztikai jegyben bekövetkező eltérés észlelése – a nyelvelsajátítás legelejét tekintve legalábbis – nem jelenik meg a percepcióban. E folyamat eredményeképpen a célnyelvi mintázatot a nyelvtanulók könnyen az anyanyelviként egyezőnek észlelhetik, és transzferálhatják produkciójukba is. Ha a nyelvtanuló az L1 és L2 között fennálló akusztikai különbségek dacára a célnyelvi mintázatot az anyanyelvi megfelelőjével helyettesíti, azt Flege (1995), valamint Flege és Bohn (2021) munkája alapján ekvivalensként való osztályozásnak nevezzük. Ugyanakkor az akcentusos ejtés problematikája produkciós nehézségekből is eredhet: ebben az esetben a nyelvtanuló észleli az L1 és L2 mintázata közötti különbséget, pusztán a dallamkontúrok akusztikai megvalósítása nem közelíti meg a natív mintázatot. Ebben az esetben például az anyanyelvi artikulációs-motorikus rutinok célnyelvi mintázattal való felülírása is megnehezítheti az akcentusmentes ejtés elsajátítását (Leather & James, 1991; Mennen, 2014). Ezen felül a dallammenetek elsajátítására még hatással lehet az adott dallamok megjelenési gyakorisága is (Mennen, 2014): abból fakadóan, hogy a jelen tanulmány izolált lexikai tónusokat vizsgál, amelyeket minden kínai nyelvtanuló már az első kínai nyelvórán, megközelítőleg azonos mennyiségben észlel és produkál, ebben a tekintetben a dallamok megjelenési gyakorisága feltételezhetően nincs közvetlen hatással az ejtésre (szemben például bizonyos tónuskapcsolatokkal, ahol a különböző tónusszekvenciák megjelenési gyakorisága szóródhat). Ehhez kapcsolódóan még azt is meg kell említeni, hogy a tónuskapcsolatok esetében koartikulációs hatásokkal is számolnunk kell, vagyis a tónus-környezet hatására a lexikai tó-

nus akusztikai megvalósulása minimális fonetikai változásokon mehet keresztül (vö. Xu, 1997), amely variabilitás a nyelvtanulók számára nehezítheti a lexikai tónusok ejtését. A szerző saját, mind tanulói, mind oktatói tapasztalatai alapján az izolált ejtésű tónusok elsajátítása (és tanítása) – a tónusszekvenciákhoz képest – jelentős figyelmet kap a kínai oktatásban, amely módszertannak része egyrészről a lexikai tónusok izolált ejtésű és hiperartikulált bemutatása és gyakorlása, hangsúlyozva a különböző tónusokra jellemző és egymástól eltérő akusztikai tulajdonságokat, ami elősegítheti a kínai tónusok közötti kontrasztok kialakítását a nyelvtanuló elméjében. Másrészről a tónuskontúrok akusztikai jegyeinek verbalizálása mellett az f0-görbék vizuális megjelenítése (l. pl. 1. ábra) is rendelkezésre áll az akusztikai különbségek nyilvánvalóvá tételére, amely technikák mind elősegíthetik (vagy éppen visszavethetik) a kínai lexikai tónusok natív ejtéshez hasonló produkciójának elsajátítását.

Összegezve, az idegen nyelvi akcentus mind a percepció, mind a produkció szintjén megjelenhet, és e jelenség potenciális felszínre kerülése alapvetően az anyanyelvi hatásból következik, aminek eredményeképpen a kínaiul tanuló magyar anyanyelvűek tónusprodukciójában akár a natív ejtéstől eltérő mintázatokra számíthatunk. Ugyanakkor fontos megjegyezni azt is, hogy – szintén a Flege & Bohn-féle (2021) Revised speech learning model-ből kiindulva – az idegennyelv-elsajátítás egy dinamikus folyamat, amely a nyelvi tapasztalat hatásával együtt és fokozatosan fejlődik. Tehát az anyanyelv mellett a célnyelvi tapasztalat mértéke is egy fontos változó az idegennyelv-elsajátítás szempontjából (Flege & Bohn, 2021), ezért a kísérletben két, eltérő mennyiségű kínai célnyelvi tapasztalattal rendelkező magyar anyanyelvű beszélő ejtését vizsgálom azon predikció tesztelése céljából, hogy vajon a több L2-tapasztalattal rendelkező nyelvtanulók jobban megközelítik-e a kínai anyanyelvi ejtést, mint a kevesebb L2-tapasztalattal rendelkező nyelvtanulók. E tanulmány kizárólag a kínai tónusok ejtésének dinamikus jellemzőit, és a produkció nehézségeit vizsgálja, vagyis az akcentusos ejtés okait nem kívánja empirikus úton feltárni, illetve az anyanyelvi hatás közvetlen vizsgálatára sem tér ki.

Az idegennyelv-elsajátítás alapjairól a kísérlet empirikus előzményeire térve kérdésként merülhet fel az atonális anyanyelvű beszélők hatékonysága a lexikai tónusok tekintetében akár más, tonális anyanyelvű személyekkel szemben. Például Lee, Vakoch és Wurm (1996) eredményei arra mutattak rá, hogy a tonális anyanyelv kizárólag akkor járulhat hozzá egy másik (L2) tonális nyelv lexikai tónusainak elsajátításához (ebben az esetben csak a percepciós diszkrimináció tekintetében), amennyiben az anyanyelvi tónusrendszer komplexebb az elsajátítandó nyelvváltozaténál. E tekintetben tehát – a Flege- és Bohn-féle (2021) Revised speech learning modelben megfogalmazottakkal párhuzamosan – az anyanyelvi mintázatok alapján prediktálható az, hogy a nyelvtanulók milyen problémákba ütköznek a lexikai tónusok elsajátításában. Hao (2012) eredményei – bizonyos szempontból ellentmondva az előbbieknek – arra utalnak, hogy habár a kantoniban 6 különböző lexikai tónus jelenik meg, ennek ellenére azonban a kantoni anyanyelvű beszélők mégsem teljesítenek jobban a mandarin tónusok percepciójában és produkciójában az atonális angol anyanyelvű beszélőkhöz képest. Emellett Hao eredményei a mandarin lexikai tónusok elsajátítását is két eltérő nehézségi szintre osztja: mind percepció, mind produkció szempontjából, mind a tonális kantoni, mind az atonális angol beszélők a T1 és T4 lexikai tónusok esetében jobban teljesítettek, mint a T2 és a T3 esetében. A T2 és a T3 esetében tapasztalt percepciós és produkciós nehézségek ezért feltételezhetően nem elsősorban anyanyelvi hatásnak köszönhetők, sokkal inkább a két lexikai tónus intrinzik akusztikai tulajdonságának, vagyis hasonló lefutásának. Nem beszélve arról az alternációról, hogy T3+T3 szekvencia esetén az első T3 T2-vé változik (T3+T3 = T2+T3) (Hao, 2012: 278).

Ha a mandarin lexikai tónusok elsajátítását a kínaiul tanuló magyar anyanyelvűek szempontjából tekintjük, akkor a négy kínai tónus izolált ejtésű megvalósulását korábban Qiuyue Ye (2013) vizsgálta disszertációjában. Az akusztikai elemzésben összesen 12 beszélő vett részt: Ye 10 nyelvtanuló (4 nő és 6 férfi), valamint két natív kínai beszélő (1 férfi és 1 nő) ejtését hasonlította össze kvalitatív minőségben a négy kínai lexikai tónus izolált ejtésű időtartama, illetve f0-kontúrja tekintetében. A nyelvtanulók 9 hónap és 3 év közötti kínai nyel-

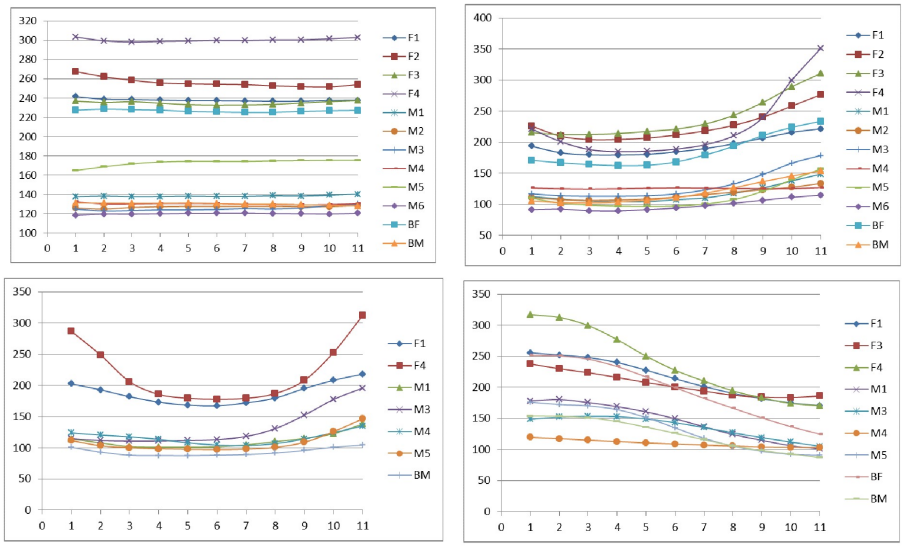
vi tapasztalattal rendelkeztek. Az akusztikai elemzés az egyéni mintázatokra fókuszált, és a nyelvi tapasztalat szóródásából következően csoportszintű produkciót Ye nem közölt, ugyanígy feltételezhetően ugyanebből az okból fakadóan statisztikai elemzést sem mutatott be. Az egyéni mintázatokat középpontba helyező eredmények a következőképpen foglalhatók össze, különös figyelmet fordítva a jelen tanulmány szempontjából is fontos akusztikai jellemzőkre, vagyis a lexikai tónusok időtartamára, valamint a f0-kontúrok megvalósulására. Időtartam tekintetében a 10 vizsgált nyelvtanuló jelentős hányada mind a négy tónus esetében a natív megvalósulásoknál rövidebbet produkált (1. táblázat, Ye 2013: 101, 138, 175, 205). Például a nyelvtanulók többsége a T4-et ejtette a többi tónushoz képest a legrövidebb időtartammal, valamint a 10 beszélő közül 8 ejtésében a natív mintának megfelelően a négy tónus közül T3 a leghosszabb. A T1 és T2 esetében a natív mintázat hasonló időtartamokat jelenít meg, azonban a 10 beszélő közül 7 a T2-t rövidebb időtartammal produkálja mint a T1-et (Ye, 2013). Ezen eredmények arra utalnak, hogy a kínaiul tanuló magyar anyanyelvűeknek nem okoz problémát a natív ejtésre jellemző időtartamokkal ejteni a kínai lexikai tónusokat.

1. táblázat. A kínai lexikai tónusok átlagos időtartama (ms) a Ye (2013) disszertációjában vizsgált 12 beszélő ejtésében (Ye, 2013: 101, 138, 175, 205). Az Fx jelölés a magyar női, az Mx jelölés a magyar férfi beszélőket jeleníti meg, míg a BF egy kínai anyanyelvű női, a BM pedig egy kínai anyanyelvű férfi beszélőt jelöl.

	F1	**F2**	**F3**	**F4**	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**BF**	**BM**
T1	415	492	382	515	283	348	281	352	597	345	421	479
T2	452	462	285	367	313	437	231	330	545	322	451	521
T3	445	574	627	697	426	463	301	481	721	310	641	599
T4	290	312	177	238	207	191	168	267	210	388	288	282

Ye (2013) f0-kontúrokra vonatkozó eredményei beszélők szintjén mutat be átlagértékeket a vokalikus szakaszban 11 ponton mérve és Excelben kirajzolva. Ye (2013) a férfi és női f0-kontúrokat Hertz értékekben mutatja be, és az

ábrákon jól elkülönülnek a férfi és női alaphangtartományok, azonban pusztán az ábrák alapján – a beszélők feltételezhetően eltérő hangszínezetét, illetve a félhangokká konvertálás hiányát is figyelembe véve – meglehetősen nehéz csoportszintű tendenciákra következtetni (4. ábra). Mindenesetre a négy lexikai tónus dallammenetének iránya minden esetben jól megfigyelhető: habár eltérő tartományokban, de a T1 szinttartó f0-mintázata jól kivehető a nyelvtanulók körében, hasonlóan a T2 emelkedésével, de e tónus ejtésében már a görbe meredekségében is látható némi variabilitás a beszélők között, hasonlóan a T3 mély, ereszkedő-emelkedő megvalósulásához. A T4 produkciója minden megjelenített nyelvtanuló esetében ereszkedő mintázatot mutat, azonban kérdésként vetődik fel, hogy a nyelvtanulók a natív ejtésnek megfelelő meredekséggel ejtik-e ezt az ereszkedő dallamot (4. ábra).



4. ábra. Ye (2013) disszertációjában résztvevő beszélők kínai tónuskontúrjai (ahol az f0-értékek a normalizált időtartam 11 pontjában vannak kinyerve és átlagértékekként vannak ábrázolva). A beszélők kódolása: az F betűt tartalmazó kódjelű beszélők nők, az M betűt tartalmazó kódjelű beszélők férfiak és a B-vel kezdődő utolsó két beszélő tekintendő a natív referenciának (Ye, 2013: 126, 163, 194, 224).

1.1. A kísérlet célja és kutatási kérdései

A tanulmányban bemutatott akusztikai elemzés célja az, hogy egy exploratív vizsgálat keretein belül dinamikus, statisztikai módszertannal megerősített eredményeket mutasson be arra vonatkozóan, hogy a kínaiul tanuló magyar anyanyelvűek lexikaitónus-produkciója eltér-e a natív kínai beszélők ejtésétől. Habár a megelőző kutatási eredmények részlegesen lehetővé tehetnék azt, hogy különkülön hipotéziseket állítsunk fel a négy kínai lexikai tónus ejtésére vonatkozóan, prediktálva az anyanyelvi mintázatok hatására bekövetkező potenciális eltéréseket a kínai anyanyelvű produkcióhoz képest, azonban – a tanulmány exploratív jellegéből fakadóan – pusztán kutatási kérdéseket fogalmazok meg. A hipotézisalkotás elmaradását a következőkkel szeretném indokolni: i.) Amennyiben a kísérlet célja az anyanyelvi hatás vizsgálata, ebben az esetben kizárólag az emelkedő T2-es és az ereszkedő T4-es lexikai tónusok hasonló magyar dallamkontúrpárjairól vannak információink, míg a magas szinttartó T1 és a mély T3 esetében nem áll rendelkezésre statisztikai elemzéssel megerősített mintázat. ii.) Ami a kínaiul tanuló magyar anyanyelvűek kínaitónus-produkcióját illeti, Ye (2013) eredményei kizárólag egyéni mintázatokat mutatnak be, eltérő számú és nemű kínai anyanyelvű és kínaiul tanuló magyar anyanyelvű beszélők ejtésében, mind az f0-kontúrok, mind az időtartam tekintetében, így tehát ezekből az eredményekből csoportszintű ejtésre, és az anyanyelvi mintától való eltérésre nem lehet közvetlenül következtetni. Ezért a jelen tanulmány a következő három kutatási kérdés vizsgálatát tűzi ki célul:

1. A kínaiul tanuló magyar anyanyelvűek a kínai beszélőkkel megegyező időtartammal ejtik a négy mandarin lexikai tónust?

2. A kínai tónusok dallamívét (f0-kontúrját) tekintve a magyar beszélők a kínai natív dallamkontúrokhoz hasonlót produkálnak?

3. További kutatási kérdés az, hogy amennyiben a kínaiul tanuló magyar anyanyelvűek eltérnek a natív kínai mintázatoktól ejtésükben, akkor megerősíthető-e a Flege- és Bohn-féle Revised speech learning model (2021) predikciója,

miszerint a több nyelvi tapasztalattal rendelkező haladó nyelvtanulók jobban megközelítik a natív mintázatokat, mint a kevesebb L2-tapasztalattal rendelkező kezdő nyelvtanulók?

Fontos kérdésként merülhet fel a tónusok rendszerszintű elemzése, vagyis a beszélői csoporton belüli lexikaitónus-kontraszt kérdése, mind az időtartam, mind a dallamkontúrok tekintetében, azonban erre a kérdésre a jelen tanulmány a terjedelmi megszorítottságból fakadóan nem terjed ki. A beszélői csoportokon belül a kínai tónusok produkciós elkülönítésére vonatkozó eredményeket egy másik, előkészületben lévő tanulmányban mutatom be (Juhász, előkészületben).

2. Módszertan

A kísérletben két eltérő nyelvi tapasztalattal rendelkező magyar anyanyelvű kínaiul tanuló beszélői csoport ejtését hasonlítottam össze egy kínai anyanyelvű beszélői csoport produkciójával. Mindhárom beszélői csoport 8 kísérleti személyből állt, mind nők. A kezdő kínaiul tanulók kínai alapszakos egyetemisták voltak, akik legalább 1 éve tanulnak mandarinul és nem rendelkeztek Kínában eltöltött ösztöndíjas nyelvi tapasztalattal (átlagéletkoruk 23,2±3,21 év volt). A haladó nyelvtanuló csoport tagjai olyan harmadéves kínai szakos hallgatók voltak, akik egy évet Észak- vagy Közép-Kínában töltöttek ösztöndíjjal (4 fő), illetve kínai mesterképzéses hallgatók voltak (4 fő), akik legalább 3 éve tanulnak kínaiul (átlagéletkoruk 24,2±2,86 év). A kínai anyanyelvű kontrollcsoport nyolc női beszélője mind északi vagy észak-keleti, mandarin dialektuscsoporthoz tartozó anyanyelvi beszélő, akik közül négyen Pekingben nőttek fel, négyen pedig Pekingtől maximum 500 km-re (átlagéletkoruk: 23,8±2,55 év). A nyolc kínai anyanyelvű beszélő mind egyetemista hallgató, négyük 5 éve, egy beszélő 3 éve, 2 beszélő két éve és egy beszélő 1 éve él Magyarországon és tanul magyar nyelven. Mindannyiuk 15–16 éve tanul angolul, amely nyelvet napi rendszerességgel használnak. A hangfelvételeket 16 bit-en, 44,1 kHz-en digitalizálva rögzítettem egy külső hangkártyával és egy omnidirekcionális kondenzátoros fejmikrofonnal.

A kísérlet anyagát izolált ejtésű CV-szerkezetű, jelentéssel rendelkező, egy szótagú szavak adták. Ezen szótagok egytől egyig a szonoráns *ma* beszédhangokból álltak, ami lehetővé tette a teljes szótagon megvalósuló f0-kontúr vizsgálatát. A *ma* hangsort a négy különböző kínai lexikai tónussal rögzítettem, így a felolvasott szavak a következők voltak: T1: 抹 mā (töröl), T2: 麻 má (rost), T3: 马 mǎ (ló), T4: 骂 mà (leszid). A felvételkészítéskor mind a kínai karakterek, mind a pinyin transzkripció meg volt jelenítve a kísérleti személyek számára. A kísérleti személyeknek ezen izolált ejtésű szótagokat random sorrendben, 6 ismétléssel kellett felolvasniuk egy képernyőről. Felmerülhet a kérdés, hogy a T1-es 抹 mā (töröl) szótag esetében miért ez a karakter került megjelenítésre a sokkal egyszerűbb és gyakrabban használt 妈 mā (anya) helyett. Ennek az az oka, hogy a jelen vizsgálatban bemutatott izolált tónusprodukció egy olyan akusztikai elemzés referenciájaként fog szolgálni, amelyben a fentebbi négy kínai szótagot egy megnyilatkozás részeként, diszillabikus egységekben is vizsgálom ugyanezen beszélők ejtésében. A 妈 mā szótag duplikálásakor azonban a létrejövő 妈妈 māma szóban a második tónus neutrális (tónustalan), így nem alkalmas egy T1+T1 konstrukció létrehozására. Megjegyzendő továbbá az, hogy ezen, laboratóriumban rögzített, kísérletvezető jelenlétében elhangzott izolált megnyilatkozások esetében hiperartikulált produkcióra, vagyis az egyes artikulációs mintázatok túlejtésére számíthatunk a spontán beszédben elhangzó megvalósulásokhoz képest (vö. Scarborough & Zellou, 2013), vagyis ebben a szituációban minden körülmény a lexikai tónusok lehetőség szerinti legjobban formált ejtését facilitálja.

A hangfelvételeket a Praat szoftverben (Boersma & Weenink, 2022) címkéztem és elemeztem: minden elemzést a teljes szótagon, vagyis a vokalikus szakaszon végeztem, és ezt az intervallumot elemeztem a szakasz időtartamaként is. Az akusztikai elemzés egy f0-kontúrokat érintő dinamikus és egy tónusidőtartamra fókuszáló statikus vizsgálatot foglal magába.

A statikus időtartamra fókuszáló elemzések esetében vizsgáltam a szótagok abszolút és relatív időtartamát is. Az abszolút időtartam a szótag valós, milliszekundumban kinyert időtartamát jelentette, míg a relatív időtartam esetében

a z-érték mentén beszélőnként normalizáltam az időtartam-értékeket, vagyis vettem az adott időtartamérték és a beszélőre számolt (négy lexikai tónusra számolt) átlag különbségét, majd ezt az értéket elosztottam a beszélőre számolt szórással (= (érték − beszélő átlaga) / beszélő szórása). Ez azt jelenti, hogy amennyiben ez a relatív érték nulla körül valósul meg, akkor közel esik a beszélő átlagához, míg például az 1-hez közeli érték arra utal, hogy a beszélő átlagához képest egy szórással hosszabb, míg −1-es érték esetében egy szórással rövidebb volt az adott megvalósulás. Az abszolút időtartamok helyett a statisztikai elemzésben elsődlegesen a relatív időtartam-értékekre fókuszálok azért, mert ezekben az adatokban az egyes beszélők közötti variabilitás, pontosabban az esetleges beszédtempót érintő eltérések szerepe kevésbé releváns, továbbá ebben az esetben az értékek relativitása a négy tónus egymáshoz való viszonyát is jobban bemutatja (még akkor is, ha a tónus-kontrasztok vizsgálata a jelen tanulmánynak nem célja). A relatív időtartamértékek statisztikai elemzését lineáris kevert modellekkel végeztem, szintén az R programban az lme4 csomag (Bates et al., 2015) segítségével. A p- és F-értékeket Satterthwaite-approximáció segítségével nyertem ki, ami az lmerTest csomagban (Kuznetsova et al., 2017) elérhető. A négy kínai lexikai tónus relatív időtartamát négy különböző lineáris kevert modellel vizsgáltam. Minden tónus esetében a relatív időtartam függő változót vizsgáltam a három beszélői csoport függvényében. Minden tónus esetében egy olyan modellt állítottam fel, amely a csoportok között kizárólag a regressziós egyenes y tengely mentén képzett elmozdulását engedi (a random metszéspont, azaz random intercept, lmer(relatív időtartam *sim* beszélői csoport + (1 | beszélő)). Ez összesen négy kevert modellt eredményezett. A függő változó három szintjét (natív/haladó/kezdő) páronként Tukey post hoc tesztekkel vetettem össze az emmeans csomag használatával (Lenth, 2020). Az adatokat a ggplot2 csomag segítségével ábrázoltam (Wickham, 2016).

A dinamikus f0-kontúrok esetében a 120 Hz felett kinyert f0-értékeket minden esetben félhangokká konvertáltam az R programban (R Core Team, 2024) a hqmisc (Quené, 2014) csomag segítségével, minden esetben 50 Hz-es referenciaértékkel. Az f0-görbék elemzéséhez az f0-értékét a vokalikus szakasz időtartamá-

hoz képest normalizáltan, 1%-onként nyertem ki, automatikusan. Az f0-görbék elemzésére generalizált additív kevert modelleket (GAMM) használtam (Wood, 2017), külön modellel vizsgálva négy lexikai tónus megvalósulását a három vizsgált beszélői csoportban. Ez azt jelenti, hogy a modellekben a vizsgált tónus ejtése a beszélői csoportok között megfigyelhető eltérésekre összpontosított. Az alapmodellben az f0 függő változó alakulását vizsgáltam a normalizált időtartam függvényében, más szóval azt, hogy az f0 értéke az időtartamra simítva hogyan változik a normalizált időtartamon belül. Az f0-görbék girbe-gurbaságából fakadóan az f0-görbéket adaptív simítással illesztettem (bs = "ad"), ami lehetővé teszi a görbe hullámzásának rugalmasabb kezelését a normalizált időtartam függvényében. A modellt a beszélői csoport parametrikus faktor változó fix hatás mellett minden tónus esetében random simítással (random smooth, bs = "fs") bővítettem. Ez két különböző random smooth funkciót jelentett: egyet a beszélőkre illesztve, egyet pedig a külön álló token-megvalósulások szerint. Az autokorreláció ellenőrzése (acf.resid()) után a modellek által becsült f0-görbéket minden esetben 95%-os konfidencia-intervallummal ábrázoltam az itsadug R csomaggal (van Rij et al., 2020).

3. Eredmények

Az eredmények bemutatását a lexikai tónusok időtartamára vonatkozó vizsgálattal kezdem, majd ezt követően mutatom be az f0-kontúrokat középpontba helyező dinamikus elemzést.

3.1. A kínai lexikai tónusok időtartamának összehasonlítása a három beszélői csoport között

A négy kínai lexikai tónus időtartamát illetően abszolút és relatív időtartam-értékeket mutatok be. Az abszolút időtartam-értékeket pusztán csak referenciaként és kvalitatív szempontból mutatom be, majd ezt követően térek rá a relatív időtartamokra, amelyeket már statisztikailag megerősített vizsgálat is kísér.

Az abszolút időtartamértékek tekintetében az 5. ábrán azt láthatjuk, hogy az adatok minden beszélői csoport esetében jelentősen szóródnak, azonban megfi-

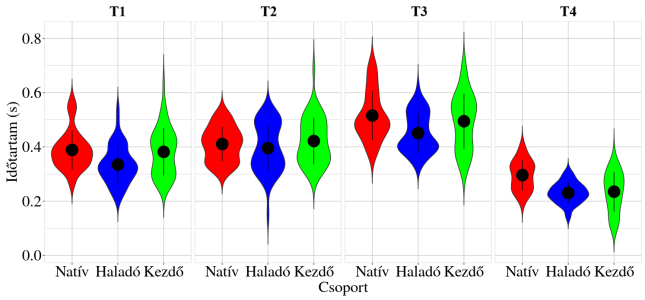
gyelhetünk egy általános tendenciát a csoportok között: a lexikai tónusok között a T3 rendelkezik minden csoport esetében a leghosszabb, míg a T4 rendelkezik a legrövidebb időtartamokkal (2. táblázat). A T1 és a T2 hasonló átlagértékekkel, de köztes időtartamokkal realizálódik a T3-ra és a T4-re jellemző átlagértékekhez képest, habár a T2 minden csoport ejtésében a T1-hez képest árnyalatnyival hosszabb időtartammal realizálódott (5. ábra). Ha a négy tónus ejtését globálisan tekintjük a csoportok között, akkor az a mintázat látszik kirajzolódni, hogy – ugyan csak tendenciák szintjén – de a csoportok között minden tónus esetében a haladó nyelvtanulók rendelkeznek a legalacsonyabb átlagértékkel, míg a kezdők ehhez képest megközelítik a natív beszélők időtartamértékeit (5. ábra).

A relatív időtartamok (6. ábra) esetében az előbb leírt időtartam-különbségek a négy lexikai tónus között nyilvánvalóbbá válik, azonban a csoportok átlagában megfigyelhető mintázatok is eltérően alakulnak (2. táblázat). Mindhárom beszélői csoport esetében a T1 és a T2 ejtésének időtartama közelítette meg a legjobban a beszélőkre külön-külön számolt átlagértékét ($z = 0$) a négy lexikai tónusra vonatkozóan. Míg a három beszélői csoport a T1-et az átlagértékhez képest egy árnyalatnyival rövidebben, addig a T2-t éppen fél szórásnyival hosszabb időtartammal ejtették. A lexikai tónusok között a T3 volt az a tónus, amelynek ejtése az átlagos ejtési időtartamhoz képest 1 szórással hosszabb átlagértékkel valósult meg, míg a T4 esetében az átlagos ejtési időtartamhoz képest 1,5 szórásnyival rövidebb időtartam figyelhető meg. A csoportok között a statisztikai próbák nem mutattak szignifikáns eltérést egyik tónus időtartamértékeiben sem. Ez azt jelenti, hogy a két kínaiul tanuló magyar anyanyelvű csoport mind a négy lexikai tónust a kínai anyanyelvűekkel megegyező időtartammal ejtette. Tendenciaszintű különbségek azonban megfigyelhetők a csoportok között: a T3 és a T4 esetében a kezdő nyelvtanulók produkálták a csoportok közötti legrövidebb relatív időtartamértékeket, míg a T1 és a T2 esetében ezzel ellenkezőleg éppen a kezdők produkálták a csoportok közötti legmagasabb relatív értékeket.

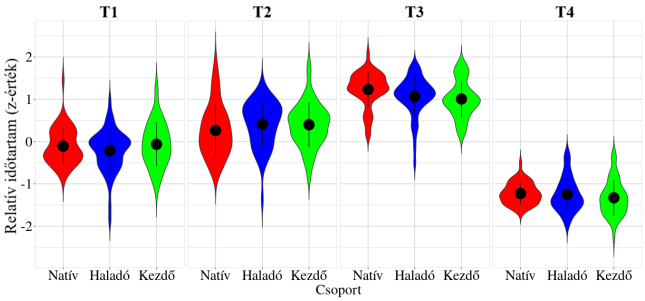
2. táblázat. A négy kínai lexikai tónus abszolút időtartamának átlaga (bal táblázat), illetve az abszolút átlagidőtartamok beszélőnként kiszámított, z-érték mentén normalizált átlagértéke.

abszolút időtartam (s)	**natív**	**haladó**	**kezdő**
T1	0,39	0,33	0,38
T2	0,41	0,40	0,42
T3	0,52	0,45	0,49
T4	0,30	0,23	0,23

relatív időtartam	**natív**	**haladó**	**kezdő**
T1	$-0,10$	$-0,22$	$-0,065$
T2	0,26	0,41	0,39
T3	1,23	1,07	1,01
T4	$-1,23$	$-1,25$	$-1,33$



5. ábra. A négy kínai tónus abszolút időtartama a három beszélői csoport ejtésében (ahol a fekete pont az átlagértéket, míg a fekete vonal a szórást jeleníti meg).



6. ábra. A négy kínai tónus relatív időtartama a három beszélői csoport ejtésében, vagyis az időtartam-értékek a beszélők szerint a z-érték mentén normalizáltak (ahol a fekete pont az átlagértéket, míg a fekete vonal a szórást jeleníti meg).

3.2. Az f0-kontúrok dinamikus összehasonlítása a három beszélői csoport között

3.2.1. A magas szinttartó T1 tónus

A három beszélői csoport T1 f0-kontúrjait összehasonlító GAMM-modell a következő eredményt hozta: A T1 modell parametrikus eredményei alapján azt mondhatjuk, hogy az egész tónuskontúr f0-értékeinek átlagát tekintve mindkét nyelvtanuló csoport ejtése a natív mintával megegyező f0-tartományban valósult meg (3. táblázat). A görbék alakját véve középpontba, a smooth együtthatók alapján mindkét nyelvtanulói csoport szignifikánsan eltért a natív kontúr alakjától (haladók: $p < 0,05$; kezdők: $p < 0,001$, 3. táblázat). Ha a T1-kontúrok alaki megvalósulását vetjük össze az 7. ábrán, akkor az látható, hogy mindkét nyelvtanulói csoport a natív ejtésre jellemző, relatíve magas f0-tartományban megvalósuló stagnáló mintázatot produkál. A haladó nyelvtanulók ejtését a natív f0-kontúrral közel teljesen átfedő egyenes jellemzi, ahol a két f0-lefutás becsült különbsége kisebb, mint 1 félhang (7. ábra). A kezdő nyelvtanulók esetében az f0-görbe a natív mintához hasonlóan szép egyenes, azonban a natív kontúrhoz képest majdnem 1 félhanggal magasabb értékekkel realizálódik (3. táblázat, 7. ábra).

3.2.2. Az emelkedő T2 tónus

A T2-re illesztett modell eredményei alapján azt láthatjuk, hogy a parametrikus eredmények szerint a görbe átlagos f0-értéke mentén egyik csoport sem tért el a natív f0-kontúrtól (3. táblázat). Továbbá a haladó nyelvtanulók a görbe alakjában, vagyis a smooth funkciók tekintetében sem tértek el a kínai anyanyelvű beszélőktől, tehát e tónust mind az f0-tartomány, mind az f0-lefutás tekintetében az anyanyelvi kínai ejtéssel megegyezően ejtették, még akkor is, ha maga a görbe a natív ejtéshez képest egész lefutásában a natív ejtésnél 1,5 félhanggal alacsonyabb f0-tartományban realizálódott. A kezdő nyelvtanulók f0-kontúrja azonban a natív ejtésnél árnyalatnyival (kevesebb, mint 1 félhanggal) magasabb f0-értékekkel indult a normalizált időtartam elején, míg a natív mintához képest közel 2 félhanggal alacsonyabb f0-értékekkel zárult (8. ábra), amely eltérések a natív mintánál több, mint 1,2 félhanggal alacsonyabb,

az egész kontúrra számított becsült átlagértéket eredményezett (3. táblázat). Továbbá azt is meg kell jegyezni, hogy a kontúr homorúbb fázisa a natív és haladó kínaiul tanuló beszélők ejtéséhez képest a normalizált időtartamon belül később következett be (7. ábra). Mindezen alaki tulajdonságok eredményezhették a szignifikáns eltérést a smooth funkciók tekintetében ($p < 0,01$) a kezdő nyelvtanulók ejtésében. Ha az emelkedés mértékét félhangokban fejezzük ki a csoportok között a görbék alapján, akkor azt mondhatjuk, hogy a natív és a haladó nyelvtanuló csoport esetében 6-6 félhangnak, tehát egy bővített kvartnak felel meg, míg a kezdők ejtésében az emelkedés megközelítőleg 5 félhangot, vagyis egy tiszta kvartot jelentett.

3.2.3. A mély ereszkedő-emelkedő T3 tónus

A T3 ejtését véve középpontba (7. ábra) egy relatíve alacsony f0-tartományban megvalósuló homorú görbét figyelhetünk meg mindhárom beszélői csoport ejtésében. A GAMM eredményei szerint az f0-görbe becsült átlagát tekintve a két nyelvtanulói csoport egyike sem tért el a natív ejtéstől, azonban – ahogy a parametrikus becsült átlagértékek is megjelenítik – a csoportok között a kezdő nyelvtanulók produkálták a legmagasabb f0-tartományban a dallamkontúrt, amely görbe kezdeti és záró fázisa a natív ejtést megközelítő becsült értékekkel realizálódott. Azonban a kezdő nyelvtanulók ejtésében a kontúrt a natív ejtéshez képest kevésbé homorú mintázat jellemezte ($p < 0,05$), vagyis összességében a natív ejtésnél magasabb f0-értékek kísérték. A haladó nyelvtanulók ejtésében a kezdők mintázatánál is lapultabb f0-kontúrt figyelhetünk meg ($p < 0,001$), amely görbe kezdeti és záró f0-értékei is a másik két csoporthoz képest alacsonyabb f0-értékekkel jelentek meg. Továbbá az f0-görbe alakját megfigyelve azt láthatjuk, hogy az adaptív smooth funkció ellenére a haladó nyelvtanulók kontúrja relatíve hullámzó. Ez a girbe-gurba forma elsősorban a normalizált időtartam középső fázisában jelentkezik, ahol a másik két csoport kontúrjának legalacsonyabb f0-értékeit figyelhetjük meg és ezzel párhuzamosan az irreguláris zöngeképzés okozta zajt is feltételezhetjük az adatokban. Az ereszkedő-emelkedő fázis nagyságát helyezve középpontba, a natív és kezdő nyelvtanulói ejtésben

megközelítőleg szimmetrikus görbéket figyelhetünk meg, amelyek a normalizált időtartam felénél érik el minimum pozíciójukat. A natív ereszkedés-emelkedés megközelítőleg 9 félhangot jelentett, amely egy nagy szextnek feleltethető meg, míg a kezdők ejtésében ez inkább csak 4 félhang, vagyis egy nagy terc hangmagasságváltozást jelentett. A haladó nyelvtanulók esetében az f0-változás leírása a görbe egyenetlenségeiből fakadóan egy fokkal nehezebb: a kezdeti ereszkedés 4 félhangra, vagyis egy nagy tercre tehető, ami nagyságrendileg megegyezik a kezdők értékeivel, illetve ha ettől a ponttól számítjuk az emelkedést is, akkor ugyanúgy 4 félhangra tehető. E lexikai tónus esetében figyelhető meg az egyetlen – GAMM szerint relevánsnak ítélt – markáns eltérés a natív és a kezdő kontúrok között: a normalizált időtartam 61,6% és 65,6% százaléka között a két f0-görbe közötti eltérés kb. 2 félhangnyira volt becsülhető (8. ábra). Habár ebben az esetben azt láthatjuk, hogy a modell predikciója szerint a fentebb említett fázis már szignifikánsan eltér, azonban Sóskuthy (2021) tanulmányára és személyes tanácsára hivatkozva ezen 8. ábrán piros sávval megjelenített eltérést kezeljük kitüntetett szereppel és bármely más 8. ábrán megjelenített eltéréstől eltérően. E döntés a szignifikánsan eltérő fázis nagyságára és a pozíciójára is alapozható: a tónuskontúrok mindössze 4%-át érintő eltérés aligha tekinthető megbízhatóan széles intervallumnak, amelyre a szignifikancia-jelölésében hagyatkozni lehetne. A prediktált eltérés inkább pontszerű különbségeket jelez, amelyek éppen abban a fázisban mutatkoznak, ahol az irreguláris zönge legnagyobb valószínűséggel feltételezhető, aminek révén a különbség relevanciája megkérdőjelezhető.

3.2.4. Az ereszkedő T4 tónus

A T4 lexikai tónus esetében mindhárom beszélői csoport esetében nyilvánvaló az f0-értékek ereszkedése a vizsgált szótag normalizált időtartamán belül. E lexikai tónus esetében a statisztikai próba eredményei szerint egyik nyelvtanuló csoport sem tért el a natív ejtéstől sem a kontúr f0-tartományában, sem a görbe alakjában (3. táblázat). Azonban a 7. ábrán azt láthatjuk, hogy a kínai anyanyelvi beszélők és a haladó nyelvtanulók f0-görbéi a normalizált időtartam legelején teljesen átfednek, majd megközelítőleg a normalizált időtartam

20%-ától a haladók egy árnyalatnyival alacsonyabb f0-átlagértékeket és ennek következtében egy fokkal kevésbé domború f0-görbéket produkálnak. Ehhez képest a kezdő nyelvtanulók f0-kontúrja a másik két csoporthoz viszonyítva kb. 1 félhanggal alacsonyabb f0-tartományból indul ereszkedésnek, azonban a kontúr záró fázisára az átlagértékek megközelítik a natív, illetve haladó csoport ejtési mintázatát (7. ábra, 8. ábra). Az ereszkedés mértéke a natív és haladó nyelvtanuló beszélőknél 8 félhangra (vagyis egy bővített kvintre/kis szextre) volt tehető, míg a kezdő nyelvtanulók esetében ez inkább 7 félhangnak (egy tiszta kvintnek) felelt meg.

Továbbá, ami összességében mind a négy lexikai tónus statisztikai eredményeivel kapcsolatban elmondható, hogy tónuskontúrtól függetlenül a beszélőkre és token-megvalósulásokra illesztett random smooth funkciók egytől-egyig szignifikáns eredményt hoztak (3. táblázat, $p < 0,001$), amely eredmények jelentős variabilitásra utalhatnak mind a beszélők mentén, mind pedig a külön-külön megvalósuló tónuskontúrok lefutását tekintve.

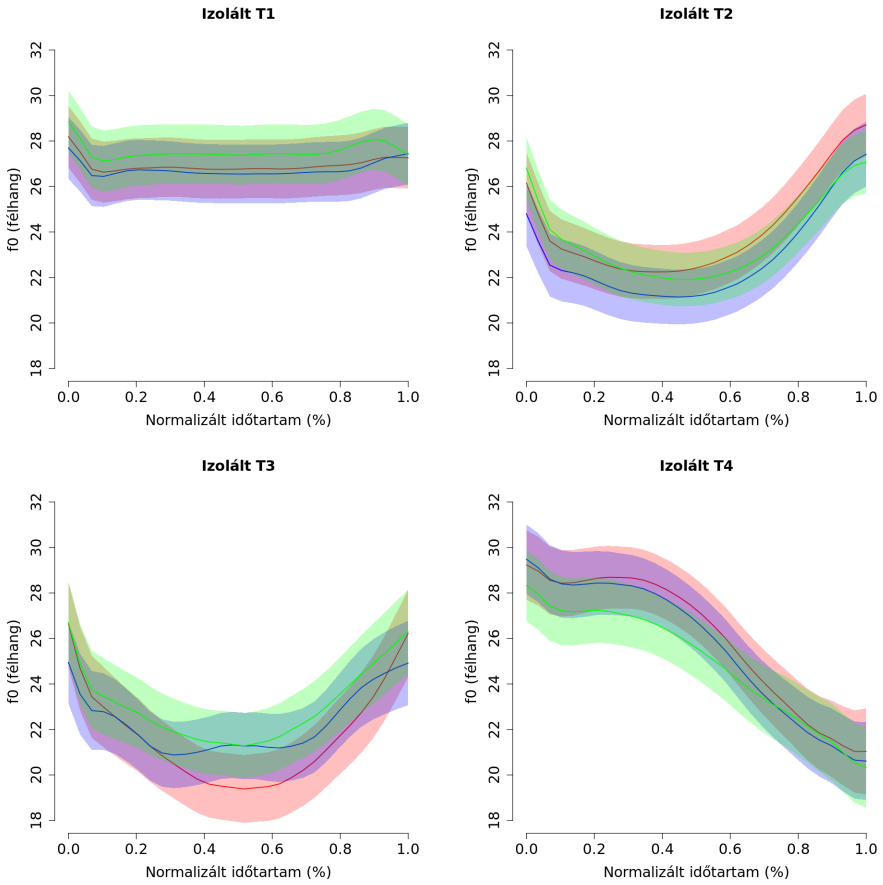
3. táblázat. A GAMM-ok parametrikus és smooth együtthatói (ahol a **beszélő** a beszélőnként, illetve a **traj.** a tokenenként illesztett random smooth-ot jeleníti meg és az n.s. jelölés a nem szignifikáns eredményt jelöli).

T1 ($R^2 = 98{,}6\%$)			
parametrikus együtthatók			
	becsült átlag	***t***	***Pr*(>|*t*|)**
natív	26,9	41,5	<0,001
haladó	26,7	−0,2	n.s.
kezdő	27,5	0,7	n.s.
smooth együtthatók			
	edf	***F***	***p***
natív	29,0	78,9	<0,001
haladó	5,6	2,2	<0,05
kezdő	7,8	7,2	<0,001
beszélő	183,4	17,1	<0,001
traj.	508,6	20,1	<0,001

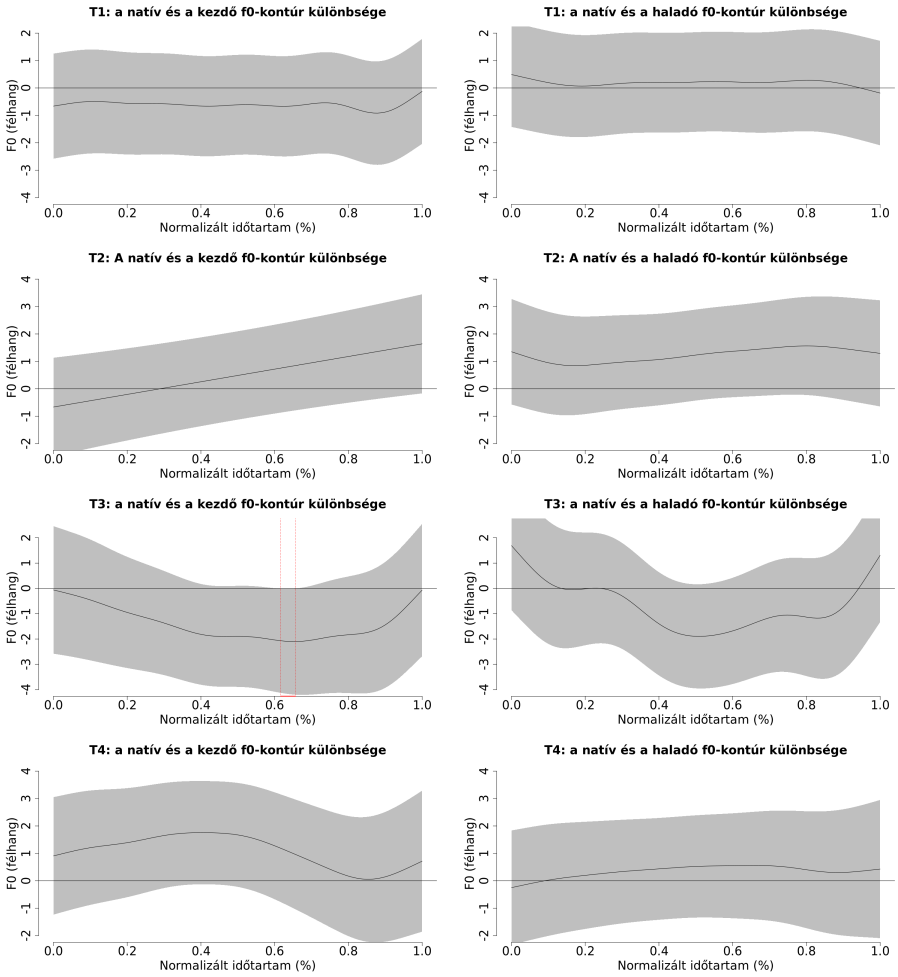
T2 ($R^2 = 99{,}0\%$)			
parametrikus együtthatók			
	becsült átlag	***t***	***Pr*(>|*t*|)**
natív	24,0	41,2	<0,001
haladó	22,8	−1,5	n.s.
kezdő	23,5	−0,6	n.s.
smooth együtthatók			
	edf	***F***	***p***
natív	29,1	306,6	<0,001
haladó	5,2	1,3	n.s
kezdő	6,6	3,0	<0,01
beszélő	165,6	5,7	<0,001
traj.	977,7	30,8	<0,001

T3 ($R^2 = 97{,}7\%$)			
parametrikus együtthatók			
	becsült átlag	***t***	***Pr*(>|*t*|)**
natív	21,9	33,6	<0,001
haladó	22,4	0,5	n.s.
kezdő	23,4	1,4	n.s.
smooth együtthatók			
	edf	***F***	***p***
natív	23,1	64,3	<0,001
haladó	7,4	13,1	<0,001
kezdő	5,6	2,03	<0,05
beszélő	155,4	7,1	<0,001
traj.	852,1	16,1	<0,001

T4 ($R^2 = 99{,}5\%$)			
parametrikus együtthatók			
	becsült átlag	***t***	***Pr*(>|*t*|)**
natív	26,8	40,8	<0,001
haladó	26,5	−0,4	n.s.
kezdő	25,7	−1,3	n.s.
smooth együtthatók			
	edf	***F***	***p***
natív	26,0	18,9	<0,001
haladó	3,7	0,4	n.s
kezdő	5,8	1,8	n.s.
beszélő	125,3	3,5	<0,001
traj.	1064,0	71,6	<0,001



7. ábra. A négy kínai lexikai tónus (T1 = balra felül, T2 = jobbra felül, T3 = balra alul, T4 = jobbra alul) GAMM-modellek eredményeként becsült f0-kontúrja, ahol a piros szín a natív kínai, a kék szín a haladó nyelvtanulók, míg a zöld szín a kezdő nyelvtanulók csoportjának produkcióját jelöli.



8. ábra. A két vizsgált nyelvtanulói csoport (haladók, kezdők) natív f0-kontúroktól szignifikánsan eltérő intervallumai a normalizált időtartam függvényében a négy kínai lexikai tónus ejtésében, ahol a bal oszlopban a kezdők, a jobb oszlopban a haladók differenciagörbéi láthatók.

4. Következtetések

A tanulmány az izolált ejtésű mandarin kínai lexikai tónusok produkciójának átfogó akusztikai fonetikai vizsgálatát tűzte ki célul a kínaiul tanuló magyar anyanyelvű beszélők ejtésében. A tanulmányban a négy kínai lexikai tónus produkcióját két akusztikai szempont szerint, vagyis a dallamívük menete, valamint

a megvalósulásuk időtartama szerint hasonlítottam össze két, eltérő mértékű L2 nyelvi tapasztalattal rendelkező kínaiul tanuló csoport ejtésében. A szótag egészének az időtartamát lineáris kevert modellekkel, az f0-görbék lefutását GAMM-okkal végeztem. Az első kutatási kérdés a lexikai tónusok időtartamára vonatkozott, vagyis hogy a kínaiul tanuló magyar anyanyelvűek a kínai beszélőkkel megegyező időtartammal ejtik-e a négy mandarin lexikai tónust. Illetve, ha a nyelvtanulók eltérnek a natív mintázatoktól, akkor az eltérésben megfigyelhető-e a célnyelvi tapasztalat mennyiségének hatása. A statisztikai elemzést a lexikai tónusok relatív időtartamán végeztem, mert az időtartam-értékek beszélők mentén való normalizációja visszaszorította a kísérleti személyek beszédtempójában megfigyelhető eltéréseket is. A lexikai tónusok időtartamában célnyelvi tapasztalattól függetlenül egyik nyelvtanulói csoport sem tért el a kínai natív ejtéstől. Ez azt jelenti, hogy a magyar nyelvtanulók képesek a natív mintának megfelelően produkálni a mandarin lexikai tónusok temporális sajátosságait, és ezen jelenségre nincs hatással a célnyelvi tapasztalat mennyisége. A relatív átlagos időtartamok kvalitatív eredményei abba az irányba mutatnak, hogy a rendszer szintjén mindkét nyelvtanuló csoport megközelítőleg fele olyan hosszan ejti a T4-et, mint az összehasonlítás referenciájaként szolgáló T3-at, tehát a natív beszélőkhöz hasonlóan különítik el e két tónust (de ezen állítást a jelen kísérlet empirikus úton nem erősítette meg). Tehát az izolált ejtésű kínai lexikai tónusokat mindkét nyelvtanuló csoport a natív időtartam-mintának megfelelően ejtette, azonban ezen tónusok időtartam alapján megvalósuló akusztikai elkülönítése nyitott kérdés marad. A négy kínai lexikai tónus közötti időtartam-kontraszt részletes, kvantitatív vizsgálatát egy előkészületben lévő tanulmányomban mutatom be. Mindenesetre egy rövid gondolat erejéig arra is érdemes kitérni, hogy mi teszi lehetővé azt, hogy a nyelvtanulók a natív mintázatnak megfelelő időtartamokkal ejtik a kínai tónusokat. Erre a jelenségre például magyarázattal szolgálhat az anyanyelvi mintázatok megfigyelése: habár a szerző nincs tudomással olyan összefoglaló tanulmányról, ahol a magyar egy szótagú megnyilatkozások időtartamát vetették volna össze akár percepciós, akár produkciós tekintetben, azonban például Juhász (2023) eredményei szignifikáns eltérést mutattak a fel-

szólító és a kijelentő monoszillabikus egységek időtartamában. Ezen eredmények – illetve emellett mindennapi tapasztalataink – alapján feltételezhető, hogy az f0-kontúr megvalósulása mellett a közlés időtartama is hozzájárulhat a jelentés kialakításához, és adott esetben feltételezhetően kontrasztív akusztikai tényező is lehet a magyar nyelvhasználatban. Ebből fakadóan számíthatunk arra, hogy a magyar anyanyelvűek érzékenyek ezen akusztikai jegy eltéréseire, ami elősegíthette a natív ejtésnek megfelelő időtartam-produkciót a kínai lexikai tónusok esetében. A jövőben mindenképpen érdemes lenne azon kutatási kérdés vizsgálata, hogy a magyar anyanyelvű beszélők észlelésében milyen hatással rendelkezik, és ejtésükben hogyan alakul a monoszillabikus megnyilatkozások időtartama.

A lexikai tónusok dallamívének megvalósulásával kapcsolatban kérdésként merült fel, hogy a magyar beszélők a kínai natív dallamkontúrokhoz hasonlót produkálnak-e vagy eltérnek a kínai dallamkontúroktól. Illetve eltérés esetében e akusztikai tulajdonság vizsgálatában is felvethető az a kérdés, hogy az eredményekben a nyelvi tapasztalat hatása is megfigyelhető-e, vagyis a haladók jobban megközelítik-e a natív mintázatokat, mint a kevesebb L2-tapasztalattal rendelkező kezdő nyelvtanulók.

A magas szinttartó T1 esetében azt láthattuk, hogy a dallammenet pozícionálása, pontosabban a kontúr f0-tartománybeli magassága nem okozott problémát a nyelvtanulóknak, hiszen mindkét csoport átlagosan ugyanolyan f0-értékekkel ejtette ezt a statikus magas lexikai tónust. Az f0-görbe alakját tekintve a haladók alapvetően nem tértek el a natív kínai ejtéstől, a két csoport között megfigyelhető eltérés nem érte el az 1 félhangot sem. A kezdő nyelvtanulók ejtésében a natív mintázatnál árnyalatnyilag (1 félhanggal) magasabb f0-értékeket láthattunk a teljes tónuskontúr megvalósulását illetően, azonban a két csoport f0-egyenese a normalizált időtartam egészében átfedett egymással. Ez azt jelenti, hogy mindkét nyelvtanulói csoport a natív ejtéssel megegyezően ejtette a T1-es tónust. A nyelvi tapasztalatra vonatkozó kutatási kérdés e tónus esetében megerősíteni látszik, hogy a haladók kisebb mértékben tértek el a natív mintától, mint a kevesebb nyelvi tapasztalattal rendelkező kezdő nyelvtanulók. A kezdő nyelvtanulók kínai anyanyelvi ejtéshez képest „túllőtt” magas akusztikai

célja (target overshoot) fakadhat a beszédfeladat minőségéből és a megfigyelés tényéből is. Empirikus kísérletek esetében a Labov-féle megfigyelési paradoxonra (1972) alapozva feltételezhetjük, hogy a beszélők máshogy beszélnek spontán helyzetben, és máshogy akkor, amikor éppen figyelik őket. A jelen kísérlet a laboratóriumi akusztikai kísérletek legkontrolláltabb formája, ahol a beszélő figyelme egyetlen szótag kiejtésére összpontosul, ráadásul idegen nyelven, úgy, hogy a kísérlet vezetője mellette ül és folyamatosan figyeli. Ezen körülmények sok beszélőben kelthetnek bizonyos mértékű szorongást a potenciális értékelés fényében (Horwitz et al., 1986), még akkor is, ha a kísérletvezetőnek ez nem célja és próbál a lehető legkellemesebb és legnyugodtabb környezetet teremteni. Emellett egy pillanatra visszatérve a laboratóriumi kísérlet feladattípusára, az izolált ejtésű felolvasási kísérlet esetében – mint korábban már említettem – hiperartikulált ejtésre számíthatunk, vagyis a prominens akusztikai jegyek túlzó megvalósítására (vö. Scarborough & Zellou, 2013). A kezdő nyelvtanulók esetében a T1-re vonatkoztatva ez feltételezhetően azt jelentette, hogy a T1 legprominensebb, magas f0-tartományban megvalósuló akusztikai tulajdonságait túlartikulálták és ezért valósult meg a kontúr a natív ejtésnél is magasabb f0-értékekkel.

Az emelkedő T2 tónus elsajátítása is sikeres volt mindkét nyelvtanuló csoport esetében: a haladó nyelvtanulók minden tekintetben a natív mintának megfelelően ejtették e kínai lexikai tónust, mégis a haladók teljes f0-kontúrja a natív mintánál kb. 1,5 félhanggal alacsonyabban realizálódott. A natív ejtésnél alacsonyabb f0-értékek, de a megegyező alakú f0-görbe megvalósulása nem az anyanyelvi magyar mintázatok eredményének tűnik, hiszen ha a magyar egy szótagú eldöntendő kérdő dallamot vesszük alapul, a tónus-görbe jelentős hányadát a kínai tónusnál magasabb f0-értékek jellemezték, szemben a haladók ejtésében megvalósuló alacsony f0-értékekkel. Ebben az esetben talán inkább egy másik irányban érdemes a mintázatok mögött megbúvó motivációt kutatni. Habár a jelen tanulmánynak nem célja a tónusok közötti kontraszt vizsgálata, azonban a bemutatott eredmények egyértelműen felvetik ennek a kérdésnek a relevanciáját. Mind a haladó, mind a kezdő kínaiul tanulók a T2-t a natív

kínai mintánál alacsonyabb f0-értékekkel, míg a T3-at pedig a natív mintánál magasabb f0-értékekkel ejtették. Ezek az akusztikai tulajdonságok felvetik a kérdést – amely kérdés vizsgálatára egy következő tanulmányban kerítek sort – hogy vajon nem a T2 és a T3 ejtésbeli megkülönböztetésének hiánya indukálja-e ezeket a nyelvtanulói mintázatokat. Hiszen ahogy Hao (2012) is összefoglalta: a mindkét tónusra jellemző ereszkedő-emelkedő mintázat könnyen eredményezheti azt, hogy e két tónus nem különül el megfelelően produkciósan (és talán percepciósan sem). A T2 és a T3 tónus esetében megfigyelhető mintázatok is összhangban állnak Hao (2012) eredményeivel, hiszen e két tónus esetében a natív ejtéstől való eltérés inkább 2 félhangra volt tehető, míg például a T1 esetében ez nagyjából maximum 1 félhang volt, vagyis a T2 és a T3 esetében jelentősebb eltérések figyelhetők meg a magas szinttartó T1-hez képest.

Az ereszkedő-emelkedő T3 esetében azt láthattuk, hogy mindkét nyelvtanuló csoport görbéjének f0-terjedelme a kínai anyanyelvű ejtéshez képest kompresszálva, és kevésbé homorúan valósult meg, amit a natív ejtéshez képest magasabb minimális f0-érték idézett elő. Továbbá, a T3 esetében azt láthattuk, hogy a natív ejtés f0-görbéje a relatíve alacsony f0-értékek ellenére is egyenletesen homorú mintázatot mutatott, szemben például a haladó nyelvtanulók tónuskontúrjával. Ezen ejtési tulajdonság például annak a képzési jegynek is tulajdonítható, hogy míg a natív beszélők esetében az irreguláris zönge produkciója jól kontrollált, addig a haladó nyelvtanulók esetében felmerülhet, hogy az irreguláris zöngeképzés révén létrejövő f0-értékek sokkal kevésbé kontrolláltak, vagyis zajosabbak az adatok. A kezdő nyelvtanulók esetében az f0-kontúr magasabb f0-tartományban valósult meg, így feltételezhetően az irreguláris zönge indukálta hatás is visszafogottabb a haladó nyelvtanulók ejtéséhez képest. Az irreguláris zönge akusztikai elemzése egy következő elemzés kutatási kérdéseként jelenhet meg, vizsgálva az irreguláris zönge intervallumának pozícióját, hosszát és az f0-értékek minőségét a kínaiul tanuló magyar anyanyelvűek ejtésében. Mindent összevetve azonban a haladó nyelvtanulók – még ha az f0-görbe alakja szempontjából nem is, de – jobban megközelítették a natív ejtést, mint a kezdő nyelvtanulók, így tehát e tekintetben a Flege- és Bohn-féle predikció

e tónus esetében is megerősítést nyert. Ha a T3-hoz kapcsolódó eredményeket a szakirodalmi forrásokkal hasonlítjuk össze, akkor a Chao (1948/1963) által lejegyzett bővített kvart (tehát 6 félhang) helyett 9 félhangnyi f0-terjedelmet figyelhetünk meg. Ezzel szemben például Chun és munkatársai (2015: 3) mérésében a megközelítőleg 160 Hz-es minimumértéket és a 280 Hz-es maximumértéket 50 Hz-es referenciaértékkel félhangokká konvertálva 20 és 29,8 félhangot kapunk, amely f0-terjedelem megegyezik a jelen kísérletben bemutatott eredményekkel. Ezen összehasonlítás alapján felmerül a kérdés, hogy vajon a Chao-féle eredmények (1948/1963), illusztrációk relatíve régi, illetve saját megfigyelésen nyugvó leírásainak használata pozitívan járulhat-e hozzá a kínaiul tanulók lexikaitónus-produkciójához. E kérdést a T4 nyelvtanulói produkciója is felveti.

Az ereszkedő T4 esetében is megerősíthető a Flege- és Bohn-féle (2021) predikció arra vonatkozóan, hogy a haladó nyelvtanulók jobban megközelítik az anyanyelvi mintázatot a kevesebb L2-tapasztalattal rendelkező kezdő nyelvtanulókhoz képest, hiszen a kezdők alacsonyabb f0-tartományból ereszkedve kisebb terjedelmű kontúrt produkáltak a natív beszélőkhöz képest. Ezeket az eredményeket potenciálisan magyarázhatjuk az anyanyelvi magyar felszólító dallamkontúr hatásával, hiszen az alacsonyabb f0-terjedelemből indul ki a kínai T4-hez képest, azonban e lehetséges hatás szerepe elég visszafogott, hiszen mindössze egy félhang különbséget eredményez. Ezen ejtési tulajdonságot akár a lexikai tónusok számkódjainak hatására is visszavezethetjük: a bevezetés részben említettem, hogy ellentétben azzal, hogy a T4 a T1-hez képest magasabb f0-tartományból indul, a kínai nyelvtanításban mindkét tónus 5-ös hangmagassági szinttel van jellemezve az 5 fokozatú skálán, ami negatív hatással lehet a nyelvtanulók produkciójára. Ebben az esetben elképzelhető, hogy pont ezt a hatást figyelhetjük meg: mind a T1, mind a T4 megközelítőleg 28,5 félhang magasságból indul, tehát a két tónus kezdeti akusztikai célja a kezdő nyelvtanuló csoportok ejtésében megegyezik. E megfigyelés megerősítéséhez azonban a már korábban említett tónus-kontrasztok vizsgálata szükséges. Ugyanakkor visszatérve a lexikai tónusok elsajátításának relatív nehézségére a nyelvtanulók körében, a T4 esetében a T1-hez hasonló mértékű eltéréseket figyelhettünk meg,

amely eltérés a haladók esetében kisebb volt mint 1 félhang, és a kezdő nyelvtanulók esetében sem érte el a különbség a 2 félhangot (szemben a T2 és a T3 esetében tapasztaltakkal, ahol akár 2 félhangot meghaladó eltérések is megfigyelhetők voltak). Mindezen eredmények összhangban látszanak lenni Hao (2012) eredményeivel, amelyek szerint a T1 és a T4 elsajátítása sikeresebb a T2-höz és a T3-hoz képest, hiszen utóbbi esetekben jelentősebb eltéréseket figyeltünk meg a natív mintázathoz képest mint a T1 és a T4 esetében. A fentebbiek alapján összegezve a célnyelvi tapasztalat jelentőségére vonatkozó kutatási kérdést, azt mondhatjuk el, hogy minden tónus esetében megfigyelhető volt az a mintázat, hogy a haladók kisebb eltérésekkel, vagyis jobban megközelítik a natív mintázatokat, mint a kevesebb célnyelvi tapasztalattal rendelkező kezdő nyelvtanulók. Tehát összességében a kutatás eredményei az f0-kontúrok esetében megerősítik Flege és Bohn (2021) predikcióját, miszerint a nyelvi tapasztalat növekedésével a nyelvtanulók célnyelvi teljesítménye jobban megközelíti a natív célnyelvi mintázatokat. Ezzel szemben az időtartam tekintetében nem látjuk ezen eltéréseket a nyelvtanulói csoportok között a lexikai tónusok produkciójában. E jelenség alapján az feltételezhető, hogy nem meglepő módon a temporális sajátosságok, pontosabban a tónus időtartamának elsajátítása könnyebb és hatékonyabban megy végbe a nyelvtanulók számára, szemben a hangszalagműködés finommotorikán nyugvó beállításával és szabályozásával. Habár az időtartam is skaláris tulajdonságként számszerűsíthető, azonban ez esetben kizárólag a zöngeképzés időtartamát (kezdetét/fenntartását/megszűnését) kell szabályozni, míg a dallamkontúrok esetében egy sokkal bonyolultabb akusztikai célt kell megtalálni, amelyhez a jelen kísérletben – az izolált ejtésű megnyilatkozásokból fakadóan – semmilyen kontextuális segítség nem állt rendelkezésre. Ezek az eredmények közvetetten összhangban állnak Flege és Bohn (2021) Revised speech learning modeljében leírtakkal, miszerint a komplexebb mintázatok (e modell esetében ez elsősorban beszédhangokra értendő) elsajátítása nehezebben megy végbe az egyszerűbb mintázatokhoz képest. A tanulmányban feltárt akusztikai eltérések jelentősége a jövőben percepciós tesztek segítségével vizsgálható. Továbbá kérdésként merül fel, hogy a négy kínai tónus – a dallamok rendszere szintjén

– mennyire különül el a kínaiul tanuló magyar anyanyelvűek ejtésében, amely kérdésre egy előkészületben lévő tanulmányban keresem a választ. A kísérlet eredményei hozzájárulnak a kínai tónusok elsajátításának és ejtési problémáinak mélyebb megértéséhez és elősegíthetik a kínai mint idegen nyelv oktatását.

Köszönetnyilvánítás

Hálával tartozom Sóskuthy Mártonnak, akinek a rendkívül részletes és segítőkész bírálatát olvasva sokat tanulhattam többek között a GAMM-ok használatáról és működéséről, amely ismereteket a tanulmány statisztikai módszertanában is alkalmaztam. Általánosságban köszönettel tartozom a tanulmány bírálóinak, akik javaslataikkal és kérdéseikkel segítettek érthetőbbé és fókuszáltabbá tenni a tanulmányt. A Kulturális és Innovációs Minisztérium EKÖP-24 kódszámú Egyetemi Kiválósági Ösztöndíj Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.

Hivatkozások

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

Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer. Computer program]. 6.3.03-as verzió. (letöltés ideje: 2022. december 17.).

Bolla, K. (1995). *Magyar Fonetikai Atlasz*. Budapest: Nemzeti Tankönyvkiadó.

Chao, Y. (1948/1963). *Mandarin Primer*. Cambridge: Harvard University Press. URL: https://doi.org/10.4159/harvard.9780674732889. doi:10.4159/harvard.9780674732889.

Chun, D., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of l2 mandarin chinese tones with learner-created tone visualizations. *Journal of Second Language Pronunciation*, *1*, 86–114. doi:10.1075/jslp.1.1.04chu.

Dallaston, K., & Docherty, G. (2019). Estimating the prevalence of creaky voice: a fundamental frequency-based approach. In *International Congress of Phonetic Sciences (ICPhS* (pp. 532–536). Melbourne, Australia.

Dorreen, K. (2017). *Fundamental frequency distributions of bilingual speakers in forensic speaker comparison. Mesterszakos szakdolgozat*. Christchurch: The University of Canterbury.

Flege, J. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233–277). Timonium: York Press.

Flege, J., & Bohn, O. (2021). The revised speech learning model (slm-r. In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge: Cambridge University Press. doi:10.1017/9781108886901.002.

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

Hao, Y.-C. (2012). Second language acquisition of mandarin chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*, 269–279. doi:10.1016/j.wocn.2011.11.001.

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental phonetic approach to speech melody*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511627743.

Horwitz, E., Horwitz, M., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*, 125–132.

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

Juhász, K. (2024). A mandarin beszédhangok produkciója kínaiul tanuló magyar anyanyelvűek ejtésében.

Juhász, K. (előkészületben). Isolated mandarin chinese lexical tone production: a systemic approach analysing tone contrast within l2 learner groups.

Klein, W. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.

Kuhl, P. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. Yeni-Komshian, J. Kavanagh, & C. Ferguson (Eds.), *Child phonology – perception* (pp. 41–66). New York: Academic Press.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. doi:10.18637/jss.v082.i13.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Leather, J., & James, A. (1991). The acquisition of second language speech. *Studies in Second Language Acquisition*, *13*, 305–341. doi:10.1017/S0272263100010019.

Lee, Y.-S., Vakoch, D., & Wurm, L. (1996). Tone perception in cantonese and mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*, 527–542.

Lenth, R. (2020). Emmeans: Estimated marginal means, aka least-squares means. URL: https://CRAN.R-project.org/package= letöltés:.

Major, R. (2001). *The foreign accent: The ontogeny and phylogeny of second language phonology*. New York: Routledge. doi:10.4324/9781410604293.

Markó, A. (2013). *Az irreguláris zönge funkciói a magyar beszédben*. Budapest: ELTE Eötvös Kiadó.

Markó, A. (2017). Hangtan. In A. Imrényi, N. Kugler, M. Ladányi, A. Markó, S. Tátrai, & G. Nagy (Eds.), *Nyelvtan* (pp. 75–203). Budapest: Osiris Kiadó.

Mennen, I. (2014). Beyond segments: towards a l2 intonation learning theory (lilt. In E. Delais-Roussarie (Ed.), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*. New York: Springer. doi:10.1007/978-3-662-45168-7_9.

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. In *Proceedings of 15th ICPhS 2003* (pp. 771–774). Bacelona: Universitat Autònoma de Barcelona.

Odlin, T. (1989). *Language transfer*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524537.

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

R Core Team (2024). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Letöltés*, . URL: https://www.R-project.org/. doi:10.32614/CRAN.package.hqmisc.

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.

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, *134*, 3793–3807. doi:10.1121/1.4824120.

Shen, X. (1989). Interplay of the four citation tones and intonation in mandarin chinese. *Journal of Chinese Linguistics*, *17*, 61–74.

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 1–19. doi:10.1016/j.wocn.2020.101017.

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The lingu-

istic proximity model. *International Journal of Bilingualism*, *21*, 666–682. doi:10.1177/1367006916648859.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. doi:10.1007/978-3-319-24277-4.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R*. New York: Chapman and Hall/CRC. doi:10.1201/9781315370279.

Xu, Y. (1997). Contextual tonal variations in mandarin. *Journal of Phonetics*, *25*, 61–83. doi:10.1006/jpho.1996.0034.

Ye, Q. (2013). Pronunciation problems of hungarian university students learning chinese: Analysis, causes and possible solutions.

Zhang, H., Wiener, S., & Holt, L. (2022). Adjustment of cue weighting in speech by speakers and listeners: Evidence from amplitude and duration modifications of mandarin chinese tone. *The Journal of the Acoustical Society of America*, *151*, 992–1005. doi:10.1121/10.0009378.

Az *izé* strukturális pozíciója és diskurzusjelölő funkciója spontán, baráti társalgásokban

Tar Cintia

SZTE BTK Általános Nyelvészeti Tanszék
SZTE Nyelvtudományi Doktori Iskola

Abstract

Hungarian native speakers often consider *izé* as a stigmatized, functionless filler word (Grétsy & Kovalovszky, 1980); however, it plays an important role in spontaneous conversations and speech planning processes. Previous literature has revealed that *izé* has two types, namely, word-substituting and time-gaining *izé*s. While the former can address word-retrieval problems and vocabulary shortages, the latter can help solve issues in speech planning (Fabulya, 2007; Gyarmathy, 2012; Gyarmathy & Neuberger, 2013; Gyarmathy, 2015; Kondacs, 2017; Marcsenkoné Kondacs, 2023). The study investigates the phenomenon from a functional discourse analytic and conversation analytic perspective, using both qualitative and quantitative methods. It examines the structural position of *izé* in Hungarian conversations. The corpus under study comprises audio recordings of Hungarian spontaneous, naturally occurring, and friendly conversations, collected and analyzed by the author. The analysis differentiates between five structural categories and the functions of word-substituting and time-gaining *izé*s in these positions. It seeks to answer the question of whether there are additional functions beyond those identified so far. It examines whether *izé* can be used in a quotative function in the Hungarian corpus, similar to the *be + like* formula in English. The analysis shows that *izé* can have a discourse-organizing, discourse-marking function, and it can influence the turn-taking system.

1. Bevezetés

1.1. A kutatás elméleti háttere

A magyar pszicholingvisztika és fonetika szakirodalma a megakadásjelenségeket két csoportra osztja: (1) bizonytalanságból fakadó és (2) hiba típusú megakadásjelenségek. A bizonytalanságból fakadó megakadásjelenségek közé tartozik az ismétlés, az újraindítás és az általuk töltelékszónak nevezett, „a közlésbe szervetlenül beékelődő, tartalmilag nem illeszkedő szavak vagy szókapcsolatok”

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

Beszédtudomány – Speech Science 2025/2. 44–81. DOI: 10.15775/Besztud.2025.2.44-81

(Gyarmathy, 2017, 88). Ez utóbbi kategóriába sorolják az *izé* nyelvi elemet is, melynek funkciója az átmeneti produkciós nehézség jelzése és a tervezési diszharmónia feloldása (Gósy, 2005; Gyarmathy, 2015, 2017). Gyarmathy (2017) szerint a megakadásjelenségek a beszéd azon jelenségei, amelyek megszakítják a spontán beszéd artikulációs és percepciós folyamatát, tervezési bizonytalanságra, illetve a folyamat hibás működésére utalhatnak, és nem társítható hozzájuk egyértelmű pragmatikai funkció. Azt azonban fontos kiemelni, hogy a terminus nem teljesen értéksemleges, hiszen azt az érzetet kelti, mintha egyfajta „rendellenes" jelenség lenne. Ezek a megakadások vagy diszfluenciák pedig a társalgás olyan természetes velejárói, amelyek a fluens beszéd érzetét idézik elő, sőt hozzájárulnak a beszédtervezés hatékonyságához (Bóna, 2023).

A pszicholingvisztika és fonetika megakadásjelenségekkel foglalkozó területének megközelítése összecseng a konverzációelemzés javítási jelenségeinek meghatározásával. A konverzációelemzés szerint a javítási jelenségek megszakítják a beszédcselekvés progresszivitását azért, hogy kezeljenek egy, a produkció, észlelés vagy megértés során felmerülő hibát, problémát (Schegloff et al., 1977). A konverzációelemzés keretében a javítási folyamatban három részt különíthetünk el: a problémaforrást, a javításkezdeményezést és a javítás végrehajtását. A problémaforrás azokat a beszédszakaszokat jelzi a társalgásban, ahol felmerülnek a javításra szoruló problémák, melyek megszakítják az interakciós folyamatot (Schegloff et al., 1977). A konverzációelemzés javítási jelensége hasonló a pszicholingvisztika és a fonetika megakadásjelenségének meghatározásához, azonban mégsem egyenlő azzal. Míg a megakadásjelenség maga akasztja meg az artikulációs és percepciós folyamatot, addig a problémaforrás a beszédcselekvés megszakításának helyét jelöli, és nem egy konkrét nyelvi jelenséget.

A problémaforrás és a javítás fogalmáról szintén elmondható, hogy nem teljesen mentesek az értékítéletektől, hiszen azt sugallhatják, hogy a spontán beszédben előforduló diszfluenciák valamiféle hibát jelentenek, amelyeket „korrigálni" kell. A konverzációelemzés ugyanakkor kiemeli, hogy ez a jelenség nem csupán a nyilvánvaló „hibákra" értendő, hanem a beszélők a segítségével gyakran finomra hangolják a fordulóikat és biztosítják a beszéd folytonosságát (Kitzin-

ger, 2013). Ebben eltér a pszicholingvisztika és fonetika szakirodalma, amely javításnak a hiba típusú megakadásjelenségeket, azaz a téves kivitelezést tekinti (pl. grammatikai hiba, malapropizmus, anticipáció, perszeveráció, metatézis) (Gyarmathy, 2015; vö. Gósy, 2005).

Fabulya (2007) szerint az *izé* egy javításkezdeményező lexikális kitöltőelem, amely maga is javításra szorulhat bizonyos esetekben (lásd: szóhelyettesítő *izé*). A konverzációelemzés keretében a beszélő a javításkezdeményezéssel jelzi, hogy valamilyen nehézség merült fel a társalgás során (Schegloff et al., 1977). Az *izé* azokban az esetekben lehet javításkezdeményező, ha a beszélő egyértelműen indikálja ezt a nehézséget, például nem jut eszébe egy keresett kifejezés (lásd: szókereső *izé*), vagy egy nagyobb szerkezeti egység megformálásához több időre van szüksége (lásd: időnyerő *izé*) (vö. Fabulya, 2007). Mivel az *izé*vel szinte bármilyen szót, kifejezést képesek vagyunk helyettesíteni (Gyarmathy, 2015), így az esetek nagy részében a feloldása elengedhetetlen feltétele az interszubjektivitás, vagyis a valóság közös értelmezése fenntartásának vagy helyreállításának (vö. Fabulya, 2007; Schegloff et al., 1977). Ennek alapján az *izé* javításra szoruló elemnek is tekinthető, hiszen a használata problémát okozhat a megértés során (vö. Schegloff et al., 1977).

Az *izé* nem csupán javításkezdeményező és javításra szoruló elem lehet, hanem maga a javítás végrehajtása is, például, ha tabukerülő funkciót tölt be a társalgásban. Lássunk erre egy példát! Vannak olyan tabukifejezések, amelyeket egy adott szituációban nem illendő kimondani, ilyenek például a trágár kifejezések vagy a szexualitással kapcsolatos kifejezések. Annak érdekében, hogy a beszélők elkerüljék a tabukifejezéseket egy társalgási szituációban, gyakran az *izé*vel pótolják azt (Fabulya, 2007). Ilyenkor nem szükséges az *izé* feloldása, hiszen a kontextusból kikövetkeztethető a jelentése. Ez összecseng a konverzációelemzés delicate típusú szókeresés javítási műveletével, (vö. Tar, 2023), amely esetben a beszélő tisztában van azzal, hogy kényes, tabukifejezést fog mondani, ezért olyan hezitációs jelenségeket produkál, mint a szünetek és az *öö*zés, amellyel kifejezi a kifejezéshez való viszonyát, vagyis, hogy kényesnek tartja azt (Lerner, 2013). A tabukerülő *izé*t abban az esetben tekinthetjük javításnak,

ha az azt megelőző beszédegység ilyen vagy ezekhez hasonló hezitációs jelenségeket tartalmaz, amelyek megszakítják a beszédcselekvés progresszivitását, és a beszélő az *izé* használatával kezeli azt a nehézséget, hogy nem produkálhat tabukifejezést. Ha az *izé* előtt nincsen hezitáció, azaz nem szakad meg a progresszivitás, akkor nem javítási jelenség.

Láthattuk, hogy a pszicholingvisztika és a fonetika az *izé*t bizonytalanságból fakadó megakadásjelenségként, a konverzációelemzés pedig lexikális kitöltőelemként értelmezi. Ezek a kitöltőelemek polifunkcionalitás tekintetében hasonlóak a diskurzusjelölőkhöz (Kosmala & Crible, 2021), azonban számos különbségről számol be a szakirodalom. A kitöltőelemekhez képest a diskurzusjelölők olyan kifejezések, amelyek a diskurzus szerkezetének és magának az interakciónak a kezelésére szolgálnak (Kosmala & Crible, 2021). Schiffrin (1987) szerint olyan szekvenciálisan függő elemek, amelyek a beszélgetés egységeit kapcsolják össze. Diewald (2013) ezzel szemben megkülönböztet két iskolát: az első iskola szerint a diskurzusjelölők szintaktikailag függő elemek, azaz konnektívumok (pl. Fraser, 1999), a második iskola szerint pedig a diskurzusjelölők szintaktikailag független elemek, amelyek a megnyilatkozás szerkezetéhez kapcsolódnak. Ez utóbbi iskola szerint a diskurzusjelölőség feltétele egyfelől a diskurzusszervező funkció, amely irányítja a diskurzust, érintheti a szekvenciális szerkezetet és szabályozhatja a beszélőváltást is (Diewald, 2013).

A diskurzusjelölők és a kitöltőelemek viszonyáról számos nézet alakult ki. Egyes munkák szerint a kitöltőelemek a diskurzusjelölők egy alcsoportját alkotják az alapján, hogy rendelkeznek-e diskurzusszervező funkcióval (pl. Davis & Maclagan, 2010). Fox (2010) szerint vannak olyan diskurzusjelölők, melyek a kitöltőelemek késleltető funkciójával bírnak, a magyarban ilyen lehet a *hát* és a *szóval* diskurzusjelölő (Schirm, 2011). A kitöltők a diskurzusjelölők egy típusát is alkothatják, és egy elhalasztott elemet helyettesíthetnek (lásd: szóhelyettesítő *izé*) (Podlesskaya, 2010). Egy adott elem kategorizálása során fontos, hogy mindig az adott kontextusban betöltött funkció a döntő a diskurzusjelölőség vagy kitöltőelemség kérdésében. Ha az adott elem például halasztó funkciót tölt be

az adott társalgásban, akkor kitöltőelem, ha pedig, tegyük fel, kezdő társalgási cselekvést vezet be, akkor diskurzusjelölő (Hayashi & Yoon, 2010).

Jelen tanulmányban az *izé*t egy olyan nyelvhasználati jelenségként értelmezem, amely funkciója tekintetében többnyire lexikális kitöltőként viselkedik, ugyanakkor megkísérelem feltárni, rendelkezik-e diskurzusszervező funkcióval, azaz képes-e diskurzusjelölői szerepet felvenni.

1.1.1. Az izé típusai és funkciói

A lexikális kitöltők, köztük az *izé* fontos szerepet töltenek be a társalgásokban, kifejezetten a spontán társalgásokban. A beszédtervezés során felmerülő nehézségek gyakran kitöltetlen szüneteket eredményeznek a beszélgetésekben, hiszen a beszélőnek több gondolkodási időre van szüksége a soron következő egység(ek) előhívásához (Horváth, 2014). Ha a kitöltetlen szünet váltásreleváns helyre esik, vagyis a cselekvés lehetséges lezárási pontjára, veszélyezteti a beszélőt, hogy elveszíti a beszédhez való jogát (vö. Kitzinger, 2013), ezt pedig az *izé* szó használatával meg tudja akadályozni (Fabulya, 2007). Ha a kitöltetlen szünet nem váltásreleváns helyen van, hanem még a cselekvés lehetséges lezárási pontja előtt, akkor a beszélő az *izé* szó használatával biztosítja partnereit, hogy eleget tesz annak a kötelezettségnek, hogy egy megszólalással legalább egy felismerhető cselekvést kell végrehajtania (vö. Sacks et al., 1974; Schegloff, 2007).

Váltásreleváns helyen, annak érdekében, hogy a beszélők a szóátvétel fenyegetettségét elkerüljék és megtartsák a beszédhez való jogot, a szünet kitöltésére van szükség (Fabulya, 2007). A szünet kitöltése történhet különböző kitöltőkkel, melyek lehetnek nemlexikális (pl. hangnyújtás, *öö*), kvázi-lexikális (pl. *uhm, hmm*), illetve lexikális kitöltőelemek, ez utóbbihoz tartozik az *izé* szó is (vö. Rieger, 2003).

A legtöbb nyelvben a lexikális kitöltőelemek névelők, mutató vagy kérdő névmások. Számos nyelvben találhatunk példát a *hogyishívják, hogymondjam* típusú kitöltőelemekre. Ezek olyan lexikalizálódott konstrukciók, amelyek általában kérdőszó és névmás vagy kérdőszó és ige összetapadásával keletkeznek

(pl. angol: *whatchamcallit* ‘hogynevezzem’; kuwaiti arab: *šisma* ‘what-name-his’) (Fox, 2010).

Vannak nyelvek, amelyekben a kitöltőelemek különálló lexikális elemmé fejlődtek. Ezek olyan határozatlan helyettesítőszók, melyeknek főnévi eredetük gyakran nyomon követhető (Fox, 2010). Idetartozik a magyar *izé* is, amely egy bizonytalan eredetű, esetleg ősi finnugor kori szó magyar képzéssel, eredeti jelentése feltehetőleg ‘dolog’, ‘valami’ (Zaicz, 2006). További példák még a baszk *zera* (vö. Urizar & Samuel, 2014), a nganaszan *əhtы* ‘dolog’ vagy az olasz *coso* ‘dolog’ (Fox, 2010).

Fabulya (2007) az *izé*, a *hogyishívják* és a *hogymondjam* lexikális kitöltőelemeket és azok feloldását Levelt (1989) beszédprodukciós modellje alapján vizsgálta. Megállapításai szerint az *izé* megjelenésének okai lehetnek (1) mentális lexikonból való előhívási nehézségek, (2) a lexikai reprezentáció hiánya, illetve (3) a nagyobb egységek megformálásához szükséges időnyerés. Funkcióik szerint kétféle *izé*t különböztet meg. A szóhelyettesítő *izé* konkrét szavak, kifejezések helyettesítésére szolgál. A másik csoportba azok az *izé*k tartoznak, amelyeket a beszélő egy bizonyos kifejezés vagy egy összetett szerkezeti egység megformálásához szükséges időnyerésre használ. Ezt Fabulya (2007) töltelék *izé*nek nevezi. Jelen tanulmányban amellett érvelek, hogy az *izé* megjelenésének minden esetben jól meghatározható oka és funkciója van a társalgásban, ezért a töltelék *izé* helyett az időnyerő *izé* terminust fogom használni. A töltelék vagy beszédtöltelék a szó nyelvművelői attitűd szerinti „felesleges” mivoltát sugallja, amely nem összeegyeztethető a dolgozatom céljával.

A szóhelyettesítő *izé*k átmenetileg vagy véglegesen helyettesítenek más kifejezéseket egy már megformált szintaktikai szerkezetben (Fabulya, 2007). A beszélő megformálja a fordulóját, sikeresen produkálja azt úgy, hogy közben egy elem előhívása során valamilyen nehézség lép fel. Ezt a nehézséget az *izé* szó használatával hidalja át.

A szóhelyettesítő *izé* tovább bontható két alcsoportra. Azokban az esetekben, amikor a beszélő egyes lexémákat helyettesít, amelyek átmenetileg nem elérhetőek a mentális lexikonból, szókereső *izé*nek nevezzük. Megfigyelhető,

hogy a szókereső *izé* gyakran toldalékolt alakban jelenik meg a társalgásban (Fabulya, 2007). Ha feloldásra kerül a sor, az *izé*t feloldó szón azonos toldalék jelenhet meg, azaz azonos szintaktikai szerepet töltenek be a társalgásban. Innen könnyen felismerheti a recipiens, mit helyettesít a beszélő.

A másik alcsoportba azok az *izé*k tartoznak, melyek egy, a beszélő mentális lexikonjából feltehetőleg hiányzó szót/kifejezést pótolnak. Ez a hiány, a szókereső *izé*vel ellentétben nem átmeneti, hanem egy teljesen hiányzó lexémáról van szó (Fabulya, 2007). Ennek feloldása általában körülírással történik, amelyhez gyakran különböző gesztusok is társulnak.

A következő táblázatban a szóhelyettesítő *izé* két csoportját mutatom be Fabulya (2007) alapján:

1. táblázat. A szóhelyettesítő *izé*k típusai (Fabulya, 2007).

szóhelyettesítő *izé*	
szókereső *izé*	**szópótló *izé***
átmenetileg	véglegesen
nehezebben érhető el a mentális lexikonból	nem érhető el a mentális lexikonból/nincs benne a mentális lexikonban/nem lehet egy szóval kifejezni
ritkán használt kifejezések, konkrét nevek	szókincshiány/nem létező kifejezés
feloldás: - keresett kifejezés megadása - közvetlenül - későbbi ponton	feloldás: - körülírás [rövid, frappáns / (tag)mondatnyi] - újrafogalmazás
feloldás hiánya: kontextusból kiderül (de általában indokolják, pl. *nem jut eszembe*)	feloldás hiánya: - kontextusból kiderül (nyelvi ökonómia) - tabu

Az időnyerő *izé*k olyan beszédtervezési nehézség esetén jelennek meg, amikor a beszélő még nem formálta meg teljesen a produkálni kívánt szerkezetet, a felépítése még hiányos állapotban van (Fabulya, 2007). Gósy (2004) szerint a beszélő nem tudja, hogyan fogalmazza meg a mondanivalóját, ezen a ponton még a tartalmi elemek válogatásának fázisában van (Gósy, 2003). A beszélőnek

több gondolkodási időre van szüksége a nagyobb szerkezet megformálásához, amely idő kitöltésére az *izé* mint lexikális kitöltőelem alkalmas (Fabulya, 2007). Az időnyerő *izé*t gyakran más késleltetésre alkalmas nemlexikális, kvázi-lexikális és lexikális kitöltőelemek kísérhetik (Fabulya, 2007; vö. Rieger, 2003). Az időnyerő *izé* mindig esetrag nélkül áll, azaz nem illeszkedik a forduló szintaktikai szerkezetébe (Fabulya, 2007). Megfigyelhető az is, hogy ez a jelenség gyakran mellékmondatok elején áll, a mellékmondatot bevezető kötőszó közelében, szomszédságában (Fabulya, 2007). Ebben nagy mértékű hasonlóságot mutat az *öö* kitöltőelemmel, amelyről Németh (2020) azt állapította meg, hogy a beszélők a kötőszó utáni tagmondatot egy egységként kezelik, és az *öö* használatával ennek a tagmondatnyi egységnek a megformálásához nyernek gondolkodási időt. Ez arra enged következtetni, hogy a beszédtervezés/fordulóalkotás során a tagmondatokat általában egy egységként kezelik a beszélők.

A dolgozat fő témájának, az *izé* lexikális kitöltőelem kvalitatív vizsgálatának megértéséhez elengedhetetlen a konverzációelemzés néhány fontos fogalmának tisztázása. A konverzációelemzés szerint az interakció alapegysége a forduló. Egy forduló addig tart, amíg az aktuális beszélő át nem adja a szót egy következő beszélőnek vagy szóátadás nélkül befejezi mondanivalóját (Sacks et al., 1974). Egyszerűbben megfogalmazva, a forduló az az egység, amit a beszélő egyszerre mond külső megszakítás nélkül. A forduló tovább bontható fordulókonstrukciós egységekre (továbbiakban TCU), amelyeknek tartalmazniuk kell legalább egy felismerhető cselekvést (Clayman, 2013). Amikor a fordulókonstrukciós egységgel a beszélő legalább egy cselekvést végrehajtott, relevánssá válik a beszélőváltás. Ezeket a pontokat váltásreleváns helyeknek nevezzük. A váltásreleváns hely a cselekvés lehetséges lezárási pontja, amelyen egy másik résztvevő szabályosan átveheti a szót az aktuális beszélőtől, de ez nem szükségszerűen történik meg (Sacks et al., 1974).

1.2. A kutatás célja és motivációja

Tanulmányomban a magyar nyelvű spontán, baráti társalgásokban megjelenő lexikális kitöltőelemet, az *izé*t vizsgálom kvalitatív és kvantitatív módszerrel

egy saját gyűjtésű, spontán, baráti társalgásokat tartalmazó korpuszon a funkcionális szemléletű diskurzuselemzés elméleti keretében, konverzációelemzéses módszerrel. Az *izé* a *Nyelvművelő kézikönyv* szerint egy „nyelvi igénytelenségre" utaló megbélyegzett nyelvi forma a nyelvhasználók megítélése szerint (vö. Grétsy & Kovalovszky, 1980). Egy frissebb kutatás azonban némiképp árnyalja ezt a vélekedést. Gyarmathy (2017) kérdőíves módszerrel többek között arra kereste a választ, hogy mennyire zavarja az *izé* az adatközlőket, amelyet egy ötfokú skála segítségével térképezett fel. Ennek alapján a megkérdezettek 34,1%-át kicsit zavarta, és mindössze 22,4% volt, akit nagyon zavart. Úgy tűnik, a nyelvhasználók megengedőbbek, toleránsabbak lettek az *izé* szó használatát illetően, azonban a kapott eredményekből az látszik, hogy csupán az adatközlők 11,8%-a az, akit egyáltalán nem zavar a kifejezés.

Habár az *izé* még mindig némileg stigma alá esik, a spontán, mindennapi társalgásokban számos oka és funkciója lehet a használatának, például szerepet játszhat a szókeresési folyamatokban, az előhívás nehézségeinek áthidalásában vagy a nagyobb szerkezeti egységek megformálásában (Fabulya, 2007). Kutatásom célja ezért az, hogy áthidalva a negatív előítéleteket, bemutassam az *izé* mint nyelvhasználati jelenség spontán, baráti társalgásokban betöltött, eddig feltárt szerepeit, azaz a szókereső, késleltető és tabukerülő funkcióját; és kísérletet tegyek eddig még nem kutatott funkciók feltárására, az *izé* függő beszédben betöltött szerepének, illetve a diskurzusszervező funkciójának bemutatására. Az *izé* nem csupán a mentális lexikonhoz való hozzáférés nehézségeinek jele, vagy a kifejezések elkerülésének egy formája, hanem számos társalgási szituációban releváns és funkcionális lehet (Fabulya, 2007).

Tanulmányomban arra a kérdésre keresem a választ, hogy:

(1) Milyen szekvenciális pozícióban jelenik meg az *izé*?

(2) Az eddig feltárt funkciók mellett milyen további szerepet tölthet be az *izé* a korpuszomban?

A jelenség vizsgálatának eddigi állása alapján hipotéziseim a következők:

A korpuszom előzetes megfigyelése alapján feltételezem, hogy:

(1) az *izé* többnyire váltásreleváns hely előtt jelenik meg, azonban megjelenhet még fordulókiterjesztés elején, forduló indításakor, párszekvencia második párrészének indításakor, illetve átfedésben is;

(2) az *izé* – pozíciója által – diskurzusszervező funkcióval bír a társalgásokban;

(3) az *izé* függő beszédben képes az angol *be + like* formulához hasonlóan idéző funkciót betölteni;

(4) amennyiben az *izé* pozíciója által diskurzusszervező funkcióval bír a társalgásokban, abban az esetben tekinthetjük diskurzusjelölőnek.

A szakirodalmi háttér és a kutatási célok bevezetése után a 2. fejezetben ismertetem a korpuszomat és a kutatás során használt módszereket. A 3. fejezetben rátérek az eredményekre, amely során bemutatom az *izé* típusait és eddig feltárt funkcióit a korpuszból vett társalgásrészletek elemzése által, majd a kategorizációs nehézségeket és az *izé*k előfordulásainak arányát a korpuszban. A 3.6. fejezetben ismertetem az *izé* idéző funkcióját, majd a 3.7. fejezetben bemutatom az *izé* strukturális pozícióhoz köthető diskurzusszervező funkciójának vizsgálati eredményeit. Végül a 4. fejezetben összefoglalom az eredményeket és levonom a következtetéseket, reflektálva a hipotéziseimre.

2. Korpusz és módszer

Kutatásom során az általam készített és feldolgozott hanganyagok, a *Kocsma-korpusz* alapján dolgoztam, amely baráti összejövetelekről készült beszélgetésekről, kocsmai és otthoni környezetben zajló hangfelvételeket tartalmaz. Jelen tanulmányhoz a teljes korpusz anyagát, összesen 7 óra 58 percnyi hanganyagot használtam fel. Ez 18 felvételt jelent, melyek átlagosan 26 percesek. A korpusz két-, három-, négy-, öt- és hatfős társalgásokat tartalmaz. A kutatásban 8 adatközlő vett részt, akik egynyelvű, köznyelvet beszélő, megyeszékhelyen és

fővárosban élő lakosok, 3 nő és 5 férfi, foglalkozásuk változatos. Az adatközlők mindegyike egy baráti társaság tagja. A hangrögzítés fázisában igyekeztem minimalizálni a megfigyelői paradoxon (Labov, 1979) adatközlőkre gyakorolt hatását és maximalizálni a korpusz spontán, természetes jellegét. A megfigyelői paradoxon minimalizálása érdekében olyan adatközlőkkel dolgoztam, akik jól ismerik egymást; a felvételeket nem stúdiókörnyezetben, hanem az általuk megszokott környezetben és élethelyzetben rögzítettem; illetve a társalgások nem irányítottak, hanem egytől egyig úgy haladtak előre, ahogy a beszélők alakították őket. Ezen tényezők csökkentették az adatközlők feszélyezettségét, és növelték a korpusz természetességének mértékét. A felvétel körülményeinek zavaró tényezői, például a háttérben szóló rádió, televízió és más beszűrődő zaj a jó minőségű hangrögzítő eszköznek köszönhetően nem, vagy nagyon kis mértékben befolyásolta a beszélgetések érthetőségét.

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

Az adatok rögzítése után transzkripciót készítettem a konverzációelemzés átírási konvencióinak megfelelően (lásd: Függelék) (Jefferson, 2004). Ezt követően az *izé*ket tartalmazó társalgásrészleteket elemeztem a konverzációelemzés szempontjai szerint. Végül kategorizáltam a korpuszban megjelent *izé*ket funkcióik szerint, kvantitatív vizsgálat alá vetettem őket, és értelmeztem a kapott eredményeket.

3. Eredmények

A 3.1–3.3. fejezetben bemutatom az *izé* három típusát, vagyis a szókereső, a szópótló és az időnyerő *izé*t néhány korpuszomból vett példa alapján. A 3.4. fejezetben a kategorizációs folyamat során felmerült nehézségeket ismertetem, majd a 3.5. fejezetben az *izé*k előfordulásainak kvantitatív eredményeit. A 3.6. fejezetben bemutatok egy új, eddig még nem tárgyalt funkciót, az *izé* idéző funkcióját. A 3.7. fejezetben áttekintést nyújtok az *izé* strukturális pozíciói által betöltött diskurzusszervező funkcióiról.

3.1. Szókereső izé

A következő társalgásrészletben Feri az *Index* nevű magyar hírportálon végbement változásokról, az újságírók tömeges felmondásáról beszél.

(1) KO_200922-24

1	Feri:	mármint, valószínűleg akik otthagyták a **izé, a:m (.) az indexet,**
2		valószínűleg ők így közösen fognak valamit csinálni amúgy.

Feri fordulójának a progresszivitása megszakad az *izé* egységnél, és szókeresést hajt végre. Ezt a nyújtottan produkált *a:m* nemlexikális kitöltőelem és a mikropauza, vagyis a 0,5 másodpercnél rövidebb szünet támasztja alá. Szókereső *izé* esetén általában a keresett kifejezés rögtön megjelenik, amint sikerül előhívni azt. Jelen példában a beszélő az *izé*vel és az utána következő kitöltőkkel és a szünet produkálásával javítást (szókeresést) kezdeményez, és ezáltal elegendő időt nyer ahhoz, hogy képes legyen előhívni a lap nevét, biztosítva ezzel az *izé* feloldását és helyreállítva az interszubjektivitást.

A második beszélgetésrészletben Feri a Youtube videók előtt megjelenő reklámokról panaszkodik partnereinek. Fordulójában több szókereső *izé*t meg lehet figyelni, vizsgáljuk meg ezeket egyesével!

(2) KO_200922-24

1	Feri:	nekem nem. nekem nagyon durva, nekem sose szokta felhozni.
2		én ilyen ilyen szarokat ho fel ilyen **izé about you**
3		meg **izé, öö flipet** hoz föl egyfo egyfolytába, ilyen politikai cuccot.
4		bá bár ma ma volt először hogy újra, (.) szembe jött velem
5		**izé, öö a:j hogy hívják a faszit,** a: (.) kovid arcát?
6	Lenke:	azt (.) győrfi pál.
7	Feri:	GYÖ GYŐRFI PÁL! ma volt az első győrfi pálom. (.)
8		öö hónapok óta először.

A forduló első *izé*je esetén Ferinek rögtön a produkált kitöltőelem után sikerül előhívnia a keresett szót: *izé about you* (online webshop). A második *izé*t egy nemlexikális *öö* kitöltő is kíséri, majd Feri ezután biztosítja az *izé* feloldását: *flipet* (internetszolgáltató). A harmadik *izé* ezeknél lényegesen komplexebb. Ebben az esetben az *izé*t ugyancsak egy nemlexikális *öö* hang követ, amelyhez járul egy, vélhetően negatív attitűdöt kifejező *a:j* indulatszó is (vö. ÉKsz., 1972). Ezután nyilvánvalóvá teszi, hogy a keresett szó, a pontos név nem elérhető számára: *hogy hívják a faszit,* majd ezt specifikálja megint csak elnyújtva a névelőt és mikropauzát produkálva: *a: (.) kovid arcát?* Ez a javítási folyamat (ön)kezdeményezése, amely egyértelműen egy külső segítségkérés, azaz egy küljavításkérés (Schegloff et al., 1977). Erre válaszképpen Lenke a következő fordulóban kis szünet után végrehajtja a (kül)javítást, azaz megadja a Feri által keresett személy nevét. Erre Feri helyesel, megismétli a férfi nevét nagyobb hangerővel, felkiáltó intonációval: *GYÖ GYŐRFI PÁL!*. Ez arra enged következtetni, hogy az *izé* feloldása, a javítási folyamat és ezáltal az interszubjektivitás helyreállítása sikeresen végbement.

A fenti példában érdemes megfigyelni a *szarok,* illetve a *politikai cuccot* kifejezéseket. Gyarmathy (2012) szerint az *izé*t a nyelvi stigmatizáltság miatt a felnőtt beszélők gyakran más elemmel, például a *cucc,* a *szar,* a *bizbasz,* a *tudodmi,* a *hogyishívják* vagy a *hogymondjam* alakokkal helyettesítik. A fenti példában a *szarokat* kifejezés azonban nem teljesen ugyanazt a szerepet tölti be, mint az *izé*, hanem Feri reklámokhoz fűződő negatív attitűdjét fejezi ki.

Feri itt a *szarokat* kifejezést vélhetően egyfajta gyűjtőnévként használja a reklámokra, míg a társalgásban egyértelmű jelei vannak annak, hogy a beszélő az *izé*t két tulajdonnév, az *About you* és a *Flip* helyettesítésére használja. A *szarokat* feloldásának hiánya miatt nem lehetünk biztosak abban, hogy a beszélő pontosan mit értett a kifejezés alatt, hiszen nem látunk bele a mentális lexikonjába. Amennyiben a beszélő, amellett, hogy negatív attitűdöt fejez ki a szóval, a reklámok helyettesítésére is használja, abban az esetben tekinthetjük a fenti példában a *szarokat* szóhelyettesítő, azon belül is szópótló formulának. A későbbiekben láthatunk majd példát a *szar* kifejezés *izé*vel azonos használatára is.

A *politikai cuccok* egység ugyancsak hasonló az *izé*hez, hiszen átmenetileg vagy véglegesen helyettesít egy lexémát. Az elem feloldásának elmaradása nem vezet félreértéshez, hiszen a kontextusból egyértelműen levezethető, hogy a beszélő valamely politikai tartalmú reklámokról beszél.

3.2. Szópótló izé

A következő példában Feri egy magyar műsorvezetőről beszél, akinek nem jut eszébe a neve.

(3) KO_201027-30

1	Feri:	én a **izére** emlékszek, a:z M1-en volt egy ilyen öö (.)
2		egy darabig asszem csak (.) ühm időjárás jelentő volt
3		aztán híradóba is bekerült. de mindig mosolygott,
4		de ilyen ilyen nagyon bárgyúan ilyen,
5		meg egy kicsit sorozatgyilkos feje volt, (.) de hogy a híradóba
6		hogyha valami halál sztoriról beszélt akkor is így mosolygott,
7		hogy így (.) és akkor meg[haltak.]
8	Gábor:	[ez para.]
9	Feri:	és ilyen nagy fülei voltak asszem.
10		(9.0)

Feri elkezd mesélni egy műsorvezetőről (1. sor), de nem emlékszik a nevére, így az *izé* szóval helyettesíti. Korábban láthattuk, hogy a szókereső esetén a

beszélő feloldotta az *izé* szót, amint sikerült előhívnia a megfelelő kifejezést. Itt viszont ismertetőjegyek és tulajdonságok halmazát adja meg: az M1 tévécsatornán volt műsora, korábban csak időjárásjelentést vezetett, majd a híradóba is bekerült, mindig mosolygott, bárgyúan, sorozatgyilkosra hasonlított. Mivel a beszélgetőpartnerek nem produkálnak válaszreakciót a váltásreleváns helyeken, így látszólag Feri nem próbálkozik tovább, hanem megfogalmazza a férfiról tett állítását: hogyha valamilyen halálesetről beszélt a híradóban, akkor is mosolygott. Gábor erre válaszreakciót produkál átfedésben: *ez para.* Feri további jellemzőt ad meg a referensről: *nagy fülei voltak,* tovább próbálkozik. Erre nem érkezik válaszreakció, hanem egy rendkívül hosszú, kilenc másodperces szünet után a beszélők újabb témába kezdenek.

Az alábbi társalgásban Feri a bolti lopásról beszél.

(4) KO_201027-31

1	Feri:	mer hogy a: **izék** a a ezek a szarok hogyha nincsen rajta
2		ez a: csippantós szar,
3	Lenke:	aha.
4	Feri:	hanem csak vonalkódos, akkor **izé** nem jelzi hogy kiviszed.

Feri a fenti megnyilatkozásában azt próbálja elmondani beszélgetőpartnereinek, hogy a bolti termékek esetén, amelyeken nincsen mágneses lopásgátló, nem jelez az áruvédelmi kapu. Feri láthatóan keresés, azon belül is szókeresési műveletet hajt végre. A névelő elnyújtásával, az *izé*vel, illetve az azt követő névelőismétléssel jelzi, hogy valamilyen probléma merült fel a szó előhívása során. Az esetek többségében az *izé* szó feloldása szükséges és elégséges feltétele annak, hogy az interszubjektivitás fennmaradjon vagy helyreálljon (vö. Fabulya, 2007). Itt az *izé* szó feloldása azonban nem történik meg, hanem a szarok kifejezéssel cseréli az adatközlő (vö. Gyarmathy & Neuberger, 2013; Kondacs, 2017), majd ugyanezt a kifejezést specifikálja egy tulajdonság megadásával, megint csak a névelő elnyújtásával: *ez a: csippantós szar.* Korábban, a (2)-es példában találkozhattunk már a *szar* kifejezéssel, ahol a szó negatív attitűdöt fejezett ki. Ezzel szemben itt vélhetően a *szar* kifejezés az *izé*hez hasonlóan egy bizonyos

lexéma póltásaként szolgál, hiszen jól látható, hogy Feri problémába ütközött a megfelelő kifejezés előhívása közben és keresést hajt végre. Habár a *szar* szó jelentéséből adódóan lehet egy megbélyegzett forma, a beszélgetőpartnerek közeli, baráti viszonya felülírja a klasszikus értelemben vett „szép", igényes beszédre való törekvést (vö. Pap, 2011).

Vannak olyan esetek, amikor a beszélők az *izé* kifejezést nem oldják fel, nem javítják, ugyanis a szövegkörnyezetből ki tudják következtetni beszélgetőpartnereik, hogy mit helyettesít.

Az alábbi részletben Jutka egy idegen nyelvű blogról beszél, amiben nem értett egy bejegyzést, lévén, hogy nem magyarul íródott.

(5) KO_200916-17

1	Jutka:	de: (0.5) azért, mert hirtelen ugye (.) azon *izé*ltem
2		hogy (.) hogy nem magyarul van. (0.5) na mindegy.

Jutka ebben az esetben az *izé*t igeként használja, melynek jelentése *gondolkodtam, tanakodtam* lehet. Itt nem szükséges a kifejezés feloldása, ugyanis a szövegkörnyezetből egyértelműen levezethető a szó jelentése. Ez feltehetőleg a nyelvi ökonómia egyik megnyilvánulási formája (Fabulya, 2007). Ebben az esetben a grice-i mennyiségmaxima sérülni látszik (Grice, 1975), hiszen a beszélő nem közöl elegendő információt. Azonban az interakció egy olyan kooperatív folyamat, amely során a beszélgetőpartnerek az erőfeszítés minimalizálására, gazdaságosságra törekednek (Coulmas, 1992). Ez azt jelenti, hogy a beszélők annyi információt közölnek, amennyiről feltételezik, hogy a másik számára feltétlenül szükséges az interakció folytatásához (Coulmas, 1992; Huszár, 2005). Mivel a fenti példában a szövegkörnyezetből tisztán elérhető a jelentés, így azt semmilyen módon nem szükséges pontosítani, konkretizálni, az *izé*t feloldani. Itt érhető tetten a nyelvi ökonómia, a gazdaságosság.

3.3. Időnyerő izé

Az *izé* másik típusa, az időnyerő *izé* akkor jelenik meg, amikor a beszélő még nem formálta meg a mondanivalóját, és még csak a tartalmi elemek válogatásánál tart (Gósy, 2004). Fontos különbség a szóhelyettesítő és az időnyerő

izé között, hogy míg a szóhelyettesítő illeszkedik a forduló szintaktikai szerkezetébe, addig az időnyerő *izé* nem, ugyanis nem egy konkrét lexémát helyettesít, funkciója pusztán az időnyerés a tervezési folyamat során.

A következő beszélgetésrészletben Jutka arról panaszkodik, hogy eldugult a lakásukban a lefolyó, és nem folyik le a víz.

(6) KO_201027-30

1	Jutka:	már második napja merjük ki a vizet a (.) öö mosogatóból,
2		hogy tegyünk bele (.) öh olyan (0.5) tisztítót.
3	Laci:	Óhh [óóhh]
4	Jutka:	[de még] mindig nem [ment le,] =
5	Lenke:	[cch basszus!]
6	Jutka:	= és öö most **izé,** öö (.) most láttam, hogy ki van szedve
7		a a az alatta lévő szekrényből minden sz:ar, alatta egy felmosó,
8		(.) gyorsan eljö(h)ttem ((nevetés)) ottho(h)nró(h)l.

Jutka egy történetmesélésbe kezd a lakásukban jelentkező problémáról: a lefolyó eldugulásáról. Laci és Lenke erre együttérzően reagálnak Jutkával átfedésben. Laci *óóhh* egysége (3. sor) váltásreleváns helyen történik, vagyis Jutka már végrehajtott a TCU-val egy felismerhető cselekvést, és így relevánssá vált a beszélőváltás (Sacks et al., 1974). Laci itt vélhetően nem akarja átvenni a szót, csak válaszreakciót, visszacsatolást produkál Jutka történetéhez.

Hasonló történik a 4–5. sorban lévő átfedésben. Mindkét esetben, Laci *óóhh* és Lenke *cch basszus!* egysége is egyértelműen egy visszacsatolás, amellyel nagy valószínűséggel egyik beszélő sem akarja átvenni a szót Jutkától. A különbség a szekvenciális elhelyezkedésben rejlik. Míg Laci és Jutka egyszerre beszélése váltásreleváns helyen történik, addig Lenke nem várja meg ezt a pontot, nem várja meg Jutka cselekvésének lehetséges lezárási pontját, hanem hamarabb produkál válaszreakciót. Miért lehet ez?

Mivel egy társalgás során a beszélők folyamatosan monitorozzák egymást (Schegloff, 2007), így a partner gyakran már a cselekvés lehetséges lezárási pontja, azaz a váltásreleváns hely előtt ki tudja következtetni a mondanivaló végét.

Ez a korai kikövetkeztetés sok esetben felismerő átfedést, azaz fordulóbelseji szóátvételt eredményezhet (Hayashi, 2013). Ez történik a fenti példában is Lenke és Jutka átfedése során. A felismerő átfedés általában nem igényel javítást, hiszen a társalgás természetes velejárója (Hayashi, 2013). Itt azonban még inkább nem szükséges javítani, hiszen a *cch basszus!* egység nem egy valódi szóátvétel, csupán egy visszacsatolás, amely nem fenyegeti Jutka beszédhez való jogát, vagyis, hogy végre tudja hajtani a kívánt fordulóját.

Láthatjuk, hogy Jutka átfedést követő TCU-ja nemlexikális kitöltőket (*öö*), *izé*t és mikropauzát tartalmaz. Ezek megjelenésének oka feltehetőleg az, hogy Jutka ezen a ponton még nem formálta meg a mondanivalóját, ezért időnyerésre van szüksége. Itt az *izé* tehát nem egy konkrét lexémát helyettesít átmenetileg vagy véglegesen (vö. Fabulya, 2007), hanem az azt követő nagyobb szerkezet megformálásának produkálását késlelteti. Ennek alapján tekintem a fenti példában szereplő elemet időnyerő *izé*nek.

A következő példában Laci arról beszél, hogy szívesebben vásárol termelőtől tojást, mintsem a nagyobb áruházakban.

(7) KO_200916-17

1	Laci:	akkor gondoltam én, nem tom, szívesebben eszek
2		**izé,** olyan helyről, ahol tudom hogy **izé,**
3		(1.0)
4	Lenke:	kapirgálós csirkék.
5	Laci:	ja.
6	Jutka:	hehehe
7	Laci:	szóval hogy, (.) örülök hogy nem kell bolti tojást.

Láthatjuk, hogy a fenti példa két időnyerő *izé*t tartalmaz. Ezek egyike sem illeszkedik a forduló szintaktikai szerkezetéhez, jelenlétük csupán időnyerésre szolgál. Az időnyerő *izé*k gyakran a mellékmondatok elején állnak, a mellékmondatot bevezető kötőszó közelében vagy szomszédságában (Fabulya, 2007). Ez jól látszik a második előfordulásban, amikor Laci a *hogy* alárendelő kötő-

szó után produkálja az *izé*t. Ez hasonlóságot mutat a *hogy öö* formulával (vö. Németh, 2020).

3.4. Az izé kategorizációs nehézségei

Az elemzés során megfigyeltem olyan kategorizációs nehézségeket, amelyek esetén nem mindig volt egyértelmű, hogy az adott *izé* melyik típusba sorolható. Ezeket pontokba szedve, egy-egy korpuszomból vett példával alátámasztva mutatom be.

A) A szópótló *izé* véglegesen helyettesít olyan elemet, amely valamilyen okból nem elérhető a mentális lexikonból. Mind a szópótló, mind a szókereső esetén előfordulhat, hogy a kontextusból való kikövetkeztethetőség miatt az *izé* feloldása elmarad (Fabulya, 2007). Tehát vannak olyan esetek, amelyekben a végleges helyettesítés és a feloldás hiánya egybeesik, így nehezen meghatározható, hogy szópótló vagy feloldás nélküli szókereső kategóriába tartozik.

(8) KO_200922-22

1 Feri: picsába. (1.0.) nem lehetek az exed (0.5) aktuális **izé,** öhöh

2 Etus: nem.

A fenti példában azért különösen nehéz a kategória meghatározása, mert az *izé* feltehetőleg egy tabukifejezést helyettesít, amely gyakran előfordul szópótló *izé*k esetén (Fabulya, 2007).

B) A szóhelyettesítő *izé* ismérve, hogy illeszkedik a szintaktikai szerkezetbe és felveszi az általa helyettesített elem toldalékát (Fabulya, 2007). Ez a toldalékfelvétel azonban bizonyos esetekben elmaradhat, és olyankor egybeesik az időnyerő kategóriával.

(9) KO_200922-24

1 Feri: hát valószínűleg teszkós **izé,** szatyorral fogom

2 befátyolni az egész karoma(h)t.

C) Az időnyerő *izé* nagyon gyakran mellékmondatot bevezető kötőszó utáni pozícióban van (Fabulya, 2007). Előfordul azonban, hogy a szóhelyettesítő *izé*

ugyanebben a pozícióban helyezkedik el, amely ugyancsak kategorizációs nehézségeket eredményez.

(10) KO_201027-31

1	Gábor:	meg volt olyan hogy **izé,** hogy ilyen patkánymérges ilyen
2		valami törm- növényi törmelék patkányméreggel felütve,

Ezeket figyelembe véve megállapítható, hogy az *izé* egyes típusait nem minden esetben egyszerű elkülöníteni egymástól. A kategorizációs folyamat során ezért érdemes a fenti tényezőket figyelembe venni és az elemzések során is szem előtt tartani.

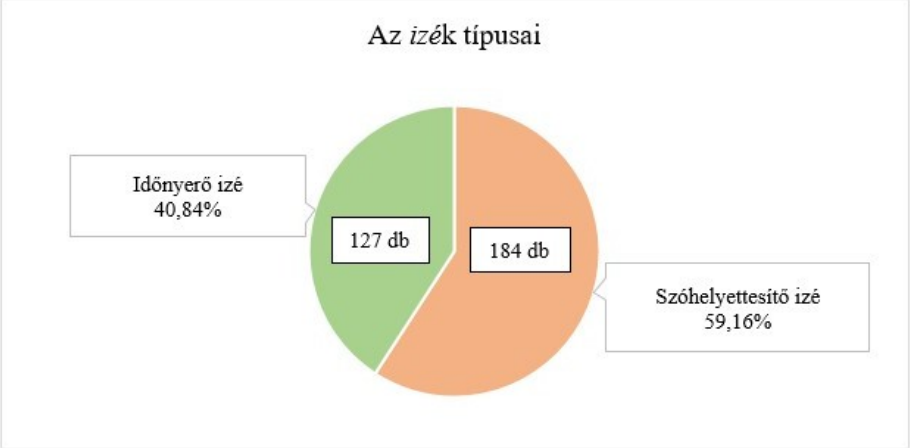
3.5. Az izé típusainak kvantitatív eredményei

Az *izé* lexikális kitöltőelemet Gyarmathy (2012) vizsgálta kvantitatív módon a BEA spontánbeszéd-adatbázis (Gósy et al., 2012) 75 óra 5 percnyi hanganyagán. A 94 adatközlőből 59 egyáltalán nem használt *izé*t, míg a többi adatközlő átlagosan 4 db *izé*t produkált a társalgásokban. Ez összesen 140 db előfordulást jelent az egész korpuszban. Ezek nagyobb része, összesen 61,7%-a szóhelyettesítő, míg a maradék 38,3% időnyerő *izé* volt. Gyarmathy (2012) és Kondacs (2017) Fabulyához (2007) hasonlóan a töltelék *izé* terminust használták. Kondacs (2017) óvodáskorúak megnyilatkozásainak vizsgálata során arra jutott, hogy a különböző *izé*k előfordulásának aránytalansága ebben a korosztályban jóval nagyobb, 95%-ban jelent meg a szóhelyettesítő, 5%-ban pedig az időnyerő *izé*.

Korpuszom 7 óra 58 percnyi anyagában az *izé* összesen 311 db előfordulást mutatott. Az előző alfejezetben ismertetett kategorizációs nehézségek ellenére az összes *izé*t osztályoztam, beleértve a nehezebben besorolható példákat is.

Az 1. ábrán az *izé* típusainak arányát mutatom be. Az ábrán jól látható, hogy a teljes korpuszban összesen 184 db, azaz 59,16% szóhelyettesítő, és 127 db, azaz 40,84% időnyerő *izé* jelent meg. Ez arányaiban megfelel Gyarmathy (2012) eredményeinek. A szóhelyettesítőn belül 143 db (77,72%) szókereső és 41 db (22,28%) szópótló *izé* volt. Minden egyes társalgás tartalmazott legalább

egy *izét*, és minden egyes adatközlő produkált legalább egyet, az adatközlők között mennyiségben azonban lényeges eltérések voltak. Ez valószínűleg annak köszönhető, hogy minden beszélőnek más és más késleltetési és hezitációs stratégiája van. Ilyen lehet például a hangnyújtás (Gósy, 2004), a diskurzusjelölők pl. *hát* (Schirm, 2011), *ilyen, így* (Vukov Raffai, 2016), a nemlexikális *öö* hang (Németh, 2020, 2021) vagy kitöltetlen szünet (Gósy, 2000). Előzetes megfigyelésem alapján azoknál a beszélőknél, akik elenyésző számú *izét* produkáltak, nagymértékben megjelent az *öö*, az *ilyen* és az *így* diskurzusjelölő, illetve a kitöltetlen szünetek. Ennek megfelelő alátámasztása további kvalitatív és kvantitatív vizsgálatokat igényel.



1. ábra. Az *izék* típusai

3.6. Az izé a függő beszédben

Az angol nyelvben megfigyelhető, hogy a *like* diskurzusjelölő gyakran idéző funkciót tölt be a függő beszédben (Romaine & Lange, 1991; Fox & Robles, 2010). Ilyenkor a beszélők a *be + like* formulát használhatják, amelynek funkciója megegyezik a *say* igével. Lássunk erre két példát:

(11) Romaine & Lange, 1991: 227

A man came up to me and said, "You really look like Princess Di." And

he looked at me and ***he's LIKE,*** *"Are you?"*

'Egy férfi odajött hozzám, és azt mondta: „Tényleg úgy nézel ki, mint Di hercegnő." Rám nézett, és azt mondta: „Te vagy az?"'

(12) Romaine & Lange, 1991: 227

And I saw her coming, and ***I'm LIKE,*** *"Nooooooooo."*

'És láttam őt jönni, és azt mondtam, "neee".'

Láthatjuk, hogy a fenti példákban a *like* ugyanúgy működik, mint a *say* ige a függő beszéd bevezetésénél. Mindkét példában a *be* ige egy formáját a *like* követi, amely felcserélhetőnek tűnik a *say* igével. A fenti példákban a *be + like* tehát idéző funkciót tölt be (Romaine & Lange, 1991). Ez az idéző funkció azonban nem csak a mások vagy saját magunk által mondottak idézését jelenti, hanem gyakran belső gondolatokat is reprezentálhatunk általa. A (11)-es példában feltehetőleg a beszélő a felidézett szituációban nem mondta ki a *Nooooooooo* egységet, csupán azt fejezi ki beszélgetőpartnerének, hogyan érezte magát, amikor látta a lányt jönni (Butters, 1982).

A korpuszom elemzése alapján az *izé* néhány esetben képes ehhez hasonló funkciót betölteni. Az alábbi társalgásrészletben Feri egy humoros történetet mesél el egy ismerőséről, aki magyar irodalomból diplomázott.

(13) KO_200922-22

1	Feri:	emlékszem amikor izé, asszem az alapszakot megcsinálta
2		és (.) hát az se most volt már,
3	Lenke:	ühüm.
4	Feri:	azt izé(h) kint voltunk a laposon izé tábortüzeztünk
5		meg így iszogattunk azt **izé,** ez az megvan az irodalom izé, diplomám
6		és akkor mostantól diplomás rőzsegyűjtő [vagyo(h)k.]
7	Lenke:	[((nevetés))]

A fenti társalgásrészlet összesen négy *izé*t tartalmaz. Az első sorban lévő egy tipikus időnyerő *izé*, kötőszó után áll. A negyedik sorban nevetve elhangzó *izé* szintén egy klasszikus időnyerő funkciójú, pozíciója szintén jellemző a kategóriára. Itt fontos megjegyezni, hogy az *azt izé(h)* jelentése 'és izé/aztán izé'

(vö. ÉKsz., 1972), ez a beszélő sajátos nyelvhasználatának a része. A szintén a negyedik sorban lévő ugyancsak egy időnyerő *izé*, amely ismét egy kötőszó után áll.

Jelen elemzés szempontjából az 5–6. sor az érdekes számunkra. Ebben az egységben Feri a szóban forgó ismerős szavait idézi, így tehát a példa függő beszédet tartalmaz. Ezt nem fejezi ki explicit módon, csupán az *azt izé* kifejezés használatával vezeti be. Ha összehasonlítjuk ezt a példát a fent tárgyalt *be + like* formulával, láthatjuk, hogy ebben az esetben is felcserélhető az *izé* a *mond* igével: *meg így iszogattunk és azt mondta, hogy ez az megvan az irodalom izé, diplomám és akkor mostantól diplomás rőzsegyűjtő vagyok.* Habár a beszélő az *azt izé* formulát többször is használja, itt sajátos, idézésre jellemző intonációval produkálja a fenti egységet. Ennek alapján feltételezem, hogy az *izé*nek lehet egy új, eddig még fel nem tárt funkciója, az idéző funkció, ez azonban további vizsgálatokat igényel.

3.7. Az izé strukturális pozícióhoz köthető diskurzusszervező funkciói

Németh (2021) tanulmányában az *öö* nemlexikális kitöltőt hezitációs jelölőnek és diskurzusjelölőnek is értelmezi annak alapján, hogy felfüggeszti-e a folyamatban lévő cselekvést, illetve részt vesz-e, és ha igen, hogyan vesz részt a kitöltő a beszélőváltás szabályozásában. Németh (2021) szerint, ha egy kitöltő hezitációs funkcióval rendelkezik, kétségkívül hezitációs jelölőnek tekinthető, azonban ez nem zárja ki azt, hogy diskurzusjelölőnek is tekinthessük. A kitöltőelemek diskurzusszervező funkciója – amely alapján tekinthetjük őket diskurzusjelölőnek – nem a cselekvés felfüggesztésén múlik, hanem a szekvenciális elhelyezkedésükön, a strukturális pozíciójukon, hiszen ennek révén vesznek részt a társalgás szabályozásában (Németh, 2021). Németh (2021) öt kategóriát különített el a szerkezeti pozíciók alapján, amely pozíciókban a kitöltőelem diskurzusszervező, azaz diskurzusjelölő funkcióval rendelkezhet. Tanulmányomban ezt az öt kategóriát használom az *izé* lexikális kitöltőelem diskurzusszervező funkciójának tesztelésére. Ez az öt kategória a következő: (1) váltásreleváns hely előtti pozíció, (2) fordulókiterjesztés indítása, (3) olyan forduló indítása, amely nem egy

párszekvencia második párrésze, (4) párszekvencia második párrészének indítása, (5) átfedésekben, azaz egyszerre beszélések során megjelenő *izé*k. Az utóbbi négy kategória mind váltásreleváns helyet jelent. Jelen tanulmányban az első három kategória releváns, ugyanis leginkább azok jelentek meg a korpuszomban.

3.7.1. Az izé előfordulása váltásreleváns hely előtt

A beszélőváltási rendszer normatív szabályai szerint amikor egy résztvevő megszólal, nem csak joga, de kötelessége legalább egy felismerhető cselekvést, azaz egy fordulókonstrukciós egységet produkálni (Schegloff, 2007). Ez azt jelenti, hogy a folyamatban lévő fordulót potenciálisan lezárhatóvá kell tenni, hiszen a beszélgetőpartnerek csak a cselekvés potenciális lezárási pontján, azaz a váltásreleváns helyen tudják szabályosan átvenni a szót (Clayman, 2013). Ha a beszélő felfüggeszti a folyamatban lévő cselekvést, az javításra és számadásra szorul a beszélgetőpartnerek felé (Németh, 2021). A beszélő gyakran az *izé* használatával biztosítja partnereit, hogy eleget fog tenni a beszélőváltás normatív szabályai szerinti kötelezettségnek és szándékozik potenciálisan lezárhatóvá tenni a cselekevést az átmeneti felfüggesztés ellenére. Nézzünk erre egy példát!

A következő társalgásrészletben Laci a főzésről beszél.

(14) KO:200916-17

1	Laci:	mer, (.) én is kifogytam az olyan kajákból,
2		amiket **izé** (.) szoktam csinálni, úgyhogy gondoltam,
3		hogy csinálok krumplipüRÉT =
4	Jutka:	[hmm]
5	Laci:	= [meg] sütök tofut, (0.5) mármint hogy csak így simán
6		(0.5) és mindkettő olyan rossz lett ((nevetés))

Laci a tagmondatot bevezető kötőszó (*amiket*) után *izé*t produkál. Ezen a ponton a beszélő még nem hajtott végre egy felismerhető cselekvést, még tehát nem következett be a váltásreleváns hely. Egy megkezdett fordulókonstrukciós egységet lezárhatóvá kell tenni a beszélőváltás normatív szabályai szerint (Sacks et al., 1974). Azonban a beszélő vélhetően valamilyen problémába ütközik a

szerkezeti egység megformálása során, amelyet az *izé* szó használatával tud áthidalni. Az *izé* funkciója a fenti példában egyfelől az, hogy jelzi, hogy több gondolkodási időre van szüksége (hezitáló funkció), amelyhez társul egy mikropauza, másfelől pedig értelmezhetjük egyfajta ígéretnek is a beszélgetőpartnerek felé, hogy eleget fog tenni a kötelességének, miszerint legalább egy felismerhető cselekvést fog produkálni, csupán több időre van szüksége (vö. Németh, 2021). Ez alapján az *izé* tekinthető diskurzusszervező funkcióval rendelkező kitöltőelemnek, amely a beszélő beszélőváltási rendszerhez való viszonyát fejezi ki: a beszélőváltási rendszerhez való igazodás szándékát, illetve a beszédhez való jog fenntartását jelzi. Ezt a szerepet a diskurzusjelölők interakciós funkciójaként értelmezhetjük, hiszen az ilyen típusú elemek nem tartalmi, hanem társalgásszervező célt, a beszélő és a hallgató közötti interakció koordinálását szolgálják (Clark, 1994). Mindezt figyelembe véve tekintem a fenti példában lévő *izé*t diskurzusjelölőnek.

Az *izé*k túlnyomó többsége strukturális pozíció szempontjából ebben a kategóriában jelent meg. Az összesen 311 db *izé*ből 296 db, azaz 95,18% a fordulókonstrukciós egység belsejében, váltásreleváns hely előtt fordult elő. A 296-ból 180 db szóhelyettesítő (60,81%) és 116 db (39,19%) időnyerő *izé* volt. A szóhelyettesítőn belül 139 db szókereső (77,22%) és 41 db (22,78%) szópótló fordult elő.

3.7.2. Az izé előfordulása fordulókiterjesztés indításakor

Amikor az aktuális beszélő végrehajt egy cselekvést, és nem jelöl ki következő beszélőt, illetve másik résztvevő sem veszi át a szót a váltásreleváns helyen, akkor az aktuális beszélőnek joga (de nem kötelessége) kiterjeszteni a fordulót (Sacks et al., 1974; Clayman, 2013). A következő társalgásrészletben a fordulókiterjesztés egy *izé*vel indul.

Az alábbi példában a beszélgetőpartnerek a szarvasmarhákról beszélgetnek.

(15) KO_201027-28

1	Jutka:	pedig én amúgy nem annyira (.) vagyok oda a bocikért,
2		[de ez nagyon]=
3	Lenke:	[én nagyon.]
4	Jutka:	= aranyos.
5		(11.3)
6	Lenke:	az egyik kedvenc állatom.
7	Etus:	nekem is.
8		(1.5)
9	Feri:	amúgy jófejek. **izé,** amikor izé, túrázni szoktam menni
10		akkor szoktam találkozni ilyenekkel így kerítésen keresztül,
11		és jönnek oda haverkodni.

A társalgásrészlet kilencedik sorában Feri kifejti a szarvasmarhákhoz fűződő szimpátiáját: *amúgy jófejek*, végrehajtva ezzel egy fordulókonstrukciós egységet, így váltásreleváns hely következik. Ezen a ponton nem jelöl ki következő beszélőt és másik résztvevő sem veszi át a szót önkijelöléssel, így Feri kiterjeszti a fordulóját egy személyes történettel. A történetet az *izé*vel indítja, amellyel jelzi a fordulókiterjesztési szándékát (Németh, 2021). Ez biztosítja, hogy a váltásreleváns helyen ne lépjen be másik beszélő a fordulókiterjesztés előtt. Mivel az *izé* ebben az esetben hatással van a társalgás szekvenciális felépítésére azáltal, hogy befolyásolja a résztvevők közötti fordulók megoszlását, ennek alapján megállapítható, hogy a fenti *izé* rendelkezik diskurzusjelölő funkcióval.

Korpuszomban az *izé* fordulókiterjesztés indításán összesen 12-szer jelent meg, melyből 4 db szóhelyettesítő és 8 db időnyerő volt.

3.8. Az izé előfordulása forduló indításakor

Az alábbi társalgásrészletben a beszélgetőpartnerek a prémium macskaeledelekről beszélnek.

(16) KO_201027-30

1	Gábor:	bazmeg jobbat esznek, mint én.
2	Lenke:	a[múgy ja.]
3	Jutka:	[háháhá.] ja:j, igen.
4	Lenke:	**izé,** folyton májast eszek.
5	Gábor:	[((nevetés))]
6	Jutka:	[((nevetés))]

A fenti beszélgetésben a résztvevők a tévében néznek egy prémium macskaeledelről szóló reklámot, mire Gábor megjegyzi, hogy a macskák jobb ételeket esznek, mint ő (1. sor), amelyre két másik résztvevő, Lenke és Jutka is egyetértően helyesel (2–3. sor). A negyedik sorban Lenke kijelöli magát a beszédre és az *izé* szóval kezdi a fordulóját. Ezzel egyrészt elhalasztja a mondanivaló első tartalmas elemét. Ebben a szakaszban feltehetőleg még gondolkodási időre van szüksége a forduló megformálásához, másrészt pedig biztosítja a szóhoz való jogát (diskurzusszervező funkció). Lenke az *izé* használatával jelzi a beszélgetőpartnerei felé, hogy igényt tart a fordulóra. Önkijelöléssel veszi át a szót. A beszélőváltás normatív szabályai szerint önkijelölés esetén mindig az első megszólaló szerzi meg a beszédhez való jogot, így potenciálisan megakadályozhatja, hogy egy másik résztvevő is önkijelölést alkalmazzon (vö. Németh, 2021).

Ebben a strukturális pozíció szerinti kategóriában az *izé* mindössze háromszor fordult elő, melyből 1 db szóhelyettesítő és 2 db időnyerő volt.

3.8.1. Az izé előfordulása párszekvencia második párrészének indításakor

A párszekvencia a szekvenciák azon típusa, amely két fordulóból áll, és az első forduló, azaz az első párrész produkálása után elvárt egy második párrész produkálása (Schegloff & Sacks, 1973).

(17) KO_201027-30

1	Jutka:	figyi már, Feri azt akartam kérdezni hogy
2		te hány nap után mentél egy röntgenre a (.) esés után.
3	Feri:	hát **izé,** öö másnap.

A fenti társalgásrészletben Jutka kijelöli Ferit mint következő beszélőt. Ilyenkor az aktuális beszélőnek, azaz Jutkának kötelessége befejeznie a fordulóját, a kijelölt beszélőnek, azaz Ferinek pedig nemcsak joga, de kötelessége adekvát következő válaszreakciót produkálnia (Sacks et al., 1974). Korpuszomban nem áll rendelkezésre olyan előfordulás, amikor a párszekvencia második párrészének indításánál jelenik meg az *izé*, csupán olyan, amelyet *hát* diskurzusjelölő előz meg és egy *öö* egység követ. Habár ebben a példában nem az *izé*, hanem a *hát* diskurzusjelölő az első elem, amellyel a beszélő „ígéretet tesz" a beszélőváltási szabályok szerinti kötelességének teljesítésére, feltételezem, hogy itt a *hát izé öö* egy egységet alkot (vö. Dér, 2016, 2017; Schirm, 2018). Németh (2020) mintázatot vélt felfedezni a *hát + öö* előfordulásában. Amikor a *hát* – a számos funkciója közül – válaszjelölő szerepet tölt be a társalgásban, azt gyakran követi *öö* (Németh, 2020). Ilyenkor a *hát*hoz kapcsolódó *öö* egyfajta ígéret lehet a forduló folytatására, azaz a beszélőváltási rendszer szerinti kötelezettség teljesítésére (Németh, 2020, 2021). A fenti példában a *hát izé* ugyancsak egy párszekvencia második párrészének indításán fordul elő, ezek együttesével jelzi a beszélő a válaszadási szándékát, így vélhetően ebben az esetben – a *hát öö*-höz hasonlóan – egy egységet alkotnak. Ezt alátámasztja az is, hogy a *hát izé* egyetlen intonációs egységként hangzik el.

A magyar diskurzusjelölő-társulásokkal foglalkozó szakirodalom (pl. Dér, 2016, 2017; Schirm, 2018) nem tér ki az *izé*re. Ez nem meglepő, hiszen egyrészt nem teljesen tisztázottak az *izé* státuszát illető kérdések, nincs egyetértés a különböző elméleti keretek *izé*-értelmezésében (lásd: 1.1). Másrészt pedig, ha az *izé*t diskurzusjelölőnek is értelmezik, akkor sem tartozik a gyakoriak közé (Gyarmathy, 2015), így a diskurzusjelölő-társulásokban sem vizsgálták az előfordulását.

Kutatásom során feltételeztem, hogy az *izé* részt vesz diskurzusjelölő-társulásokban, különösen a *hát,* az *és* és az *így* diskurzusjelölőkkel. Az eredményeim azonban ezt nem igazolták, a *hát izé* egyszer, az *így izé* kétszer fordult elő, tehát nem mutattak mintázatot, inkább egyedi előfordulások voltak. Az *és izé* ennél valamivel többször, összesen ötször jelent meg a korpuszomban. Külön érdekesség,

hogy háromszor előfordult az *azt izé* 'oszt izé', 'aztán izé' jelentésben, igaz, ugyanattól a beszélőtől.

Korábban már tárgyaltuk, hogy az *izé* egyik típusa, az időnyerő *izé* nagyon gyakran tagmondatot bevezető kötőszó után áll (Fabulya, 2007). Korpuszomban összesen 85-ször fordult elő az *izé* kötőszó után, melyből a *hogy izé* a többihez képest kiemelkedően magas, 22 db előfordulást mutatott. Továbbá összesen 15 db előfordulást mutatott az *izé* vonatkozó névmással, melyből a leggyakoribbak az *amikor izé* és az *amiket izé* voltak.

Ebben a fejezetben láthattuk, hogy az *izé* a strukturális pozíciója révén képes az *öö*-höz hasonlóan diskurzusszervező funkciót betölteni a társalgásban. Ennek alapján beigazolódott a hipotézisem, miszerint az *izé* amellett, hogy lexikális kitöltőelem, diskurzusjelölőnek is tekinthető azáltal, hogy befolyásolja a társalgás szekvenciális rendeződését és a beszélgetőpartnerek fordulóinak megoszlását a társalgásban.

4. Konklúzió

Tanulmányomban megvizsgáltam a magyar nyelvű spontán, természetes társalgásokat tartalmazó korpuszomban az *izé* kettős arcát, vagyis a szóhelyettesítő és az időnyerő funkciót betöltő *izé*ket, melyeket néhány saját példa alapján ismertettem. Az *izé* típusait a korpuszban számszerűsítettem, amely során arra az eredményre jutottam, hogy a 311 db előfordulásból 184 db, azaz 59,16% szóhelyettesítő, és 127 db, azaz 40,84% időnyerő *izé* jelent meg, a szóhelyettesítőn belül pedig 143 db (77,72%) szókereső és 41 db (22,28%) szópótló *izé* volt. Ez arányaiban megfelel a korábbi eredményeknek (vö. Gyarmathy, 2015). A kategorizálás során felmerültek bizonyos kategorizációs nehézségek, melyek a következők voltak: a végleges helyettesítés gyakran egybeesik a feloldás hiányával; a szóhelyettesítő *izé* nem mindig vesz fel toldalékot; a szóhelyettesítő *izé* gyakran az időnyerő *izé* tipikus pozíciójában van.

Hipotézisem szerint az *izé* diskurzusjelölő, diskurzusszervező funkcióval is rendelkezik a szekvenciális pozíciója révén. Az *izé*ket öt kategóriába soroltam

a szerkezeti pozíciók alapján, amely előfordulások esetén a kitöltőelem diskurzusszervező, azaz diskurzusjelölő funkcióval rendelkezhet (vö. Németh, 2021). Ez az öt kategória a következő volt: (1) váltásreleváns hely előtti pozíció, (2) fordulókiterjesztés indítása, (3) olyan forduló indítása, amely nem egy párszekvencia második párrésze, (4) párszekvencia második párrészének indítása, (5) átfedések, azaz egyszerre beszélések. Kutatásom során feltételeztem, hogy az *izé* mind az öt kategóriában meg fog jelenni. A beszélőváltási rendszer normatív szabályai szerint amikor egy résztvevő megszólal, nemcsak joga, de kötelessége legalább egy felismerhető cselekvést produkálni, vagyis a folyamatban lévő fordulót potenciálisan lezárhatóvá kell tenni, hiszen a beszélgetőpartnerek csak a váltásreleváns helyen tudják szabályosan átvenni a szót (Clayman, 2013). Ha a beszélő felfüggeszti a folyamatban lévő cselekvést, az javításra és számadásra szorul (Németh, 2021). A beszélő gyakran az *izé* használatával biztosítja azt, hogy eleget fog tenni a beszélőváltás normatív szabályai szerinti kötelezettségnek, és szándékozik potenciálisan lezárhatóvá tenni a cselekevést az átmeneti felfüggesztés ellenére. Az *izé* a fordulókonstrukciós egység belsejében, váltásreleváns hely előtti pozícióban így vesz részt a társalgás szervezésében, azaz itt érhető tetten a diskurzusjelölő funkciója. Korpuszomban az *izé*k túlnyomó többsége, a 311 db előfordulásból 296 db ebben a kategóriában jelent meg, a típusainak megoszlása 60,81%–39,19% volt a szóhelyettesítő *izé*k javára az időnyerővel szemben.

A második kategóriába a már váltásreleváns hely után, a fordulókiterjesztés indításán megjelenő *izé*k tartoznak. Amikor az aktuális beszélő végrehajt egy cselekvést, és nem jelöl ki következő beszélőt, illetve másik résztvevő sem veszi át a szót a váltásreleváns helyen, akkor az aktuális beszélőnek joga (de nem kötelessége) kiterjeszteni a fordulót (Clayman, 2013). Ha az *izé* ebben a pozícióban van, azzal a beszélő egyrészt jelzi a fordulókiterjesztési szándékát, másrészt biztosítja, hogy a váltásreleváns helyen ne lépjen be másik beszélő. Az *izé* ebben az esetben hatással van a társalgás szekvenciális felépítésére azáltal, hogy befolyásolja a résztvevők közötti fordulók megoszlását, így diskurzusjelölő

funkcióval rendelkezik. Korpuszomban ez 12 db előfordulást mutatott, melyből 4 db szóhelyettesítő és 8 db időnyerő *izé* volt.

A harmadik kategóriába tartozó *izé*k a forduló indításán találhatóak. Ilyenkor nincs kijelölve beszélő, így bármely résztvevő önkijelöléssel átveheti a szót (Clayman, 2013). Ha a beszélő *izé*vel indítja a fordulóját, azzal jelzi, hogy igényt tart a szó jogára, de elhalasztja az első tartalmas elemet, feltehetőleg több gondolkodási időre van szüksége. A beszélőváltás normatív szabályai szerint önkijelölés esetén mindig az első megszólaló szerzi meg a beszédhez való jogot, így potenciálisan megakadályozhatja, hogy egy másik résztvevő is önkijelölést alkalmazzon (vö. Németh, 2021). Így vesz részt az *izé* a diskurzus szervezésében ebben a kategóriában. Az *izé* háromszor jelent meg ebben a pozícióban.

A negyedik kategóriában az *izé* a párszekvencia második párrészének indításakor jelenik meg. Párszekvencia esetén az első párrész elhangzása után elvárt a második párrész produkálása, ebben az esetben tehát mindig van kijelölt beszélő, akinek kötelessége adekvát válaszreakciót produkálnia (Sacks et al., 1974). Korpuszomban nem áll rendelkezésre olyan előfordulás, amikor az *izé* pontosan ebben a strukturális pozícióban jelenik meg, csupán olyan, amelyet a *hát* diskurzusjelölő előz meg. A *hát* mint válaszjelölő diskurzusjelölő a példámban feltehetőleg egy egységet alkot az *izé*vel, ugyanis a kettő együttesével jelzi a beszélő, hogy eleget kíván tenni a kötelezettségének, csupán gondolkodási időre van szüksége. Ezt erősíti az is, hogy a *hát izé* egyetlen intonációs egységben hangzik el.

A harmadik hipotézisem tehát, miszerint az *izé* mind az öt strukturális pozícióban meg fog jelenni, nem teljesült, hiszen a negyedik kategóriában megelőzte egy *hát* diskurzusjelölő, az ötödik kategóriában, átfedésben pedig egyáltalán nem jelent meg.

Kutatásom során feltételeztem, hogy az *izé* részt vesz diskurzusjelölő-társulásokban, különösen a *hát,* az *és* és az *így* diskurzusjelölőkkel. Az eredményeim azonban ezt nem igazolták, a *hát izé* egyszer, az *így izé* kétszer fordult elő, tehát nem mutattak mintázatot, inkább egyedi előfordulások voltak. Szintén egyedi előfordulásként szerepelt, de elemzésem megmutatta, hogy az *izé*, az eddig feltárt

funkciói mellett szerepet játszhat a függő beszédben. Ezt angol nyelvű kutatásokra alapoztam, amelyekben a *be + like* idéző funkcióját vizsgálták (Romaine & Lange, 1991; Fox & Robles, 2010). Az angolban ezt úgy lehet tesztelni, hogy a *be + like* formulát kicseréljük a *say* igével és koherens megnyilatkozást kapunk. Kutatásomban hasonlóan jártam el, az *izé*t a mond igére cseréltem. A bemutatott példámban a beszélő egy ismerős szavait idézi, azonban ezt nem fejezi ki explicit módon, csupán az *izé* szó használatával vezeti be azt. Ez megerősítette a hipotézisemet, miszerint az *izé* függő beszédben képes idéző funkciót betölteni.

Jelen tanulmány hozzájárul a magyar nyelvben egyedülálló nyelvi jelenség, az *izé* hazai szakirodalmának bővítéséhez. Tanulmányomban amellett érveltem, hogy az *izé* kitöltőelem megbélyegzettsége, stigmatizáltsága ellenére számos funkciót képes betölteni a spontán társalgásokban, a beszédtervezési folyamatban és a diskurzus szervezésében. Az *izé* sokszínűségének bemutatása révén tanulmányom elősegíti a jelenség megértését és elfogadását, ezáltal hozzájárulhat a nyelvi stigmatizáltság csökkentéséhez.

Hivatkozások

Butters, R. (1982). Editor's note [on be like 'think'. *American Speech*, *57*, 149.

Bóna, J. (2023). *Fluencia és diszfluencia a beszédben: A beszéd folyamatossága a szupraszegmentális szint temporális jellemzői és a megakadások tükrében (Akadémiai doktori értekezés*. Budapest: Eötvös Loránd Tudományegyetem.

Clark, H. (1994). Managing problems in speaking. *Speech Communication*, *15*, 243–250.

Clayman, S. (2013). Turn-constructional units and the transition-relevance place. In J. Sidnell, & T. Stivers (Eds.), *The handbook of conversation analysis* (p. 150–166). Oxford: Wiley-Blackwell.

Coulmas, F. (1992). *Die Wirtschaft mit der Sprache: Eine sprachsoziologische Studie*. Frankfurt: Suhrkamp.

Davis, B., & Maclagan, M. (2010). Pauses, fillers, placeholders and formulaicity in alzheimer's discourse: Gluing relationships as impairment increases. In N. Amiridze, B. Davis, & M. Maclagan (Eds.), *Fillers, pauses and placeholders* (p. 189–215). Amsterdam/Philadelphia: John Benjamins.

Diewald, G. (2013). Same same but different" – modal particles, discourse markers and the art (and purpose) of categorization. In L. Degand, B. Cornillie, & P. Pietrandrea (Eds.), *Discourse markers and modal particles: Categorization and description* (p. 19–46). Amsterdam: John Benjamins.

Dér, C. I. (2016). Diskurzusjelölő-társulások a magyar spontán beszédben. In T. Gecső (Ed.), *Ikon, nyelvi jel, szimbólum: Nem természetes jelek a kommunikációban* (p. 21–26). Budapest: Tinta Könyvkiadó.

Dér, C. I. (2017). A hát multifunkcionalitása a beszédműfajok és a diskurzusjelölő-társulások függvényében. *Beszédkutatás*, (p. 169–184).

Fabulya, M. (2007). Izé, hogyishívják, hogymondjam: Javítást kezdeményező lexikális kitöltőelemek. *Magyar Nyelvőr*, *131*, 324–342.

Fox, B. (2010). Introduction. In N. Amiridze, B. Davis, & M. Maclagan (Eds.), *Fillers, pauses and placeholders* (p. 1–9). Amsterdam/Philadelphia: John Benjamins.

Fox, B., & Robles, J. (2010). It's like mmm: Enactments with it's like. *Discourse Studies*, *12*, 715–738. doi:10.1177/1461445610381862.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, *31*, 931–952.

Grice, P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics* (p. 41–58). New York: Academic Press volume 3.

Grétsy, L., & Kovalovszky, M. (Eds.) (1980). *Nyelvművelő kézikönyv I*. Budapest: Akadémiai Kiadó.

Gyarmathy, D. (2012). Az izé funkciófüggő realizációi. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* (p. 178–194). Budapest: Akadémiai Kiadó.

Gyarmathy, D. (2015). Izé – a szó, amivel minden megmagyarázható. Élet és tudomány. URL: `https://eletestudomany.hu/ize-a-szo-amivel-minden-megmagyarazhato`.

Gyarmathy, D. (2017). *Megakadásjelenségek a magyar spontán beszédben*. Budapest: MTA Nyelvtudományi Intézet.

Gyarmathy, D., & Neuberger, T. (2013). The hungarian filler word izé in children's and adults' spontaneous speech. Paper presented at the XV. Summer School of Psycholinguistics, Balatonalmádi, May 26–31.

Gósy, M. (2000). A beszédszünetek kettős funkciója. *Beszédkutatás*, (p. 1–14).

Gósy, M. (2003). A spontán beszédben előforduló megakadásjelenségek gyakorisága és összefüggései. *Magyar Nyelvőr*, *127*, 257–277.

Gósy, M. (2004). A spontán magyar beszéd megakadásainak hallás alapú gyűjteménye. *Beszédkutatás*, (p. 6–18).

Gósy, M. (2005). *Pszicholingvisztika*. Budapest: Osiris Kiadó.

Gósy, M., Gyarmathy, D., Horváth, V., Gráczi, T., Beke, A., Neuberger, T., & Nikléczy, P. (2012). Bea: Beszélt nyelvi adatbázis. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* (p. 9–25). Budapest: Akadémiai Kiadó.

Hayashi, M. (2013). Turn allocation and turn sharing. In J. Sidnell, & T. Stivers (Eds.), *The handbook of conversation analysis* (p. 167–190). Oxford: Wiley-Blackwell.

Hayashi, M., & Yoon, K.-E. (2010). A cross-linguistic exploration of demonstratives in interaction: With particular reference to the context of word-formulation trouble. In N. Amiridze, B. Davis, & M. Maclagan (Eds.), *Fillers, pauses and placeholders* (p. 33–66). Amsterdam/Philadelphia: John Benjamins.

Horváth, V. (2014). *Hezitációs jelenségek a magyar beszédben*. Budapest: ELTE Eötvös Kiadó.

Huszár, A. (2005). *A gondolattól a szóig: A beszéd folyamata a nyelvbotlások tükrében*. Budapest: Tinta Könyvkiadó.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (p. 13–23). Philadelphia: John Benjamins.

Kitzinger, C. (2013). Repair. In J. Sidnell, & T. Stivers (Eds.), *The handbook of conversation analysis* (p. 229–256). Oxford: Wiley-Blackwell.

Kondacs, F. (2017). Az óvodások megnyilatkozásairól az izé kapcsán. In T. Váradi (Ed.), *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből 2017* (p. 109–119). Budapest: MTA Nyelvtudományi Intézet.

Kosmala, L., & Crible, L. (2021). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech*, . doi:10.1177/00238309211010862.

Labov, W. (1979). A nyelv vizsgálata társadalmi összefüggésben. In C. Pléh, & T. Terestyéni (Eds.), *Beszédaktus – kommunikáció – interakció* (p. 365–398). Budapest: Tömegkommunikációs Kutatóközpont.

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* (p. 95–134). Cambridge: Cambridge University Press.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Marcsenkoné Kondacs, F. (2023). A hát, az izé és az ugye diskurzusjelölők jelentéseinek ábrázolása a kibővített natural semantic metalanguage modellben

(doktori disszertáció. *Szegedi Tudományegyetem*, . Nyelvtudományi Doktori Iskola.

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

Németh, Z. (2021). A nemlexikális öö hang mint diskurzusjelölő magyar nyelvű társalgásokban. *Beszédtudomány*, *2*, 173–206. doi:10.15775/Besztud.2021.173-206.

Pap, A. (2011). Adalékok a nyelvi benyomáskeltés stratégiáihoz: A leechi udvariassági elvek megvalósulása a magyarok nyelvhasználatában. *Magyar Nyelvőr*, *135*, 78–89.

Podlesskaya, V. (2010). Parameters for typological variation of placeholders. In N. Amiridze, B. Davis, & M. Maclagan (Eds.), *Fillers, pauses and placeholders* (p. 11–32). Amsterdam/Philadelphia: John Benjamins.

Rieger, C. (2003). Repetitions as self-repair strategies in english and german conversations. *Journal of Pragmatics*, *35*, 47–69.

Romaine, S., & Lange, D. (1991). The use of like as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech*, *66*, 227–279.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

Schegloff, E. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.

Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*, 361–382.

Schegloff, E., & Sacks, H. (1973). Opening up closings. *Semiotica*, *8*, 289–327.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

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

Schirm, A. (2018). Diskurzusjelölő-társulások a szögedi szociolingvisztikai interjúban. *Alkalmazott Nyelvtudomány*, *18*, 1–16.

Tar, C. (2023). Az önjavítási jelenségek magyar nyelvű spontán, baráti társalgásokban. *Beszédtudomány*, (p. 75–104). doi:10.15775/Besztud.2022.75-104.

Urizar, X., & Samuel, A. (2014). A corpus-based study of fillers among native basque speakers and the role of zera. *Language and Speech*, *57*, 338–366. doi:10.1177/0023830913506422.

Vukov Raffai, E. (2016). A diskurzusjelölő-választások életkori sajátosságai az így, ilyen, hát, mondjuk, ugye esetében. *Magyar Nyelvőr*, *140*, 483–497.

Zaicz, G. (2006). *Magyar etimológiai szótár*. Budapest: Tinta Könyvkiadó.

Függelék

A példákban használt átírási konvenciók (Jefferson, 2004)

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

Difficulties in the perception of Mandarin Chinese vowels [ɤ] and [ɿ]/[ʅ] by Hungarian learners of Mandarin

Libo Fan

University of Szeged

Abstract

This study investigates how Hungarian learners of Mandarin Chinese identify the vowels [ɤ], [ɿ], and [ʅ], which are absent from Hungarian phonology. It focuses on how learners distinguish the mid vowel [ɤ] from high vowels [ɿ]/[ʅ] at different stages of learning and explores the factors influencing their perception. Two main variables are considered: the quality of input (native vs. non-native Chinese teachers) and the quantity of input (beginner vs. intermediate learners). Consonant context as a variable also appears in the analysis, along with qualitative formant measurements on the samples of the perception test, investigating and establishing a possible explanation for the results. Participants included 21 beginners and 10 intermediate learners. Beginners were divided into three subgroups based on teacher type: native Chinese, Hungarian L2 speakers of Chinese, and both. Intermediate learners were taught by both types of teachers. An X(AB) perceptual identification test was used to investigate the perception of Chinese vowels [ɤ] and [ɿ]/[ʅ] among Hungarian learners of Chinese. Results showed that [ɤ] was identified less accurately than the high vowels. Learners taught by a native speaker performed better, highlighting the importance of input quality. Surprisingly, intermediate learners did not outperform beginners, which may be due to orthographic interference and fossilisation of pronunciation skills rising from the elimination of pronunciation training in advanced classes. Overall, the study suggests that both teacher background and the writing system affect the perceptual identification.

1. Introduction

Mandarin Chinese (hereafter Chinese) has been the official language of China for a few decades. It is used in schools and universities, and on national radio and television broadcasts (Duanmu, 2007:4). An increasing number of Hungarian speakers learn Chinese from year to year (Simay et al., 2020), which raises questions on the similarities and differences between the two languages, with

Email address: flbwyf@gmail.com (Libo Fan)

Beszédtudomány – Speech Science 2025/2. 82–111. DOI: 10.15775/Besztud.2025.2.82-111

the aim to improve and adapt the teaching methods to the specific requirements of the target learners. The present study explores one of these questions: the perception, specifically the identification, of Chinese vowels [ɤ]/[ɿ]/[ʅ], that are neither part of the vowel phoneme system, nor appear as allophones in the speakers' mother tongue. In general, our goal is to describe the possible difficulties of Chinese learners with Hungarian as their mother tongue, and to discuss the possible explanations for them, in order to make the results usable in the future improvement of teaching methodology.

1.1. Rationales of the question of [ɤ] and [ɿ]/[ʅ] perception

An understanding of how learners acquire a new phonological system must account for both the linguistic differences between the native and target language, and the universal facts of phonology (Gass et al., 2013). When learning non-native languages, the influence of previous linguistic experience is particularly significant (Strange, 1995). Contrastive Analysis Hypothesis (CAH) assumed that learners tend to transfer the patterns of their native language structure to the foreign language, and it is also assumed that this is the major source of difficulty or ease in learning the structure of a foreign language. Similar structures are assumed to be easy to learn, while different ones are considered to be difficult (Lado, 1957:59). We treat all non-native languages as L2 languages in this study. Hungarian is L1 and the phrase "second language" L2 of Hungarian learners refers to Chinese in the present research; in other words, Chinese is L2, and Hungarian learners of Chinese are analysed. We have to note, however, that Chinese is the 3rd or 4th language for most speakers, since English and/or German (and often another foreign language) is compulsory in primary and secondary education in Hungary.

Eckman (1977) proposed the Markedness Differential Hypothesis (MDH), grounded in a phonological theory of markedness. According to this proposal, the most difficult structures to learn are those being both different and more marked at the same time compared to the corresponding native language structure. Chinese vowels [ɤ]/[ɿ]/[ʅ] that are the focus of the present study neither

exist in Hungarian, nor are common in natural languages. Therefore, these vowels can be considered difficult for Hungarian learners of Chinese.

In the present experiment, there are several reasons to analyse the identification between [ɤ] and [ɿ]/[ʅ]. First, the Chinese [ɤ] and [ɿ]/[ʅ] vowels, which are not part of the Hungarian vowel system, are allophones of vowel phonemes in Chinese /ə/ and /i/, respectively, and appear in the same consonantal contexts in Chinese. Second, based on Fan's questionnaire results (Fan, 2024), Hungarian learners of Chinese do not consider Chinese vowels [ɤ]/[ɿ]/[ʅ] to be difficult, while Chinese teachers reported [ɤ] causing various problems to the learners, and eight out of twenty participating teachers claimed that their students also face difficulties learning [ɿ]/[ʅ] vowels.

The Speech Learning Model (SLM) proposed by Flege (1995), the revised SLM (SLM-r) proposed by Flege & Bohn (2021), and the Perception Assimilation Model (PAM) stated by Best (1995) and Best et al. (2001) suggest that segments that are not part of the language learners' mother tongue would be more difficult to learn, as the students need to acquire both their perception and production. The production of the segments in question was addressed by Juhász (2020). Her results showed that the L2-learners' pronunciation of [ɤ] was significantly different from that of the Chinese native speakers, but [ɿ] and [ʅ] did not show any significant difference. The present research aims to broaden the scope of the investigation related to this issue, focusing on the perception of the Chinese L2 sounds at hand.

1.2. Chinese and Hungarian vowels: phonological and phonetic aspects: with focus on [ɤ] and [ɿ]/[ʅ]

As mentioned before, the L1 phonological system directly affects the acquisition of L2 speech sounds (SLM, PAM). Thus, in the next section, L1 and L2 (i.e., Hungarian and Chinese) vowel systems are compared to discover the problematic aspects of the [ɤ] and [ɿ]/[ʅ] speech sounds. The Hungarian vowel system includes 14 phonemes /i, iː, u, uː, y, yː, ø, øː, ɑ, aː, o, oː, ɛ, eː/ that do not have allophonic variation. The Chinese vowel system includes five phonemes

/a, ə, i, u, y/ with 9 contextual allophones altogether (Figure 1). Zhu's (2010) phonological-phonetic theory is applied in my research, as the present study only focuses on monophthongs in open syllables. [e] in Figure 1 only occurs in diphthongs, while other vowel variants, appearing in diphthongs and triphthongs, are not shown in the present study either. As the present study focuses on the perception of mid [ɤ] and high [ɿ]/[ʅ] speech sounds, we describe these in more detail.

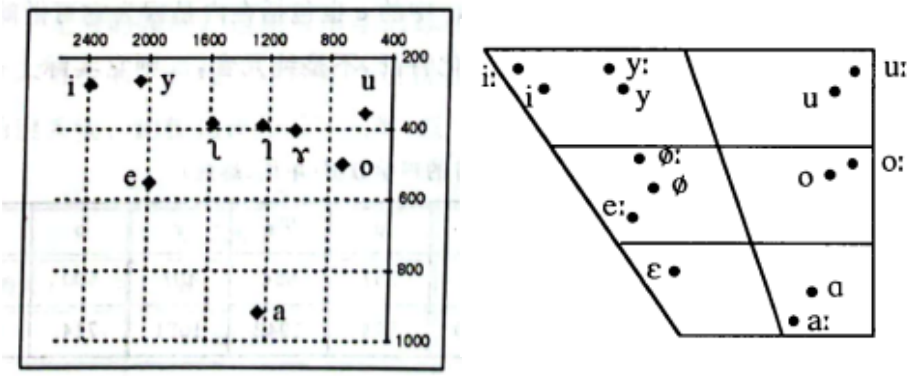


Figure 1: Mandarin vowels (left) (Zhu, 2010: 268); and Hungarian vowels (right) (Szende, 1994).

The phonological category and behaviour of the two apical segments [ɿ] and [ʅ] are subject to debate. Even though their acoustic structure is characterized by a formant structure indexing their vocalic nature from a purely phonetic viewpoint, some scholars regard them as syllabic fricatives, hence the segments are transcribed with the following IPA graphemes (even though they are the same speech sounds): [z] and [z̩] (Chao, 1934; Hartman, 1944; Pulleyblank, 1984; Lin, 1989; Wiese, 1997; Duanmu, 2000, 2007. cf: Lee-Kim, 2014). The second proposal is the approximant account, wherein [ɿ] and [ʅ] are written as [ɹ] and [ɻ], respectively (Lee-Kim, 2014). In the present study, we adhere to the traditional Chinese phonological analysis, i.e. we treat them as apical vowels (Cheng, 1966, 1973; Lee & Zee, 2001), since [ɿ] and [ʅ] are in complementary distribution with the high front vowel [i]. [ɿ] only occurs after the non-retroflex

consonants [ts, tsʰ, s], while [ɿ] only appears after the retroflex segments [tʂ, tʂʰ, ʂ, ɻ], and [i] occurs in other environments. This speech sound variation based on the preceding consonant can be summed up as contextual allophony with a clear-cut complementary distribution (Duanmu, 2000). Based on X-ray images taken by Zhou & Wu (1963), the tongue tip/blade gesture of [ɿ]/[ʅ] inherited from the preceding dental and retroflex consonants remains nearly unchanged for the following voiced period. The retroflex or non-retroflex feature of the preceding consonant spreads onto the vowel, [ʅ] is called the retroflex apical vowel, while [ɿ] is called the dental apical vowel (Cheng, 1973:13). Zhang (2003) performed an identification test and a discrimination test for isolated synthesized [ɿ] and [ʅ] stimuli. The results of the discrimination test show that the perception of [ɿ] and [ʅ] by native Chinese speakers is non-categorical, compared to the extreme categorical perception of stops. The identification test shows that the patterns of the first two formants determine the categorical boundary for phonemic distinction between [ɿ] and [ʅ], but F3 is also a potent cue for identifying the two apical vowels. Zhang states that maybe the listeners use some cues other than those manipulated in the identification of the stimuli as phonemes for discriminating the stimuli (Liberman et al., 1961).

The question of [ɤ] is generally clear. It is a mid back, unrounded vowel (Duanmu, 2007:37; Lin, 2007:73; Chen et al., 2019) and is generally regarded as a contextual allophone of the mid central vowel /ə/ in the nucleus of open consonant-vowel (CV) syllables. [ɤ] enjoys a more expansive phonotactic context, it does not only occur after [ts, tsʰ, s] and [tʂ, tʂʰ, ʂ, ɻ], but also some other consonants like [k, kʰ, x, t, tʰ, n, l]. Therefore, the mid vowel [ɤ] and the apical vowels [ɿ]/[ʅ] have a different status in the Chinese vowel system. [ɿ] and [ʅ] occur in limited contexts, [ɤ] occurs in more contexts. For the present study, it is important that [ɿ]/[ʅ] and [ɤ] share their consonantal contexts: [ɤ] can appear in all syllables where the /i/ allophones in question can.

Despite being allophones of the same vowel /i/, [ɿ] and [ʅ] segments are produced at different places of articulation (post-alveolar & dental), hence they are denoted by two different IPA-symbols and considered two distinct segments for

L2 learners to acquire. However, these vowels are not perceptually differentiated by Chinese natives (as it is found by Zhang, 2003), resulting from the fact that that they can be predicted by the consonantal context because they are allophones appearing in complementary distribution. Therefore, we will consider the apical vowels as one category within the identification task, and the test only contrast the following two pairs: [ɿ] with [ɤ] and [ɻ] with [ɤ]. Precisely the present research only focuses on the syllables starting with consonants [tʂ, tʂ^{h}, ʂ, ʐ, ts, ts^{h}, s], which can precede both the /i/ and /ɤ/ allophones in question, and ending with vowels [ɤ, ɿ, ɻ] (Table 2).

1.3. Orthographic considerations of non-native perception

Wang (2001) studied the vowel perception of Japanese and Korean learners of Chinese. She chose the identification of two or three vowels specifically for the relation of the vowel system of the given mother tongue and Chinese in a similar (X(AB)/X(ABC)) identification test as in our study. The results of Wang's research (2001) also raise the question of the perception of [ɻ] by listeners with Korean and Japanese as their mother tongue. This vowel does not exist in these two languages. When [ɻ] was played, they had to choose an option from A[ɤ]/B[ɻ]. The Korean students chose [ɤ] nearly 3 times more than [ɻ]. On the contrary, Japanese students did not choose [ɤ] at all. Wang pointed out that this might be affected by the orthographic symbol <i>, since in Chinese, <i> represents [i].

L2 orthographic input has been confirmed to show effects on the acquisition of non-native phonological and phonetic patterns. L2 orthographic input interacts with the acoustic input, influencing L2 learners' mental representations of L2 phonology, and orthography-induced pronunciations may be part of the acoustic input for instructed learners (Bassetti, 2008). Erdener & Burnham (2005) found that, while all L2 learners were more capable of repeating L2 words when they saw graphemes of the words, the effect was stronger or weaker depending on the level of phonological transparency of both L1 (native language) and L2 orthographies. It is found that native users of transparent

L1 writing systems are more negatively affected by a less transparent L2 orthography. In our case, the Hungarian writing system is phonologically highly transparent, whereas Pinyin, which is used in language teaching as a base for Chinese, is more opaque. Pinyin is the transcription of Chinese characters using the graphemes of the Latin alphabet. < > is used to represent graphemes in the present study. In Chinese, the grapheme <e> represents /ə/, therefore also its allophones: [ɛ], [ə], [ɤ], [e], [o] (Xu, 1980:33, cf: Duanmu, 2007:37). And the grapheme <i> represents /i/, therefore [j], [i], [ɿ] and [ʅ]. However, in Hungarian, <e> represents /ɛ/, therefore [ɛ], and <i> represents /i/, and therefore its allophones. Thus, in our case, the grapheme <i> denotes two different segments in Chinese (the apical vowels [ɿ] and [ʅ]). As we can see in terms of orthography, these apical vowels are differentiated from the mid back vowel because they are denoted by different graphemes, i.e., [ɿ] and [ʅ] are denoted by the grapheme <i> and the mid vowel [ɤ] is denoted by <e>. However, while <i> in Hungarian also denotes /i/, but there is no allophonic variation, <e> denotes [ɛ] also without allophonic variation. Thus, in L2 Chinese, learners are required to associate different speech sounds with the same grapheme, by recognizing the determining role of the onset consonant, which might pose difficulty since the segment-to-grapheme link in the L1 is almost exclusive and apparent and not characterized by this variation. The present study also considers the phonetic context. Based on PAM (Best, 1995), the consonants surrounding a vowel affect how that vowel will be perceptually assimilated. Flege & Bohn (2021) also state that it is important to note that the context in which input is assessed may also influence how well the input is consolidated and thus indirectly influence speech learning.

Finally, the SLM also proposed that L2 learners gradually "discern" L1-L2 phonetic differences as they gain experience using L2 in daily life, and that the accumulation of detailed phonetic information with increasing exposure to statistically defined input distributions for L2 sounds will lead to the formation of new phonetic categories for certain L2 sounds (Flege, 1995). Furthermore, in SLM-r, Flege & Bohn (2021) state that the quality of input has been largely

ignored in L2 speech research even though it may well determine the extent to which L2 learners differ from native speaker. In the present study, the learner's experience with the quantitative input (two groups with different L2 experience) and the qualitative dimension (students with teachers of different L1, Hungarian and Chinese) are considered as well.

The relationship between perception and production is that many L2 production errors have a perceptual basis. Flege (1995) suggested that L2 production accuracy is limited by perceptual accuracy. And the PAM also holds that the pronunciation difficulty encountered by L2 learners is determined by perceptual limitations (Best, 1995; Best et al., 2001). However, the SLM-r (revised Speech Learning Model) assumes that L2 segmental production and perception coevolve without precedence (Flege & Bohn, 2021).

1.4. Formant values of [ɤ] and [ɿ]/[ʅ]

The [ɤ] and [ɿ]/[ʅ] vowels in question were studied in the pronunciation of native speakers and language learners of Chinese by Fan (submitted). We assume that there is a direct link between the characteristics of Chinese vowel formants and the perceptual traits of Hungarian L1 speakers. The production results of a native Chinese speaker showed that the first formants are lower in the two apical vowels [ɿ]/[ʅ], while the second formants are higher than the mid vowel [ɤ]. The second and third formants of the retroflex high vowel [ʅ] are closer to each other, mainly due to a shift in both values.

Zhu's (2010) data (Figure 1, left) show no difference in the first formant between the two high vowels, while Fan's (submitted) do.

We also did a qualitative comparison between acoustic values of Chinese [ɤ], [ɿ], [ʅ] and the Hungarian perceptual vowel map/space (Figure 3).

We can see the Hungarian perceptual vowel map/space and the Chinese acoustic data. Comparing Figure 2 and Figure 3, we can see that the acoustic measurements of [ɤ] and [ɿ]/[ʅ] are mapped onto the same Hungarian perceptual category of <ö/ő> ø. And it can be seen that the three analysed vowels [ɤ] and [ɿ]/[ʅ] are quite close to each other in Figure 1 as well. Hungarian learners

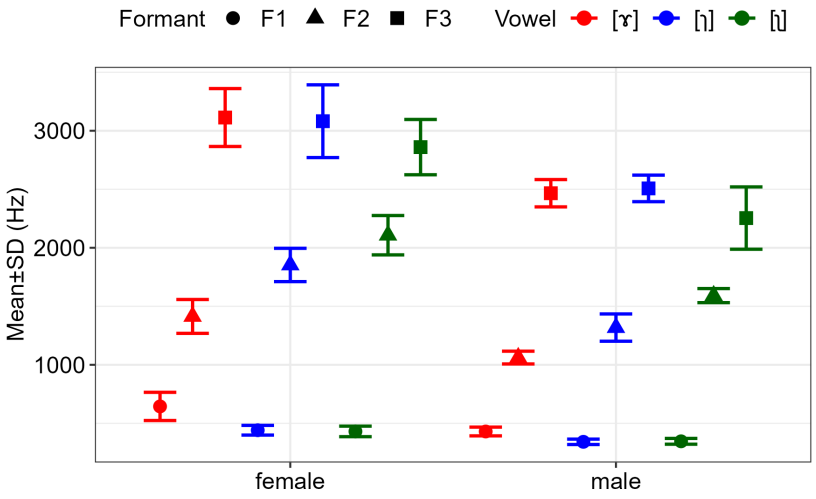


Figure 2: Formants of the high and mid vowels (Fan, submitted).

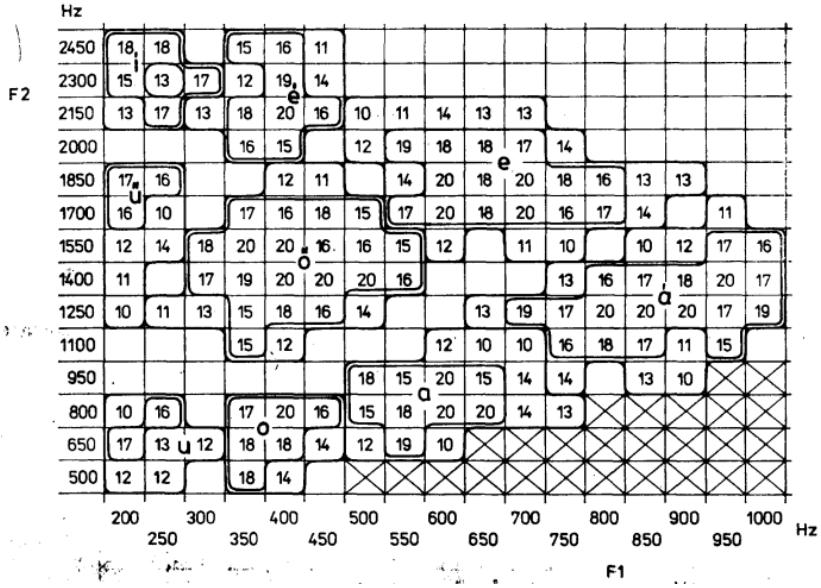


Figure 3: Plot of Hungarian vowels (Kiss, 1985).

of Chinese produce [ɤ] in a more acoustically palatalized way (Juhász, 2020). These facts might result in interference in the perception/identification of the

vowels, which is the motivation for designing a two-alternative forced-choice test in 2.2.

1.5. Hypotheses

Based on the theoretical considerations, the production study by Juhász (2020), and Fan's (2024) questionnaire and interview results, the following hypotheses were formulated along the two variables investigated in this perceptual analysis (i.e., quality and quantity of the L2 input):

i) The correct identification of [ɤ] is lower than that of the apical vowels [ɿ] and [ʅ], as their production was found to be native-like, suggesting a possibly more stable perception.

ii) The perception of the attested vowels may not show a difference between advanced learners and beginners, as the ratio of pronunciation and perception teaching is lower in the case of the former, and they have been more exposed to Pinyin.

iii) The learner groups having a native speaker as at least one of their teachers will have higher correct response ratios, as they experience native pronunciation in a higher ratio.

iv) Identifying [ɤ] and [ɿ] vowels after non-retroflex [ts, ts^h, s] could induce a lower identifying rate than [ɤ] and [ʅ] identification after retroflex [tʂ, $tʂ^h$, ʂ, ɻ], as the F2 and F3 show a larger difference between [ɤ] and [ʅ].

2. Methods

2.1. Subjects

31 native Hungarian speakers participated in a discrimination task (Table 1). All participants were born and raised in Hungary. They were divided into four groups: The advanced group included 10 subjects who had been learning Chinese for five semesters. During the first two semesters, they were taught online by a native Chinese and a native Hungarian L2-speaker of Chinese, and

from the third semester, they started attending physical classes with the same teachers. The rest of the participants (21) were beginners, who had been learning Chinese for one semester. The beginners were divided into three groups: 7 of them were taught exclusively by a native Chinese speaker, 6 of them by a native Hungarian L2-speaker of Chinese, and 8 of them by both a native Hungarian L2-speaker of Chinese and a native Chinese speaker. The native Chinese teacher was consistent across all three groups that included a native teacher.

No participant reported any hearing problems. One student in the advanced group had spent half a year in Taiwan as an exchange student studying economics, but she did not learn Chinese. The remaining participants had never been to China before. All participants use Hungarian as their primary language in daily life. They all could also speak English. Some participants had learned other languages as well. None of the sounds included in the task ([ɤ]/[ɿ]/[ʅ]) are part of the English sound system either, thus in the present study, the interference of English is not considered.

All the participants were university students aged between 18 and 22 years.

Table 1: Participants.

	teacher(s)	**no. of students**	**group name**	**no. of semesters**
beginner groups	native Chinese speaker	7	BegChi	1
	native Hungarian speaker	6	BegHu	1
	joint teachers	8	Begmix	1
advanced group	joint teachers	10	Intmix	5

2.2. Stimuli

All stimuli consisted of CV-syllables ending in [ɤ], [ɿ] or [ʅ], with the high vowels depending on the expected vowel after the onset (Table 2). Each stimulus represents a lexical item in Chinese. To exclusively attest vowel differentiation and exclude the possible effect of the tones, the present study introduces only

the results for items in falling tone 4. Tone was limited to tone 4 because all 14 test items are meaningful words in Chinese with tone 4, while not all of them exist with other tones. The vowels [ɿ] or [ʅ] appear after seven consonant onsets [ts, ts^{h}, s, tʂ, tʂ^{h}, ʂ, ɻ], hence 14 items are analysed in this study. Besides the 14 stimuli recorded for the present study, 114 further items (test items for different vowel allophones and tones and distractors) were played three times each (N = 384) in randomized order. The actual test was preceded by a short training period with 3 non-test items.

The test items were recorded by 6 native speakers in a sound-treated room using a head-mounted microphone via Speech Recorder (Christoph & Klaus, 2004). After recording the sounds, a professional native Chinese voice actor was asked to judge the six speakers' pronunciation. The voice actor had passed a Mandarin Chinese proficiency test, was born in Hebei and finished his university in Beijing, and he had dubbed Chinese textbooks as well, therefore he can be accounted as a high proficiency speaker with a well based judgmental reliability. The speaker whose reading was used for the perception test was chosen based on the scores given by the native actor. The one closest to the ideal Chinese pronunciation was selected. The male speaker is from Mainland China, and he had been studying in Hungary for 3 years at the time of recording. During these three years, he was mostly in a Chinese-speaking environment, except for attending classes at the university.

Table 2: Stimuli.

non-retroflex		**retroflex**	
Pinyin	**IPA**	**Pinyin**	**IPA**
<zè> – <zì>	[tsɤ] – [tsɿ]	<zhè> – <zhì>	[tʂɤ] – [tʂʅ]
<cè> – <cì>	[ts^{h}ɤ] – [ts^{h}ɿ]	<chè> – <chì>	[tʂ^{h}ɤ] – [tʂ^{h}ʅ]
<sè> – <sì>	[sɤ] – [sɿ]	<shè> – <shì>	[ʂɤ] – [ʂʅ]
		<rè> – <rì>	[ɻɤ] – [ɻʅ]

In Table 2, the retroflex feature indexes the consonant context, since [tʂ, tʂʰ, ʂ, ɻ] are retroflex sounds.

2.3. Experiment design

The present research examined identification of the vowels discussed above in a two - alternative forced-choice (2AFC) test by adult native Hungarian speakers. The perception test was administered in a quiet room through headphones using Praat ExperimentMFC (Boersma & D., 2022).

The subjects were instructed to select the word they heard as soon as possible after listening to the stimulus (Table 2). After the subject listened to a single stimulus, they had to click one of two responses. In other words, the participant had to select one from two possible answers displayed in Pinyin: <zè>/<zì>, <cè>/<cì>, <sè>/ <sì>, <zhè>/<zhì>, <chè>/<chì>, <shè>/<shì> and <rè>/<rì>. This means that the participants did not have to decide between the two high vowels, but between the mid and one of the high vowels. The response and the reaction time were recorded. The listeners also had to judge how sure they were in their answer on a 1 to 5 scale where 1 meant absolutely unsure, 5 meant absolutely sure. The present study will not analyse the goodness responses.

2.4. Statistics

The answers were analysed in R (R Core Team, 2022). The correctness of the answers was analysed using Binomial Generalized Linear Mixed Models (BGLMM), and for the reaction times Generalised mixed effects linear models were run for logistic distribution (GLMM, lme4: Bates et al., 2015; lmerTest packages: Kuznetsova et al., 2017). A Tukey post hoc test was administered to attest the effects of the interactions (emmeans package: Lenth, 2021). The models were built in a top-down selection method: the simplest model was chosen and that was still not significantly different from the possible largest, converging model.

The correctness of the answer was set as the dependent variable in the BGLMMs, while reaction time was the dependent variable in the GLMMs. The factors were the following: phoneme category (i.e., mid or high vowel), learner group (advanced, beginner with Chinese teacher, beginner with Hungarian teacher, beginner with joint teachers), and tongue tip position (retroflex or not). The models including all three factors did not converge, therefore the tongue tip position was eliminated and attested separately. The p-value of the final model was extracted by Anova (car package: Fox & Weisberg, 2019). In order to analyse the possible effect of the retroflex context and the retroflex feature of the vowel on the correctness of the identification, the results for the mid and high vowels were tested separately using two further BGLMMs. The correctness of the answer was set as the dependent variable, and the retroflex feature and the learner group were set as factors. The model selection and the extraction of p-value were run as described above.

All figures were drawn using ggplot2 (Wickham, 2016).

2.5. Formant analysis

The vowel formants of the test items were also measured to be used in the interpretation of the contextual effect in the perception results. The first three formants (F1, F2, F3) of the stimuli in Table 2 were measured in Praat automatically by a script. The vowels were labelled from the start to the end of the F2 in the oscillogram and spectrogram. In the case of [ɻʏ] and [ɻʉ], the spectrogram had to be set to the range of 0–8kHz in order to let the higher frequency frication appear in the view range. In the case of these sequences, the loss of this higher frequency frication was considered the start of the vowel. The formant range was set to 5 kHz in general, except in the case of the four [ʅ] vowels, where the F2 and F3 fall close to each other leading to mis-measurements. After manual checking, the formant range was set to 4.5 kHz in these four cases. The further settings were left as standard (5 formants for male voice). The measurements were taken at the mid 40 ms of the vowels.

3. Results

3.1. Reaction times

Reaction times between the correct and incorrect answers did not show any differences (Figure **??**, left). Reaction times for the correct answers showed some tendencies: reaction times for the correct answers for <i> [ɪ]/[ɪ] were shorter than for <e> [ɤ] (Figure **??**, right). Accordingly, the best fitting Generalised Mixed Effects Logistic Regression included the interaction of the vowel and the group, a random slope on the vowels by learners, and a random intercept by the stimuli. However, the results did not indicate significant differences by any of the factors or their interaction. As there was no significant difference, reaction time is not included in the latter results.

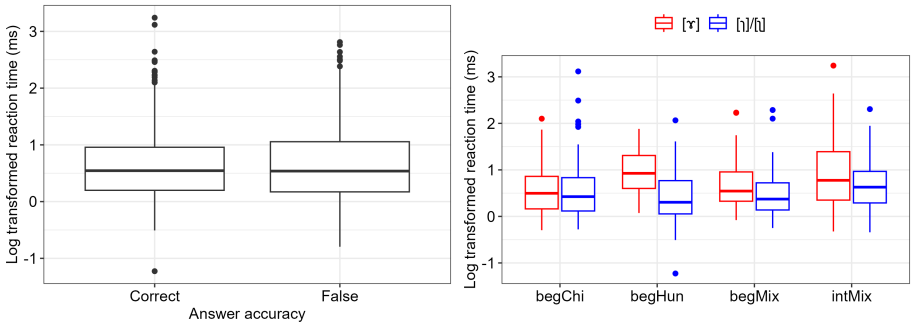


Figure 4: The reaction times (log-transformed, ms). Left: Reaction times for the correct vs. false answers, right: Reaction times for the correct answers [ɤ] vs [ɪ]/[ɪ] among the learner groups.

3.2. Accuracy rate of [ɤ] and [ɪ]/[ɪ]

The accuracy rate was calculated speaker-by-speaker for the mid and the high vowels separately. The ratio of correct answers revealed a noticeable difference in performance when identifying the two vowel phonemes. The results were grouped by the vowels and the groups. The mean value is not used in the present research because the data are not normally distributed, thus the median and the interquartile range are used for description. The best fitting

model was the one that included the interaction of the vowel and learner group with a random intercept by learners.

Based on the statistical results of the generalized linear mixed model, there is a significant difference between the accuracy rates of <e> [ɤ] and <i> [ɿ]/[ʅ] (results for the vowel factor: $\chi^2(1, 1302) = 64.99$, $p < 0.001$) (Figure 5). In addition, the highest accuracy rate for <i> [ɿ]/[ʅ] is 0.76, suggesting that the apical vowels also cause some problems. The median of <e> [ɤ] from different groups is lower than that of <i> [ɿ]/[ʅ] (Figure 6). The accuracy rates of <e> [ɤ] of begChi, begHun, begmix and intmix are 0.60, 0.30, 0.64 and 0.58, respectively. The median range of <e> [ɤ] is from 0.30 to 0.64, but the range of <i> [ɿ]/[ʅ] is from 0.67 to 0.76. This result is in agreement with Juhász's production test (2020).

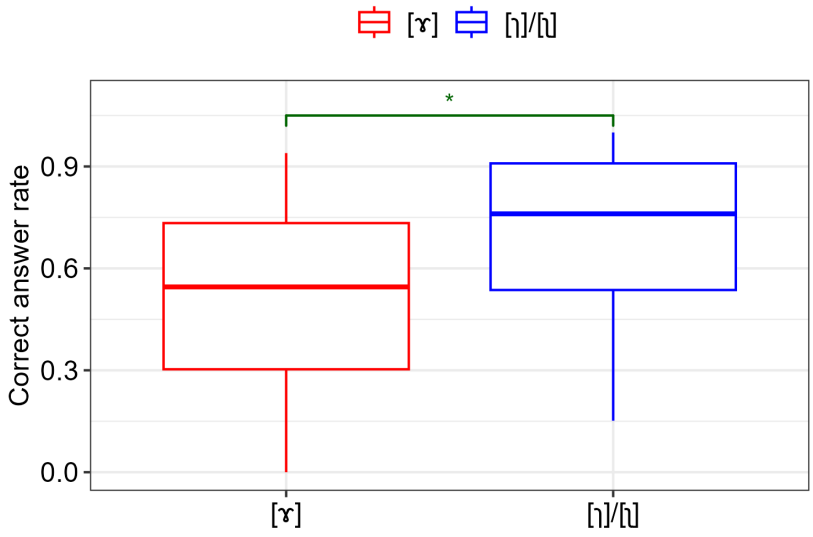


Figure 5: The rate of answer correctness for the mid vs. high vowel regardless of learner group.

The interaction of the two factors was also significant ($\chi^2(3, 1302) = 13.38$, $p = 0.004$) (Figure 6). According to the Tukey post hoc test, there is a significant difference (i) between the begChi and begHun, and also (ii) between

the begmix and begHun groups' results in the case of <e> [ɤ]. This means that among the beginner groups, the accuracy rate may be influenced by the teachers. The begChi and begmix (with a native Chinese teacher) groups have significantly higher accuracy rates than the begHun group does (with a native Hungarian teacher) for [ɤ]. Comparatively, the accuracy rate of <e> [ɤ] in the begHun group is 0.30, while it is 0.60 and 0.64 in the begChi and begmix groups, respectively. As for apical vowels, the accuracy rates of the begChi and begmix (with a native Chinese teacher) groups were also higher than in the begHun group (with a native Hungarian teacher). Comparatively, the accuracy rate of <i> [ɿ]/[ʅ] in the begHun group is 0.67, while it is 0.77 and 0.75 in the begChi and begmix groups, respectively.

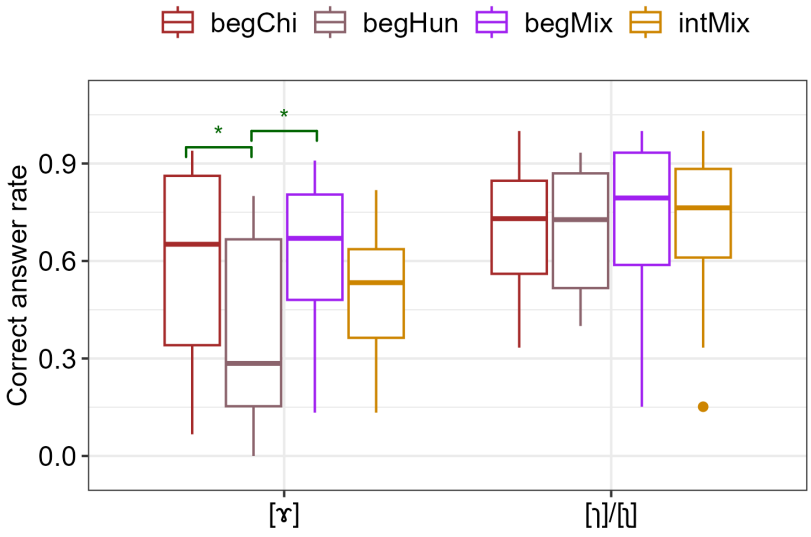


Figure 6: Correct answer rate for the vowels within the learner groups.

The advanced (intmix) group's results did not show significant differences from any of the beginner groups, which means that their results were not significantly better despite having more experience with the language.

3.3. Phonetic context for vowel perception

The possible influence of tongue tip position (retroflex or not) is shown in Figure 7 and Figure 8. We have to emphasize that the feature of being retroflex or not is intrinsic for the high vowels, i.e., these apical vowels are distinguished by this specific feature, while in the case of the mid vowel, this appears only as a contextual coarticulatory effect and thus is not supposed to appear throughout the entire duration of the vowel. Based on the ultrasound study of Lee-Kim (2014), both the tongue tip raising and tongue back retraction of [ɿ] and [ɻ] are maintained throughout the entire syllable. The accuracy rate of identifying the non-retroflex vowel [ɿ] is higher than that of the retroflex vowel [ɻ] (Figure 7). While this seems to be a difference within the groups with a native teacher (Figure 8), the best fitting generalized linear mixed model was the one that included only the factor of retroflection, but not the group, and included a random intercept by the learner. There was a significant difference between the retroflex and non-retroflex vowel ($\chi^2(1, 651) = 9.536$, $p = 0.002$).

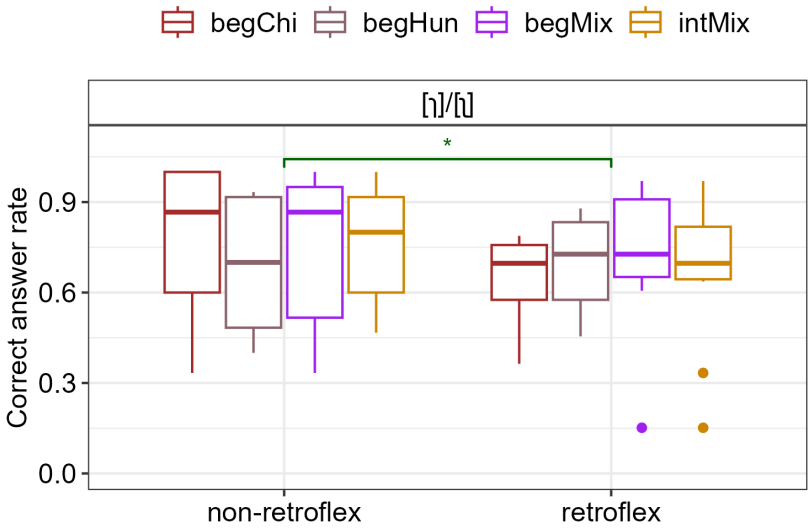


Figure 7: Correct answer rate of [ɿ]/[ɻ].

In the case of the mid vowel (Figure 8), the best fitting generalized linear mixed model was the one that included the retroflex and group factors with their interaction and random intercept by the learner. There was a significant difference between the groups ($\chi^2(3, 651) = 8.081$, $p = 0.044$), and also the interaction of the group and retroflection was significant ($\chi^2(3, 651) = 9.081$, $p = 0.028$), but no significant difference was found between the vowels. Based on the Tukey post hoc test, there is a significant difference between the begChi and begHun groups, and between the begHun and begmix groups, where the begHun group has lower correct answer rates. Although the effect of the group factor was significant meaning that there is a general difference between the begHun and the other two beginner groups, the post hoc test for the interaction of the two factors made it clear that if the contextual effect is considered, the lower perception rate in the begHun group from the other two beginner groups appears in the vowels following a retroflex consonant, while the difference does not reach the level of significance in the non-retroflex context.

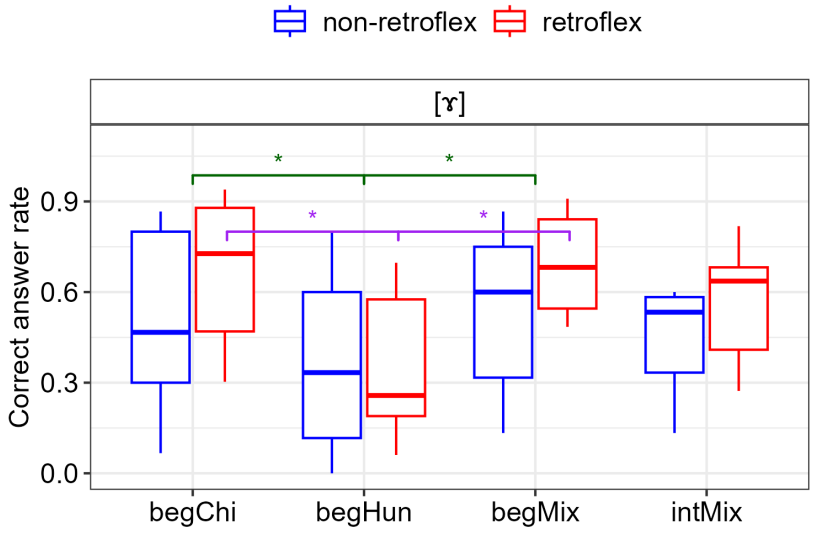


Figure 8: Correct answer rate of [ɤ].

3.4. Formant analysis

The mixed results for the contextual effect on the perception raised the question of how the formants of the mid vowel change between retroflex and non-retroflex contexts as compared to the high vowels.

Figure 9 shows the formant frequencies for the test items, and Table 3 shows the mean values. The present data lies in 14 items, and the data is only from one person, therefore we do not intend to draw general conclusions. However, in these specific data, we can see that the F2 and F3 values get closer in a retroflex context/pronunciation. What is more important for the results above, is that the formant values of the mid and high vowel are closer in a non-retroflex context, and further apart following a retroflex consonant. The larger difference must appear due to the inherent retroflex feature in the high vowel. While not willing to draw large conclusions in general for Chinese vowels, for the present data we can say the following. The formant values of the mid and high vowels lay considerably closer to each other in the present stimuli in the non-retroflex scenario compared to the retroflex scenario, in correspondence with Lee & Zee (2001). We assume that phonetic context may influence these vowels' perception by Hungarian learners of Chinese. As the results in 3.3 showed, the correct identification for [ɤ] shows higher differences across the learner groups in the retroflex context than in the non-retroflex one. However, it is the other way for the /i/ allophones. The retroflex context resulted in somewhat higher correct answer rates for the beginner groups taught at least partially by a native speaker, while the third beginner group had lower accuracy rate in this context. The perception of the high vowel was significantly lower when it was retroflex in general, regardless of the speaker groups. At the present point, we can only draw the conclusion that the effect of the context should be addressed in studies more focused on this specific question.

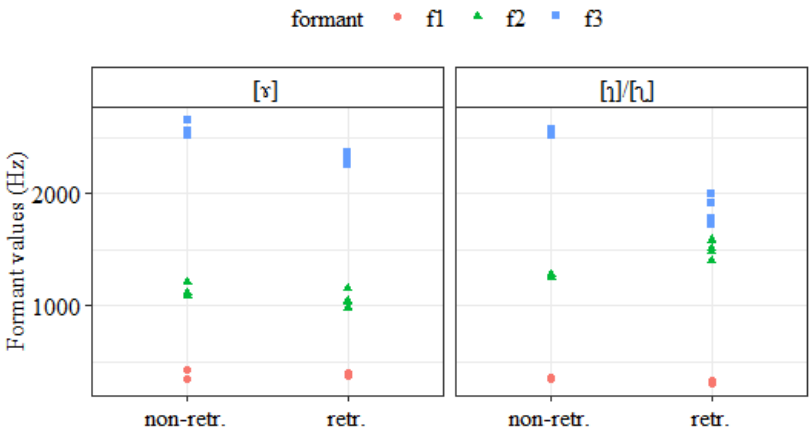


Figure 9: The mean formant frequencies (Hz) and their ratios in the test items.

Table 3: The mean formant frequencies (Hz) and their ratios in the test items.

	[ɤ], non-retroflex	**[ɤ], retroflex**	**[ɿ]**	**[ʅ]**
F1 (Hz)	403	384	350	321
F2 (Hz)	1120	1048	1259	1495
F3 (Hz)	2564	2315	2547	1852
F1/F2 ratio	0.34	0.34	0.28	0.21
F2/F3 ratio	0.44	0.45	0.49	0.81

4. Discussion

4.1. Effects of native language and markedness on the identification of Chinese vowels

The perception test's results showed that the accuracy rates of the mid vowel were lower in all listener group than that of the high vowels, which suggests that <e> [ɤ] is more difficult than <i> [ɿ]/[ʅ] for Hungarian learners of Chinese. These results are in correspondence with Juhász's (2020) results in the production domain. We assume that besides the effect of the difference in the vowel system of the target and the mother language – as suggested by CAH

and MDH –, there must be further factors influencing the acquisition of these segments. Orthographic interference may be raised as a possible reason.

[ʏ] is more difficult than [ɿ] and [ʅ] for Hungarian listeners, and [ʏ] is often perceived as [ɿ]/[ʅ]. The perception result corresponds with the production result of Juhász (2020): the formant values of [ʏ] are significantly different from native Chinese, and Hungarian learners produced the [ʏ] with higher F2 values than Chinese native speakers. There may be a connection between the perception and the production of speech sounds; however, perception and production may also develop mutually and synchronously. This is stated in the L2-learning models (Flege & Bohn, 2021).

Hungarian students in the present study confounded [ʏ] and [ʅ], which was found for Korean students (Wang, 2001) as well. In terms of sound value, [ʏ] in Chinese and [u] in Japanese are similar. However, Japanese students were more likely to perceive [ʏ] as [a]. [ʏ] in Chinese and [ø]/[øː] in Hungarian are similar, but the production result only partly suggested that Hungarian students' production is close to [ø]/[øː]. But the affirmed result is that [ʏ] is realized with higher F2 by Hungarian speakers (Juhász, 2020) and [ʏ] is proven to be confounded with [ʅ]/[ɿ] by Hungarian students in the present study. Compared to the Korean and Hungarian students, Japanese students did not perceive [ʅ] as [ʏ] at all in the selection of "[ʏ]/[ʅ] (maybe the students regard [ʅ] as [i])". Therefore, it seems that predicting and explaining the perceived difficulties of [ʏ]/[ʅ] based solely on acoustic patterns between the native language and the target language is not enough.

4.2. Effects of linguistic experience on the identification of Chinese vowels

The advanced group did not have better results than the beginner groups in our mid-high differentiation task, which is in agreement with Juhász's (2020) production test results.

As discussed in 1.3, orthographic input plays an important role in second language acquisition. More experienced, i.e. advanced learners of Chinese spend more time receiving more orthographic input, while the amount of focused pro-

nunciation (and thus perception) training decreases. Generally, Chinese pronunciation practice is in the first semester, the proper production of the Chinese segments is more highlighted and emphasized in the first year of studying, i.e., Chinese teachers are more careful and articulate more effectively to differentiate these segments. However, as time goes by, everyday communication does not require this efficient distinction between these speech sounds, thus if the mental discrimination of the categories is not established in the beginning in the L2 learner's mind, it is likely that advanced learners will face difficulties when trying to tell them apart – because "high-quality and well differentiated" input to help them discriminate these sounds becomes more and more absent. In other words: this result suggests that if the correct perceptual discrimination is not founded in the beginning of L2 acquisition, then the lack of distinction in the L2 learner's mind persists and fossilises, and may probably deteriorate as well (but this is just a hypothesis which should be addressed in another analysis). Thus, Chinese learners' mental representations of Chinese phonology may be negatively influenced by more Chinese orthographic input. Hungarian advanced learners of Chinese get more orthographic input than beginners, and Pinyin is more opaque than Hungarian orthography – as mentioned above.

4.3. Effects of consonant context on the identification of Chinese vowels

Based on Figure 7, [ɤ] and [ɿ] are situated closer in terms of their formant values compared to the distance between [ɤ] and [ʅ]. We expect that identifying [ɤ] and [ɿ] after non-retroflex [ts, ts^h, s] could induce a lower correctness rate, compared to [ɤ] and [ʅ] after retroflexes [tʂ, $tʂ^h$, ʂ, ɻ], because their second and third formants are closer in these contexts. From the results of the present study, it is only proved that [ɤ] after non-retroflexes [ts, ts^h, s] induces a lower identifying rate than [ɤ] after retroflex [tʂ, $tʂ^h$, ʂ, ɻ] in groups of Begmix, BegChi and intMix, whereas [tsɿ, $ts^hɿ$, sɿ] does not induce a lower identifying rate than [tʂʅ, $tʂ^hʅ$, ʂʅ, ɻʅ] in groups of Begmix, BegChi and intMix. This may be caused by the inherent difficulty for retroflex perception (Tabain et al., 2020). In the syllables [tʂʅ, $tʂ^hʅ$, ʂʅ, ɻʅ], both consonant contexts and the apical vowel

are retroflex, which is perhaps the major contributor to their difficulty. The magnitude of the acoustic change is less apparent (as compared to a dental C + velar V sequence), which might pose problems since there is no dynamic formant change to be used as an acoustic cue to anchor the vowel. The accuracy rate of the non-retroflex vowel [ɿ] is higher than that of the retroflex vowel [ʅ], implying that the retroflex vowel [ʅ] might be more difficult than the non-retroflex vowel [ɿ].

4.4. Effects of a Chinese teacher on the identification of Chinese vowels

The present results showed that experience with a native language teacher may have an effect on Hungarian listeners' perception. The accuracy rate was higher in the beginner groups with a native Chinese teacher than the group without one, which also implies that Hungarian teachers of Chinese have an accent and possibly also have difficulties in discriminating these sounds in production. However, based on the results of the phonetic context for vowel perception, it is interesting that BegHun is in reverse to the other groups with a native Chinese teacher, as seen in Figure 6: [ɤ] after non-retroflex [ts, ts^h, s] induces a higher identifying rate than [ɤ] after retroflex [tʂ, $tʂ^h$, ʂ, ʐ] in the BegHun group. In addition, [tsɿ, $ts^hɿ$, sɿ] also induces a higher identifying rate than [tʂʅ, $tʂ^hʅ$, ʂʅ, ʐʅ] in the BegHun group. This result suggests that experience with a native speaker through merely a native Chinese teacher may already have an impact on identification.

5. Conclusion

This study investigated the perception of Mandarin Chinese vowels [ɤ], [ɿ], and [ʅ] by Hungarian learners, drawing on empirical results and theoretical models of L2 phonological acquisition. The data clearly show that [ɤ] is significantly more difficult for learners to identify than the apical vowels [ɿ]/[ʅ], a finding aligned with previous production studies and perception-based research.

While Contrastive Analysis Hypothesis (CAH) and Markedness Differential Hypothesis (MDH) offer general predictions about difficulty in learning unfa-

miliar and marked L2 sounds, they fall short of explaining the present findings. All three vowels in question are both absent from Hungarian and relatively marked in terms of phonetic rarity, yet learners showed a clear asymmetry in performance. This suggests that markedness alone cannot predict perceptual difficulty in this context.

Another finding was that the accuracy rate in the advanced, intMix group was not proven to be significantly higher than in the beginner groups, moreover, in some aspects it showed a tendency of lower values. Our interpretation – as explained in the Discussion – is that their phonetic memory is not kept awake by focusing on the phonetic-phonemic level only during the first semester; if accurate perceptual discrimination is not established at the onset of L2 acquisition, the inability to distinguish sounds may persist, become fossilised, and potentially deteriorate further over time. Tusor (2016) states that Hungarian learners of Chinese always rely on Pinyin transcription when they are studying Chinese characters and pronunciation. The students who are not made aware that the sounds written in Pinyin are not the same as the sounds represented by the letters of the English and Hungarian alphabets will certainly tend to pronounce these speech sounds incorrectly. The incorrect pronunciation may persist and become fossilised even after abandoning Pinyin.

Furthermore, the study confirms that input quality, as predicted by SLM-r (Flege & Bohn, 2021), is a key determinant in L2 perception: learners taught by native Chinese instructors had significantly higher identification accuracy, particularly for the problematic [ʏ]. This underlines the pedagogical importance of early exposure to native input, especially for phonologically subtle or unfamiliar segments.

In summary, these findings highlight the importance of early, high-quality phonetic training, ongoing perceptual practice, and a critical approach to orthographic input in teaching Chinese to Hungarian learners. Future work should further explore how these variables interact longitudinally and whether targeted interventions can mitigate fossilisation and orthography-driven misperception.

References

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske, & M. YoungScholten (Eds.), *Input Matters in SLA* (p. 191–206). Clevedon, UK: Multilingual Matters.

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

Best, C. (1995). A direct realist view of cross-language speech perception. speech perception and linguistic experience. In Strange (Ed.), *Speech perception and linguistic experience: theoretical and methodological issues* (p. 171–204). Timonium: New York Press.

Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*, 775–794.

Boersma, P. W., & D. (2022). Praat: doing phonetics by computer [Computer program].

Chao, Y. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philosophy, Academia Sinica*, *4*, 363–397.

Chen, Y., Zhang, J., Sieg, J., & Chen, Y. (2019). Is [ɤ] in mandarin a transitional vowel? — evidence from tongue movement by ultrasound imaging. *Journal of Chinese Linguistics*, *47*, 371–405.

Cheng, C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Cheng, R. (1966). Mandarin phonological structure. *Journal of Linguistics*, *2*, 135–158.

Christoph, D., & Klaus, J. (2004). Speech recorder - a universal platform independent multi-channel audio recording software. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04*. Lisbon, Portugal: European Language Resources Association (ELRA.

Duanmu, S. (2000). *The phonology of standard Chinese*. New York: Oxford University Press.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. (2nd ed.). Oxford: Oxford University Press.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*, 315–330.

Erdener, V., & Burnham, D. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*, 191–228.

Fan, L. (2024). Difficulties of chinese vowel finals: A study on hungarian learners and teachers. In P. Sz. Simon, & L. A (Eds.), *15th International Conference of J. Selye University. Language and Literacy Section. Conference Proceedings* (p. 23–41). Komárno, Slovakia.

Fan, L. (submitted). Perception and production of chinese vowel finals by hungarian learners – some relevant difficulties.

Flege, J. (1995). Second language speech learning: Theory, findings, and problems. speech perception and linguistic experience: Issues in cross-language research.

Flege, J., & Bohn, O. (2021). The revised speech learning model (slm-r. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (p. 3–83). Cambridge: Cambridge University Press.

Fox, J., & Weisberg, S. (2019). An r companion to applied regression.

Gass, S., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course*. New York: Routledge.

Hartman, L. (1944). The segmental phonemes of the peiping dialect. *Language*, *20*, 28–42.

Juhász, K. (2020). A mandarin illabiális veláris magánhangzó [ɤ], illetve az alveoláris [ɹ] és posztalveoláris [ɻ] approximánsok produkciója kínaiul tanuló magyarok körében. *Alkalmazott nyelvtudomány*, *20*.

Kiss, G. (1985). A magyar magánhangzók első két formánsának meghatározása szintetizált hangmintákat felhasználó percepciós kísérlet segítségével. Nyelvtudomány Közlemények,.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Lado, R. (1957). Language across cultures. Ann Arbor:.

Lee, W., & Zee, E. (2001). An acoustical analysis of the vowels in beijing mandarin. In P. Dalsgaard, B. Lindberg, H. Benner, & Z.-H. Tan (Eds.), *7th European Conference on Speech Communication and Technology (EUROSPEECH 2001 Scandinavia* (p. 643–646). Aalborg: International Speech Communication Association.

Lee-Kim, S.-I. (2014). Revisiting mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association*, *44*, 261–282.

Lenth, R. (2021). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*, 5–1.

Liberman, A., Harris, K., Eimas, P., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech*, *4*, 175–195.

Lin, Y.-H. (1989). Autosegmental treatment of segmental processes in chinese phonology.

Lin, Y.-H. (2007). *The Sounds of Chinese with Audio CD* volume 1. Cambridge: Cambridge University Press.

Pulleyblank, E. (1984). *Middle Chinese: A study in historical phonology*. Vancouver, BC: University of British Columbia Press.

R Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Simay, A., Fan, L., & Szemle, K. (2020). A kínai nyelv magyarországi tanításának rövid története és jelene (brief historical overview and present state of chinese studies in hungary.

Strange, W. (1995). *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press.

Szende, T. (1994). Hungarian. *Journal of the International Phonetic Association*, *24*, 91–94.

Tabain, M., Butcher, A., Breen, G., & Beare, R. (2020). A formant study of the alveolar versus retroflex contrast in three central australian languages: Stop, nasal, and lateral manners of articulation. *The Journal of the Acoustical Society of America*, *147*, 2745–2765.

Tusor, N. (2016). A kínaiul tanuló magyar anyanyelvűek tipikus kiejtési hibái.

Wang, Y. (2001). A preliminary investigation on the perception of high vowels in mandarin chinese by korean and japanese students. *Language teaching and linguistic studies*, *6*, 8–17.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wiese, R. (1997). Underspecification and the description of chinese vowels. In J. Wang., & N. Smith (Eds.), *Studies in Chinese phonology* (p. 219–249). Berlin: Mouton de Gruyter.

Xu, S.-R. (1980). *Phonology of Standard Chinese (*普通话语音知识*)*. Beijing: Wenzi Gaige Chubanshe.

Zhang, Y. (2003). The influence of acoustic properties on perception of apical vowels in beijing mandarin. In *In. Proceedings of the Sixth National Conference on Modern Phonetics* (p. 109–114).

Zhou, D., & Wu, J. (1963). *Putonghua fayin tupu [Articulatory diagrams of Standard Chinese*. Beijing: Shangwu yinshuguan.

Zhu, X. (2010). *Phonetics (*语音学*)*. Beijing: The Commercial Press.

Beszéddallam reprodukciója kottakép alapján

Gocsál Ákos[1,2], Partos Dorka[1]

[1] *Pécsi Tudományegyegyem Művészeti Kar*
[2] *ELTE Nyelvtudományi Kutatóközpont*

Abstract

In 1967, Iván Fónagy and Klára Magdics published figures of hundreds of intonation contours, using musical notation, reflecting a wide range of emotional states and attitudes. The purpose of this study was to examine whether they can be reproduced by means of speech manipulation software. The intonation curves of ten sentences, selected from the book, were reproduced in Praat, after musical notes were converted to fundamental frequency values. A pilot study with five sentences (reflecting apathy, joy, sadness, fright, and mock) was conducted to test whether listeners can decode the emotions and attitudes expressed. Thirty-five students (aged 18-27 years), all native speakers of standard Hungarian, indicated on 5-point scales how intensely they felt the sentences they had heard expressed the emotions (1 = not at all, 5 = very strongly). Each sentence was evaluated by all five emotions, and an extra emotion not included in the test sentences (surprise). In general, the random distribution of scores suggested that listeners were unable to identify the intended emotions, but the sentences were perceived as not conveying other qualities. It is concluded that prosodic features other than intonation are also necessary for recognizing emotions.

1. Bevezetés

Jelen műhelytanulmányunkban egy olyan módszert ismertetünk, amelynek alkalmazásával kottaképeken megjelenített beszéddallamot a Praat program segítségével természetes bemondásra illesztünk.

A beszéddallam képi megjelenítése iránti igény nem új a beszédkutatás szakirodalmában. Sir Joshua Steele már 1775-ben kidolgozott egy, a zenei notációhoz hasonló jelrendszert a dallam és a prozódia más elemeinek lejegyzéséhez (idézi Murdoch, 1883), ám ez nem terjedt el. Viëtor (1897:103) basszuskulcsban kottaként ábrázolta a német *du* szó különféle érzelmekkel ejtett változatait, Scripture (1906) pedig már az alaphang kontúrját ábrázoló „beszédgörbéket”

Email addresses: gocsal.akos@pte.hu (Gocsál Ákos), pdorka2002@gmail.com (Partos Dorka)

Beszédtudomány – Speech Science 2025/2. 112–131. DOI: 10.15775/Besztud.2025.2.112-131

közölt, amelyek megrajzolásához oszcillogramokon (kimogramokon) mért hullámhosszok alapján számolt frekvenciaadatokat. Jones (1909) a kétféle ábrázolásmódot egyesítette úgy, hogy a lehallgatott beszéd dallamgörbéit az ötvonalas rendszerben, basszuskulcsban vagy violinkulcsban ábrázolta. Ő azonban nem kimogramokat használt, a lejegyzés során a beszéddallam frekvenciáit hangvilla hangjához viszonyítva határozta meg.

A beszédkutatás hazai szakirodalmában már Regner (1861), Gombocz (1907) és Tolnai (1915) is említi a beszéddallam jelentőségét – Gombocz francia nyelvű mondat kottaképét is közli –, a magyar nyelvre vonatkozóan azonban csak Csűry (1925) munkájában találunk először képi ábrázolást. Csűry ötvonalas, kottaszerű, de nem zenei lejegyzést alkalmazott. Néhány évvel később pedig Hegedűs (1930) Scripture-höz hasonlóan dallamgörbéket rajzolt, azonban ezek elemzésekor egy-egy kiemelt frekvenciaértékhez zenei hangokat is közölt.

A dallamok lejegyzésében jelentős előrelépés történt az 1960-as években. Magdics (1963) részletes elemzést nyújtott a beszéd és a zene összehasonlításáról, és altkulcsban több mondat kottaképét is közölte. Arra a következtetésre jutott, hogy bár a dallam, a ritmus, a hangközök stb. sok szempontból különböznek, mégis e tényezők alapján összehasonlítások végezhetők. Magdicsnak e munkája nyújt elméleti alapot Fónagy és Magdics (1967) nagyszabású munkájához, amely sok száz kottapéldán keresztül mutatja be magyar beszéd dallamformáit. Ezeknek a dallamformáknak a lejegyzése Magdics (1963) példáihoz hasonlóan szintén altkulcsban történt. Munkájukat élesen bírálta Boros (1969) a zenei kontroll hiánya és számos, a lejegyzés során adódott pontatlanság miatt. Bartók (1969) ugyan hitelesnek tartotta a közölt dallamíveket, azonban felvetette, hogy ezek a dallamok nem feltétlenül tükrözik jól a köznyelvi beszédet, mivel a mondatokat színészek bemondásában rögzítették.

Boros a későbbiekben zenei kották formájában közölt beszéddallamokat (Boros, 1971, 1975), míg Bartók (1974) nyolcvonalas rendszert használt, hozzátéve, hogy ez a vonalrendszer semmi esetre sem tekinthető „zeneinek”. Bartók radikálisan eltérő véleményt fogalmazott meg Fónagy és Magdics munkájával szemben, ugyanis kijelentette, hogy bár a beszédben szubjektív módon meghatározott

hangközök hasonlíthatnak a zenei hangközökre, azokkal sohasem lehetnek azonosak, mert a beszédben zenei hangközök nem léteznek.

A mai beszédelemző szoftverek alkalmazásával könnyen megjeleníthetők a dallamgörbék, így jól támogatható az intonációs rendszerek fonológiai, azaz diszkrét kategóriákkal, szimbólumokkal való leírása (Hirst & Cristo, 1998:16), de a görbék megrajzolásával könnyen tanulmányozhatók a prototipikustól (Markó, 2017), eltérő mintázatok is (Markó, 2013; Tóth, 2016). Napjainkra a beszéddallam zenei megalapozottságú, kottaszerű ábrázolása már visszaszorult a beszédkutatás szakirodalmában, azonban a dallamok zenei hangközökkel való párhuzamba állítása időről időre felmerül (Bowling et al., 2010; Robledo et al., 2016; Gocsál & Urbanics, 2022), sőt, ismeretesek olyan szerzők, akik határozottan állást foglaltak a beszéd kottával történő leírása mellett (Chow & Brown, 2018). Ők azzal érvelnek, hogy a kotta az egyetlen olyan eszköz, amely a dallamot és a ritmust egységes rendszerben tudja ábrázolni, illetve csak a kotta segítségével tudja egy beszélő egy másik beszélő prozódiáját reprodukálni.

Az itt ismertetendő módszerünknek is ez a kiindulópontja. Célunk az, hogy a Fónagy és Magdics (1967) által közölt kották alapján szoftveres manipuláció alkalmazásával reprodukáljuk a lejegyzett dallamot, majd validálni szeretnénk azokat, azaz azt szeretnénk megállapítani, hogy a hallgatók reprodukált dallam észlelésekor képesek-e visszakódolni az adott dallam által Fónagy és Magdics (1967) szerint kifejezett érzelmet, attitűdöt. Ugyanakkor nem kívánunk elvi állásfoglalást tenni a zenei lejegyzés helyességéről vagy helytelenségéről, hanem az motivál bennünket, hogy megvizsgáljuk, a Fónagy és Magdics által közölt, kottakép formájában rendelkezésre álló több száz dallamforma gyakorlati célokra felhasználható-e.

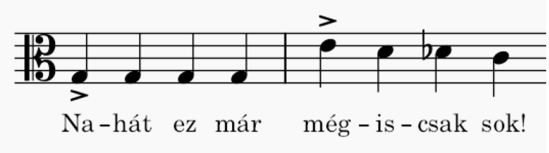
2. Az alkalmazott eljárás

2.1. A mondatok kiválasztása

Az eljárás kidolgozásához kiválasztottunk Fónagy és Magdics (1967) kötetéből tíz olyan kottát, amelyek a szöveges ismertetés alapján jelentősen eltérő

érzelmet, attitűdöt, szándékot fejeznek ki. Ezek az alábbiak: 1. felháborodás előkészítő résszel (583. kotta, p. 239.), 2. közöny (364. kotta, p. 192.), 3. öröm (367. kotta, p. 193.), 4. szomorúság (372. kotta, p. 194.), 5. rémület (432. kotta, p. 205.), 6. nyugtatás (445. kotta, p. 208.), 7. gúny (502. kotta, p. 221.), 8. fojtott hangú fenyegetés (550. kotta, p. 232), 9. sértődöttség, duzzogás (612. kotta, p. 245.), 10. elgondolkodás, töprengés (624. kotta, p. 248.). A mondatok kiválasztása önkényesen történt, azonban törekedtünk arra, hogy minél változatosabb érzelmeket és attitűdöket tükrözzenek, és szerepeljenek közöttük alapérzelmek is (öröm, szomorúság).

Az alábbiakban a fenti 1. példán keresztül ismertetjük az eljárást. Első lépésként a MuseScore v3.6.2.5 programban rögzítettük a dallamot (MuseScore, 2021), így az lehallgathatóvá vált (1. ábra).



1. ábra. A *Nahát, ez már mégiscsak sok!* mondat kottaképe (Fónagy & Magdics, 1967:239 alapján).

A következő lépésben táblázatos formában rögzítettük a kottáról leolvasható zenei hangokat, illetve a hozzájuk tartozó frekvenciaértékeket egészre kerekítve (1. táblázat).

1. táblázat. A *Nahát, ez már mégiscsak sok!* mondat zenei hangjai.

szótag	*Na*	*hát*	*ez*	*már*	*még*	*is*	*csak*	*sok*
zenei hang	G	G	G	G	E’	D’	D♭’	C’
frekvencia (Hz)	196	196	196	196	330	294	277	262

A frekvenciaértékek tanulmányozása során – összevetve azokat Gráczi et al. (2020) által közölt, mai alapfrekvencia-adatokkal – azt valószínűsítettük, hogy a bemondás női beszélőtől származik. A Fónagy és Magdics (1967:314) által adott rövid módszertani leírásból azonban kiderül, hogy az egyes dallamformák

lehallgatását, „kiéneklését", majd első kottázását követően a lejegyző mindegyik dallamot közös „alaphangra" transzponálta, emiatt a frekvenciaadatok nem adnak támpontot a beszélő nemének legalább bizonyos valószínűséggel történő megítéléséhez. Ez az említett „alaphang" a szerzők közlése szerint a beszélő által a legkisebb erőfeszítéssel képezhető hang volt, amelyet a végső lejegyzéskor minden beszélő esetében az altkulcson a legalsó vonalra (F hang) helyeztek. Ez a transzponálási mozzanat kétségkívül összehasonlíthatóvá teszi a különböző beszélők által produkált dallamokat, azonban – egyetértve Bartók (1969) korabeli kritikai észrevételével, és kiegészítve azt – az „alaphang" itt csak egy technikai fogalom, és semmi esetre sem tekinthető annak az átlagos alapfrekvenciának, amelyet egyébként ilyen esetekben mérni szokás. Ezt alátámasztja az, hogy a kották túlnyomó többségében a jelzett F hang alatt már nem jelentek meg hangok, az átlagos alapfrekvencia viszont éppen attól „átlagos", mert egy szélesebb terjedelem középértéke, tehát törvényszerű, hogy alatta is előforduljanak hangok. Bartók (1969) úgy vélte, hogy a szerzők a vizsgált beszédszakaszok legmélyebb hangját tekintették „alaphangnak", azonban ez sem pontos megállapítás. A kötet számtalan olyan kottát közöl, amelyikben egyetlen F hang sincs, ugyanakkor előfordul F alatti hang is (pl. 714/b. kotta, p. 294.).

Mindez tehát kérdéseket vet fel azzal kapcsolatban, hogy a kottákon megjelenített beszéddallamok reprodukciójakor milyen alapfrekvenciát állítson be a kutató, illetve a beszélő létező, „igazi" alapfrekvenciájához a kottán ábrázolt hangokat hogyan viszonyítsa. Jelen munkánkban erre két megoldást mutatunk be.

2.2. A mondatok rögzítése

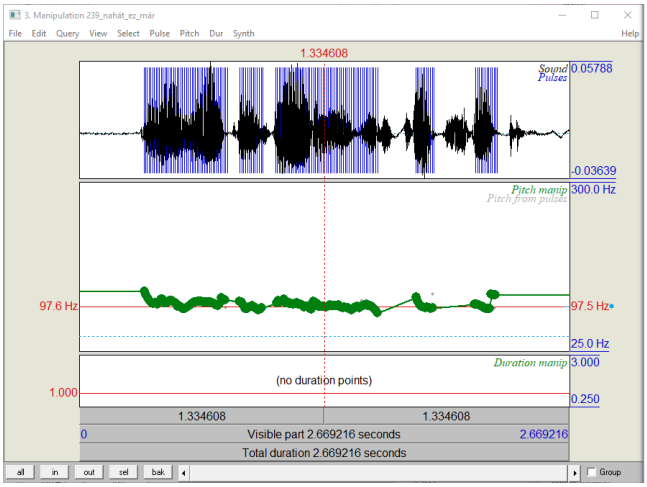
A kiválasztott mondatokat egy magyar anyanyelvű, 53 éves férfi beszélő bemondásában rögzítettük. A beszélő a sztenderdnek tekinthető köznyelvi nyelvváltozatot beszéli, nem dohányzik, beszédhibája, hangképzési rendellenessége, beszédszervi betegsége nincs.

A bemondások rögzítése TASCAM DR-44WL típusú digitális diktafonnal, .wav formátumban (16 bites felbontásban, 44 kHz-es mintavételezéssel), csendes

szobában, zavaró körülmények nélkül történt. A beszélő a kiválasztott mondatokat semleges érzelmi állapotban, dallamvariáció nélkül, egyenletes alapfrekvenciával mondta el.

2.3. A dallamgörbe vizsgálata

A fenti eljárással rögzített *Nahát, ez már mégiscsak sok*! (239. kotta) mondatból a Praat 6.0.52. programban (Boersma & Weenik, 2019) létrehoztunk egy manipulációs objektumot. Minden beállítást a Praat által felkínált alapértéken tartottunk (2. ábra).

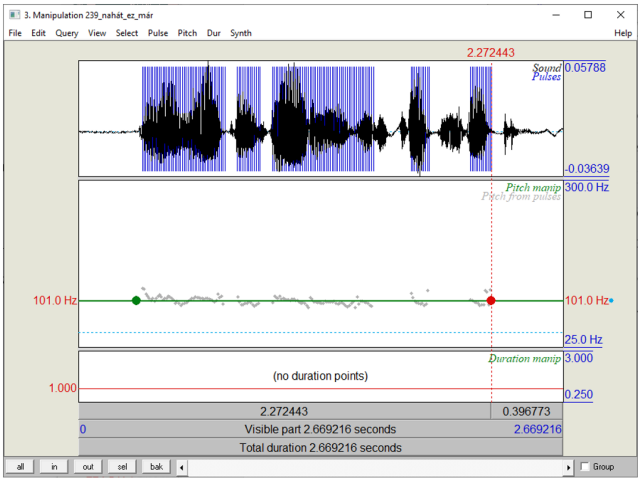


2. ábra. A *Nahát, ez már mégiscsak sok!* mondat a manipulációs ablakban.

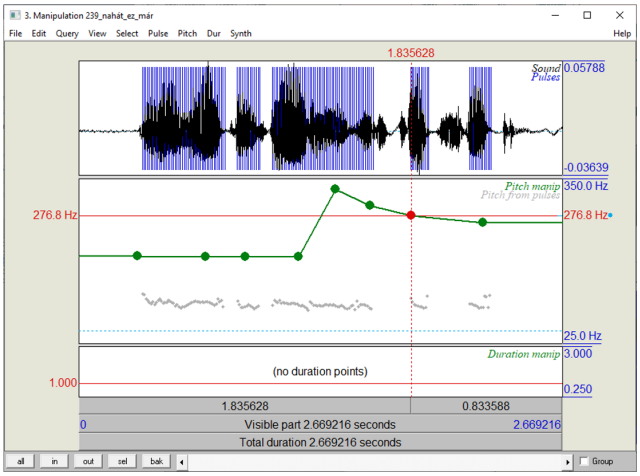
Látható, hogy a beszélő, bár igyekezett dallamvariáció nélkül beszélni, ez nem sikerült teljesen, kisebb ingadozások továbbra is vannak a dallamban. Első lépésként ezeket az ingadozásokat küszöböltük ki, megfelelő beállításokkal monotonná alakítottuk az összes dallamot, majd az alapfrekvenciát a vizsgált mondat alapfrekvenciájának mediánjára állítottuk (100 Hz, 3. ábra).

Ezt követően először az 1. táblázat szerinti értékekre állítottunk be az egyes szótagokon az alapfrekvenciát (4. ábra).

Az így kapott dallamgörbe azonban túl magas, természetellenes hangzású volt, így a férfi beszélő számára természetes hangfekvésre transzponáltuk. A



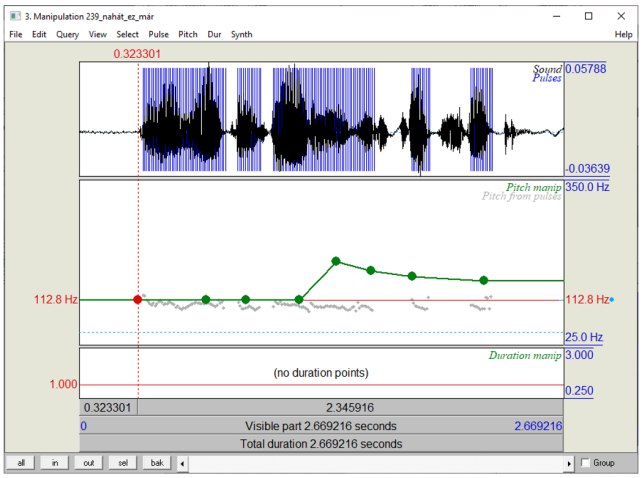
3. ábra. A *Nahát, ez már mégiscsak sok!* mondat egyenletes alaphangon.



4. ábra. A *Nahát, ez már mégiscsak sok!* mondat dallamgörbéje az 1. táblázatban közölt frekvenciaértékek alkalmazásával.

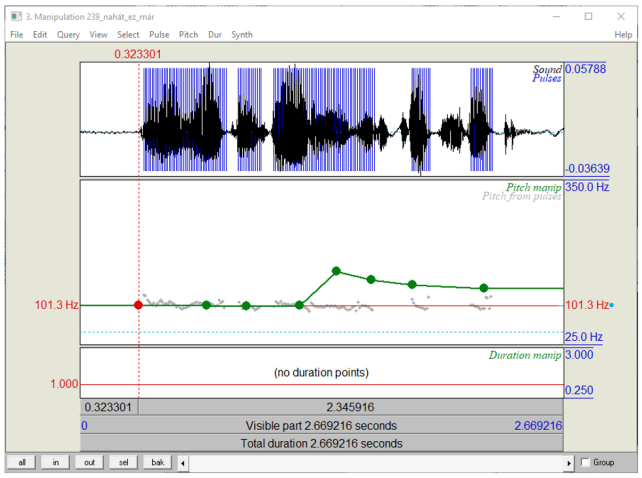
transzponálást két változatban végeztük el. Az első változatnál a szerzők által korábban említett F „alaphangot” vettük alapul (176 Hz), és ezt szállítottuk le a bemondáson mért 101 Hz-es mediánértékre (5. ábra).

A második változatnál a mondat elején, több szótagon is egyenletesen megjelenő G hangot vettük alapul (196 Hz), és ezt szállítottuk le a bemondáson mért



5. ábra. A *Nahát, ez már mégiscsak sok!* mondat első transzponált változata.

101 Hz-es mediánértékre, azaz 11,48 félhanggal transzponáltuk a bemondást lefelé (6. ábra).



6. ábra. A *Nahát, ez már mégiscsak sok!* mondat második transzponált változata.

Valószínűsíthető, hogy mindkét megoldás működőképes, a hallgatók el tudják fogadni. Az első esetben azonban a magasabb hangfekvés nagyobb feszültségi állapotot tükrözhet, a második, mélyebben megvalósuló dallam nyugodtabb beszélő benyomását kelti.

2.4. A módszer tesztelése

A következőkben pilot kutatást végeztünk annak megállapítására, hogy a fenti módszer alkalmazásával rekonstruált beszéddallamok valóban közvetítik-e – Fónagy és Magdics szóhasználatát megtartva – az adott érzelmeket, attitűdöket a hallgatók számára.

A 2.1. alfejezetben említett mondatok közül ötöt választottunk ki a pilot kutatáshoz, ezek dallama közönyt, örömöt, szomorúságot, rémületet és gúnyt fejez ki. Két esetben módosítottunk a Fónagy és Magdics által közölt mondatokon azzal a céllal, hogy a hallgatókat a mondatokban előforduló szavak, kifejezések ne befolyásolják a kifejezett érzelem megítélésében. Így az eredeti, közönyt kifejező *Nem érdemes ezzel foglalkozni.* mondat helyett a *Ma is fogunk ezzel foglalkozni.*, illetve a szomorúságot tükröző *Kidobtak az állásomból.* helyett *Beszéltek az állásomról.* mondatokat használtuk. A pilot kutatáshoz használt mondatokat és a kottaképekből visszafejtett frekvenciaadatokat a 2. táblázat tartalmazza.

A frekvenciaértékek beállítását követően minden mondatot egységesen – a fenti második megoldás alapján – 11,48 félhanggal lefelé transzponátunk, így a kottaképeknek megfelelően a dallam és a hangfekvés is megmaradt.

A kutatáshoz tesztlapot készítettünk, amely nem csak a fenti öt érzelmet, attitűdöt tartalmazta, hanem egy hatodikat, a meglepetést is. A kutatás résztvevői ötfokú skálán jelölték, hogy az adott érzelem, attitűd milyen mértékben tükröződik az elhangzott mondatban (1 = semennyire, 5 = nagyon erősen). A résztvevők mindegyik mondatot értékelték mind a hat érzelemre vonatkozó skálán.

A kutatásban 35 nappali tagozatos egyetemi hallgató vett részt (életkoruk 18–27 év), mindannyian magyar anyanyelvűek, hallószervi betegségről, hallássérülésről nem számoltak be. A lehallgatás a Pécsi Tudományegyetem Művészeti Karának egyik tantermében, zavaró körülmények nélkül, több kisebb csoportban, számítógépről, jó minőségű, Xiaomi típusú Bluetooth hangszórón lejátszva történt. A kísérletvezető a jelen pilot kutatásban nem szereplő, a 2. fejezetben ismertetett mondat lejátszásával bemutatta a feladatot, ezzel együtt

2. táblázat. A pilot kutatáshoz használt mondatok és frekvenciaadataik.

1. közöny	*ma*	*is*	*fo*	*gunk*	*ve*	*le*	*fog*	*lal*	*koz*	*ni*
zenei hang	D'	H	H	H	A	A♭	H	A	G	G
frekvencia (Hz)	294	246	246	246	220	233	246	220	196	196
2. öröm	*kép*	*zeld*	*még*	*sem*	*ő*	*vizs*	*gáz*	*tat*		
zenei hang	349	294	349	246	262	277	262			
frekvencia (Hz)	F'	D'	F'	H	C'	C#'	C'			
3. szomorúság	*be*	*szél*	*tek*	*az*	*ál*	*lá*	*som*	*ról*		
zenei hang	A	F	F	F	G	F	F♭	F		
frekvencia (Hz)	220	175	175	175	196	175	164	175		
4. rémület	*be*	*szél*	*jen*	*sán*	*dor*	*mi*	*tör*	*tént*		
zenei hang	C'	C'	C'	C'	C'	D'	C'	C'		
frekvencia (Hz)	262	262	262	262	262	294	262	262		
5. gúny	*mi*	*cso*	*da*	*ér*	*ző*	*szív*				
zenei hang	G	F	F	G	F	G				
frekvencia (Hz)	196	175	175	196	175	196				

megbizonyosodott arról, hogy a résztvevők mindegyike számára egyértelmű a feladat és a tisztán hallják a lejátszott mondatokat. Egy hallgató a tesztlap megtekintésekor jelezte, hogy nem ismeri a *közöny* szót, számára a kutatás vezetője példákkal illusztrálva elmagyarázta a szó jelentését.

A tesztmondatok lejátszása minden csoportban azonos körülmények között történt. Először a hallgatók a közönyt, majd – egy, a kutatók által előzetesen megállapított, véletlenszerű sorrend szerint – a többi érzelmet, attitűdöt kifejező mondatot hallgatták meg (öröm, szomorúság, rémület, gúny) és értékelték a fenti módon. Minden mondat lejátszása egymás után kétszer történt, az első lehallgatás után a hallgatók rögzítették az adott mondattal kapcsolatos benyomásaikat, ezt követően még egyszer meghallgathatták ugyanazt a mondatot, korrigálhatták válaszukat.

A számításokat az *R* szoftverben (4.4.1 verzió, R Core Team, 2024) végeztük el a χ^2 (goodness-of-fit) próba alkalmazásával. Minden esetben nullhipotézis-

ként azt feltételeztük, hogy az egyes skálaértékre azonos arányban érkeznek válaszok. Ahol ettől szignifikáns eltérés mutatkozott, a kiugró cellaértékeket a standardizált reziduumok (SR) vizsgálatával kerestük meg. Kiugrónak tekintettük az abszolút értékben 1,96-nál nagyobb SR-értékeket. Egyes esetekben nem szignifikáns χ^2 ellenére is határértéken túli SR-értékek adódtak, ekkor a nem szignifikáns χ^2 értéket tekintettük mérvadónak. A hatás nagyságát a Cramér-féle V segítségével állapítottuk meg (*rcompanion* csomag: Mangiafico, 2024), a kapott V-értékek értelmezését Mangiafico (2016) alapján végeztük.

3. Eredmények

Mindegyik mondat esetében megvizsgáltuk, hogy hány fő választotta az egyes skálaértékeket a különböző érzelmeknél. Előfordult, hogy egy vagy több skálaértéket egyetlen hallgató sem választotta, így a nullhipotézisnek megfelelően a következő elméletileg várt cellaértékekkel dolgoztunk: öt kategória esetében 7, négy kategóriánál 8,75, három kategóriánál 11,67.

3.1. A közönyt kifejező dallammal megvalósított mondat

A skálaértékek eloszlását és a számítások eredményét a 3. táblázat tartalmazza. Itt, illetve a 4., 5., 6. és 7. táblázatokban félkövér kiemelés jelzi azokat az észlelt érzelmeket, amelyek esetében a skálaértékek eloszlása a nullhipotézisnek megfelelően nem különbözik szignifikánsan az egyenletes eloszlástól.

Az észlelt közöny esetében kapott nem szignifikáns p-érték azt mutatja, hogy kiegyenlített azon válaszadók aránya, akik szerint a lejátszott mondat kifejezi a közönyt, illetve akik szerint nem. Az összes többi észlelt érzelem esetében a skálaértékek arányos megoszlását feltételező nullhipotézis elvethető. Az adatok eloszlásából világosan látszik, hogy a válaszadók ezekben az esetekben jellemzően a legalacsonyabb skálaértéket jelölték meg. Ahol szignifikáns eltérés adódott, a V értékek erős összefüggést jeleznek, továbbá az SR-értékek megerősítik az 1-es skálaérték kiugróan gyakori ($> 1{,}96$), illetve a magasabb skálaértékek elvártnál szignifikánsan ritkább ($< -1{,}96$) előfordulását.

3. táblázat. Közöny.

észlelt érzelem	**1**	**2**	**3**	**4**	**5**	χ^2	***df***	***p***	***V***
közöny	**2**	**8**	**8**	**11**	**6**	**6,285**	**4**	**0,178**	**0,212**
SR	−2,11	0,42	0,42	1,69	−0,42				
gúny	27	3	3	2	0	50,829	3	< 0,001	0,695
SR	7,12	−2,24	−2,24	−2,63					
rémület	23	5	4	2	1	47,143	4	< 0,001	0,580
SR	6,76	−0,84	−1, 26	−2,11	−2,53				
szomorúság	13	7	10	4	1	12,857	4	< 0,05	0,303
SR	2, 53	0	1,26	1,26	−2,53				
öröm	23	7	4	1	0	33	3	< 0,001	0,560
SR	5,56	−0,68	−0,85	−3,02					
meglepetés	28	5	2	0	0	33,686	2	< 0,001	0,703
SR	5,85	−2,39	−3,46						

3.2. Az örömöt kifejező dallammal megvalósított mondat

A 4. táblázat foglalja össze az adatokat és a számítások eredményeit.

4. táblázat. Öröm.

észlelt érzelem	**1**	**2**	**3**	**4**	**5**	χ^2	***df***	***p***	***V***
közöny	19	8	7	1	0	19,286	3	< 0,001	0,428
SR	4,00	−0,29	−0,68	−3,02					
gúny	28	4	2	1	0	57	3	< 0,001	0,736
SR	7,51	−1,85	−2,63	−3,02					
rémület	18	8	1	5	3	25,429	4	< 0,001	0,426
SR	4,64	0,42	−2,53	−0,54	−1,69				
szomorúság	23	3	7	2	0	32,543	3	< 0,05	0,556
SR	5,56	−2,24	−0,68	−2,63					
öröm	**10**	**6**	**5**	**10**	**4**	**4,571**	**4**	0,334	**0,180**
SR	1,26	−0,42	−0,84	1,26	−1, 26				
meglepetés	**3**	**3**	**11**	**10**	**8**	**8,285**	**4**	0,081	**0,243**
SR	−1,69	−1,69	1,69	1,26	0,42				

A nullhipotézis az öröm és a meglepetés észlelésekor igazolódott, azaz a megfigyelt eloszlások ezekben az esetekben nem különböznek szignifikáns mértékben az adatok elvárt, egyenletes eloszlásától. A hallgatók egy része szerint tehát örömöt és meglepetést fejez ki a lejátszott mondat, más részük szerint viszont nem. A meglepetés esetében azonban a p-érték tendenciaszerű kapcsolatra utal, a 3-as érték közel szignifikáns mértékben gyakrabban, az 1-es és 2-es érték viszont

közel szignifikáns mértékben ritkábban fordult elő, mint az elméletileg várt gyakoriság. A többi esetben viszont elvethető a nullhipotézis, azaz az elhangzott mondatot egyértelműen nem közönyösnek, gúnyosnak, rémületet vagy szomorúságot kifejezőnek ítélték a hallgatók. A szignifikáns esetekben az 1-es skálaérték kiugróan gyakori előfordulását igazolták a magas SR-értékek ($> 1{,}96$), a V pedig erős hatásra utal.

3.3. A szomorúságot kifejező dallammal megvalósított mondat

Az 5. táblázat foglalja össze az adatokat és a számítások eredményeit.

5. táblázat. Szomorúság.

észlelt érzelem	**1**	**2**	**3**	**4**	**5**	χ^2	***df***	***p***	***V***
közöny	**11**	**4**	**10**	**3**	**7**	**7,142**	**4**	**0,178**	**0,128**
SR	1,69	−1,26	−1,26	−1,69	**0**				
gúny	23	3	4	4	1	46,571	4	$< 0{,}001$	0,576
SR	6,76	−1,69	−1,26	−1,26	−2,53				
rémület	12	12	8	2	1	16	4	$< 0{,}01$	0,338
SR	2,11	2,11	0,42	−2,11	−2,53				
szomorúság	**8**	**8**	**6**	**11**	**2**	**6,285**	**4**	**0,178**	**0,212**
SR	0,42	0,42	−0,42	1,69	−2,11				
öröm	26	5	3	1	0	46,257	3	$< 0{,}001$	0,663
SR	5,56	−0,68	−0,85	−3,02					
meglepetés	21	5	5	2	2	36,286	4	$< 0{,}001$	0,509
SR	5,91	−0,84	−0,84	−2,11	−2,11				

A nullhipotézis a közöny és a szomorúság esetében tartható meg, azaz a hallgatók egy része szerint a szándékunk szerint szomorúságot kifejező dallammal kimondott mondat kifejezi a közönyt és a szomorúságot, más részük szerint viszont nem, és arányuk kiegyenlített. A szignifikáns p-értékek arra utalnak, hogy a hallgatók szerint a mondat egyértelműen nem fejez ki gúnyt, rémületet, örömöt, meglepetést. Ezekben az esetekben az SR-értékek az 1-es skálaérték kiemelkedően magas, míg a magasabb skálaértékek elvártnál szignifikánsan ritkább előfordulását igazolták. A V a szignifikáns esetekben erős összefüggést mutat.

3.4. A rémületet kifejező dallammal megvalósított mondat

A 6. táblázat foglalja össze kapott eredményeket.

6. táblázat. Rémület.

észlelt érzelem	**1**	**2**	**3**	**4**	**5**	χ^2	***df***	***p***	***V***
közöny	**4**	**3**	**8**	**7**	**13**	**8,857**	**4**	**0,064**	**0,251**
SR	−1,26	−1,69	0,42	**0**	2,53				
gúny	**11**	**5**	**8**	**10**	**1**	**9,42**	**4**	**0,051**	**0,259**
SR	1,69	−0,84	0,42	1,26	−2,53				
rémület	24	4	4	3	0	35,514	4	$< 0{,}01$	0,581
SR	5,95	−1,85	−1,85	−2,24					
szomorúság	23	9	3	0	0	18,057	2	$< 0{,}001$	0,508
SR	4,06	−0,95	−3,10						
öröm	30	2	2	1	0	68,886	3	$< 0{,}001$	0,81
SR	8,29	−2,63	−2,63	−3,02					
meglepetés	23	8	1	3	0	33,914	3	$< 0{,}001$	0,568
SR	5,56	−0,29	−3,02	−2,24					

A rémületet kifejező mondat esetében két nem szignifikáns, de tendenciaszerű összefüggés adódott. A hallgatók tendenciaszerűen inkább közönyösnek érezték a mondatot, a gúny esetében viszont az 5-ös skálaértéktől eltérő vélemények voltak tendenciaszerűen túlsúlyban. A hallgatók szerint a mondat egyértelműen nem fejez ki rémületet, szomorúságot, örömöt, meglepetést. A szignifikáns esetekben a V erős összefüggést mutat.

3.5. A gúnyt kifejező dallammal megvalósított mondat

A 7. táblázat foglalja össze kapott eredményeket.

Az adatok alapján az állapítható meg, hogy a hallgatók nem szignifikánsan, de tendenciaszerűen gúnyosnak ítélték a szándékunk szerint gúnnyal ejtett mondatot. Az összes többi esetben szignifikánsan elutasítóak voltak, azaz a mondat nem fejezett ki közönyt, rémületet, szomorúságot, örömöt, meglepetést a kísérleti személyek számára. A V-értékek erős összefüggést mutatnak.

3.6. Az eredmények összefoglalása

Általános megállapításként azt fogalmazhatjuk meg, hogy a feltételezett érzelmek – bár felismerésük nem volt egyértelmű – általában nagyobb mértékben

7. táblázat. Gúny.

észlelt érzelem	**1**	**2**	**3**	**4**	**5**	χ^2	***df***	***p***	***V***
közöny	10	12	9	2	2	12,571	4	$< 0{,}05$	0,299
SR	1,26	2,11	0,84	−2,11	−2,11				
gúny	**7**	**4**	**8**	**3**	**13**	**8,857**	**4**	**0,064**	**0,251**
SR	**0**	−1,26	0,42	−1,69	2,53				
rémület	31	2	1	0	0	51,235	2	$< 0{,}001$	0,868
SR	7,15	−3,39	−3,75						
szomorúság	23	4	4	4	0	30,943	3	$< 0{,}001$	0,542
SR	5,56	−1,85	−1,85	−1,85					
öröm	22	3	4	5	1	41,429	4	$< 0{,}001$	0,544
SR	6,33	−1,69	−1,26	−0,84	−2,53				
meglepetés	12	8	10	2	3	10,857	4	$< 0{,}05$	0,278
SR	2,11	0,43	1,26	−2,11	−1,69				

tükröződtek a lejátszott mondatokban, mint a többi öt érzelem vagy attitűd többsége. A gúny esetében volt a felismerés a legpontosabb, itt közel szignifikáns eredményt adott a számítás. A közönyös, az örömteli és a szomorú dallammal ejtett mondatok esetében a nullhipotézis igazolódott, azaz a skálaértékek eloszlása nem különbözött szignifikánsan a véletlen eloszlástól. Amikor azonban a hallgatók arról nyilatkoztak, hogy ugyanezek a mondatok milyen mértékben tükröznek más érzelmeket, a legtöbb esetben alacsony skálaértékek adódtak, azaz a hallgatók elutasították, hogy ezeket a további érzelmeket az adott dallammal megvalósított mondat kifejezné. De hasonló, elutasító eredményt kaptunk a – szándékunk szerint – rémületet kifejező mondat érzelmi töltetének visszakódolásakor is: a válaszokból egyértelműen az derül ki, hogy a mondat a hallgatók szerint nem tükröz rémületet.

4. Következtetések

Műhelytanulmányunkban annak a lehetőségét vizsgáltuk, hogy a Fónagy Iván és Magdics Klára által 1967-ben közölt kottaképek alapján szoftveres úton rekonstruálhatók-e beszéddallamok. A bemutatott módszerrel ez technikailag megvalósítható, azonban kérdéses, hogy a hallgatók megbízhatóan vissza tudják-e kódolni a dallamok által, az említett szerzők szerint kifejezett érzelmeket, atti-

tűdöket. Pilot kutatásunk eredményei azt mutatják, hogy önmagában a dallam rekonstrukciója nem ad elegendő támpontot ahhoz, hogy a hallgató egyértelműen azonosítsa a meghatározott érzelmet, ahhoz azonban általában igen, hogy milyen érzelmet, attitűdöt biztosan *nem* fejez ki. Az érzelmek megfelelő visszakódolásánál megfigyelt bizonytalanság arra utalhat, hogy az érzelmet tükröző dallam bár lényeges, de nem egyetlen prozódiai eleme az érzelmes beszédnek, és vélhetően önmagában még nem elegendő az érzelmek pontos észleléséhez. Emellett bizonyos esetekben tévesztéseket is megfigyeltünk. Az örömöt kifejező mondat tendenciaszerűen inkább meglepetést tükrözött a hallgatóknak. A rémületet kifejező pedig szintén tendenciaszerűen inkább a közöny érzetét keltette, de a gúny elutasítása sem volt egyöntetű, csak tendenciaszerű. A tévesztések okainak feltárása további feladatokat jelent. Kutatásunkban más prozódiai és akusztikai tényezőket, paramétereket nem változtattunk meg, márpedig ismeretes az intenzitás, az energia- vagy a tempóviszonyok szerepe a különböző érzelmek kifejeződésekor (Scherer, 2003, 2013). Ezek beemelése a további kísérletekbe feltehetően javítja az érzelmek felismerésének sikerességét.

Mindezek mellett az sem zárható ki, hogy az 1967-ben rögzített beszéddallamokat a mai hallgató kissé máshogy észleli, mint a korabeli beszélők, de az is elképzelhető, hogy a zenei hangokat alapfrekvenciára átfordító módszerünk sem hibátlan. Hiányossága még a kutatásunknak, hogy nem alkalmaztunk semleges érzelmi töltetű mondatokat, amelyekkel összehasonlításokat lehetett volna végezni. Erre Fónagy és Magdics munkája lehetőséget teremt, mivel minden esetben közölték a semleges érzelmi állapotban megvalósított mondatok dallamának kottaképét. További kutatásokban erre ki kell térni.

Amennyiben sikerül további fejlesztőmunkával a hallgatók számára felismerhetővé tenni a kották formájában lejegyzett érzelmeket, akkor javaslatot tehetünk a közölt dallamok különféle alkalmazásaira. Chow és Brown (2018) is említ néhány ilyen alkalmazást, például a gépi beszéd javítását, vagy az intonáció tanítását az idegennyelv-oktatásban. Ehhez hozzátehetjük még a művészi beszéd tanításának támogatását, vagy a beszélő érzelmi állapotának, attitűdjeinek észlelésével kapcsolatos kísérleteket. Elképzelhető ugyanis, hogy ugyanazon

mondat dallamában egy adott helyen már félhangnyi különbség is hordozhat olyan jelentéstöbbletet, amely alapvetően más kontextusba helyezi a mondatot. Az itt vázolt módszer ilyen kutatásokra is alkalmas, és a beszédprodukció és a beszédészlelés eddig még kevéssé kutatott területeire is bepillantást nyújthat.

Hivatkozások

Bartók, J. (1969). A beszéddallam lejegyzésének kérdéseihez. In D. Pais, & L. Benkő (Eds.), *Dolgozatok a hangtan köréből. Nyelvtudományi Értekezések 67* (p. 88–92). Budapest: Akadémiai Kiadó.

Bartók, J. (1974). Egyéni és társadalmi érvényű elemek a köznyelvi hanglejtésben. In *Általános Nyelvészeti Tanulmányok X. A nyelv hangdomíniuma* (p. 5–20). Budapest: Akadémiai Kiadó.

Boersma, P., & Weenik, D. (2019). Praat: Doing phonetics by computer. v. 6.0.52. URL: `URL:www.praat.org`.

Boros, R. (1969). A magyar beszéddallam jegyzése. In D. Pais, & L. Benkő (Eds.), *Dolgozatok a hangtan köréből Nyelvtudományi Értekezések 67* (p. 76–87). Budapest: Akadémiai Kiadó.

Boros, R. (1971). Néhány hétköznapi beszéddallam. *Magyar Nyelvőr*, *95*, 234–236.

Boros, R. (1975). Beszéddallamok. *Magyar Nyelvőr*, *99*, 41–46.

Bowling, D., Gill, K., Choi, J., Prinz, J., & Purves, D. (2010). Major and minor music compared to excited and subdued speech. *The Journal of the Acoustical Society of America*, *127*, 491–503. doi:`10.1121/1.3268504`.

Chow, I., & Brown, S. (2018). A musical approach to speech melody. *Frontiers in Psychology*, *9*, 247. doi:`10.3389/fpsyg.2018.00247`.

Csűry, B. (1925). A szamosháti nyelvjárás hanglejtésformái. *Magyar Nyelv*, *XXI*, 1–21.

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

Gocsál, Á., & Urbanics, T. (2022). Vocal pleasantness ratings explained by a musical approach. In M. Nourbakhsh, & N. Chahartagh (Eds.), *Proceedings Book. The Second International Conference on Laboratory Phonetics & Phonology (ICLPP* (p. 59–66). Tehran, Iran: Booy-e Kaqaz (BOOKA) Publication.

Gombocz, Z. (1907). A mondat zenei hangsúlyáról. *Uránia*, *8*, 129–131. URL: `http://real-j.mtak.hu/10730/1/MTA_Urania_1907.pdf`. URL:.

Gráczi, T., Gósy, M., Krepsz, V., Markó, A., Huszár, A., Damásdi, N., & Gocsál, Á. (2020). Az alapfrekvencia jellemzői az életkor és nem függvényében. In A. Fóris, A. Bölcskei, J. Bóna, T. Gráczi, & A. Markó (Eds.), *Nyelv, kutúra, identitás. Alkalmazott nyelvészeti kutatások a 21.századi információs térben. III. Fonetika*. Budapest: Akadémiai Kiadó. URL: `https://mersz.hu/hivatkozas/m675nyki3f_23` elérés forrás.

Hegedűs, L. (1930). *Magyar hanglejtésformák grafikus ábrázolása*. Királyi Magyar Egyetemi Nyomda.

Hirst, D., & Cristo, A. (1998). A survey of intonation systems. In D. Hirst, & A. Cristo (Eds.), *Intonation Systems. A Survey of Twenty Languages* (p. 1–44). Cambridge: Cambridge University Press.

Jones, D. (1909). Intonation curves. a collection of phonetic texts, in which intonation is marked throughout by means of curved lines on a musical stave. URL: `https://ia800903.us.archive.org/23/items/intonationcurves00jonerich/intonationcurves00jonerich.pdf` uRL:.

Magdics, K. (1963). From the melody of speech to the melody of music. *Studia Musicologica Academiae Scientiarum Hungaricae*, *4*, 325–346. URL: `http://www.jstor.com/stable/901395`. URL:.

Mangiafico, S. (2016). Summary and analysis of extension program evaluation in r, version 1.20.07. URL: https://rcompanion.org/handbook revised 2024.

Mangiafico, S. (2024). rcompanion: Functions to support extension education program evaluation. r package version 2.4.36. URL: https://cran.r-project.org/web/packages/rcompanion/index.html.

Markó, A. (2013). Kérdő megnyilatkozások a spontán beszédben. *Beszédkutatás*, (p. 42–59).

Markó, A. (2017). Hangtan. In G. Nagy (Ed.), *Nyelvtan* (p. 75–203). Budapest: Osiris Kiadó.

Murdoch, J. (1883). *A Plea for Spoken Language*. Cincinnati, New York: Van Antwerp, Bragg & Co. URL: https://play.google.com/books/reader?id=WmQQAAAAYAAJ&pg=GBS.RA2-PA54&hl=hu.

MuseScore, B. (2021). Musescore 3.6.2.5. URL: https://musescore.org/hu.

R Core Team (2024). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.r-project.org/.

Regner, T. (1861). A magyar nyelv kiejtése. *Akadémiai Értesítő*, *II*, 363–447. URL: http://real-j.mtak.hu/35/1/AkademiaiErtesito_1861-1862_Nyelv.pdf. URL:.

Robledo, J., Hurtado, E., Prado, F., Román, D., & Cornejo, C. (2016). Music intervals in speech: Psychological disposition modulates ratio precision among interlocutors' nonlocal f0 production in real-time dyadic conversation. *Psychology of Music*, *44*, 1404–1418. doi:10.1177/0305735616634452.

Scherer, K. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227–256.

Scherer, K. (2013). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech and Language*, *27*, 40–58.

Scripture, E. (1906). Researches in experimental phonetics. the study of speech curves. URL: https://archive.org/details/researchesinexp01scrigoog uRL:.

Tolnai, V. (1915). Adatok a magyar hanglejtéshez. *Magyar Nyelv*, *11*, 51–59.

Tóth, A. (2016). Kérdő funkciójú megnyilatkozások kisiskolások beszédében. *Beszédkutatás*, (p. 43–58).

Viëtor, W. (1897). Kleine phonetik des deutschen, englischen und französischen. URL: https://openlibrary.org/works/OL13122732W/Kleine_Phonetik_des_Deutschen_englischen_und_franz%C3%B6sischen uRL:.

Changes in the results of voice biometric software using different methods (GMM-UBM, i-vector) in the case of different speech tasks and voice sample durations

Attila Fejes[1,2], Dávid Sztahó[3]

[1]*Special Service for National Security Institute for Expert Services*
[2]*Doctoral School of Law and Political Sciences, Széchenyi István University*
[3]*Budapest University of Technology and Economics*

Abstract

During forensic speaker comparison, the audio forensics expert appointed to perform the investigation works with audio recordings of different types and durations. Distinct speech samples and durations affect the probability data. In order to evaluate biometric identification results, the probability value of the data obtained must be determined so that the expert's report can be accurate and interpreted by other actors in the public proceedings. In the present study, the speech samples of 78 speakers from the forensic voice sample database were compared within the framework of the FORENSICSpeech research project (Beke et al., 2020). The samples include three different types of speech: spontaneous, read, and narration speech. The recording of the samples was repeated after an average of two weeks, and then the audio files were cut into 20, 40, 60, 80, 100, and 120 seconds in duration using automatic editing. The aim of this study is to show how different speech styles and durations affect voice biometric identification results.

Results show that EER (Equal Error Rate) and FRR (False Reject Rate), Cllr (Log likelihood ratio cost) values decrease with increasing duration; however, in the 20–120-second range, the change is not continuous. Similarly, the lowest EER, FRR, Cllr, and Cllr- min values occur in the case of spontaneous speech, followed by narration, while the speech samples of information exchange give the highest Cllr values. The data as a whole is characterized by the fact that the more advanced i-vector method tends to provide more efficient, lower error-rate person identification results.

1. Introduction

The application of biometric methodology is an important element in speech-based speaker identification in forensic science. Biometrics determines the likelihood of the sample owner's identity using certain biological or behavioral char-

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

Beszédtudomány – Speech Science 2025/2. 132–153. DOI: 10.15775/Besztud.2025.2.132-153

acteristics. In the case of voice biometrics, voice is the feature whose uniqueness makes it possible to use as a biometric feature. This is due to the fact that every human being has different biological (physical) dimensions; no two human bodies are exactly alike, including all organs involved in speech production. Speech is also influenced by the individual's personality, sociocultural environment, education, emotional and intellectual intelligence, and a number of factors that may not appear together in the case of another person (Anil et al., 2011; Gráczi et al., 2022; Leemann et al., 2025).

The advantage of voice biometrics is that the result is independent of the expert performing the test; its validity and error rate can be accurately measured, and it can be well-automatized, so it is suitable for mass data processing. Technology provides a probabilistic result, so it does not define a categorical identity or difference. The probability of identity is determined in different forms by the high-tech systems available today, calculated from Score data (Morrison, 2013; Kelly et al., 2019), Likelihood Ratio (LR) (Van der Vloed, 2016; Zhang. & Tang, 2018), and its decimal-based logarithmic value (LLR) (Jessen et al., 2019). In our research, we performed measurements with the Batvox software, working with two distinct version numbers and different biometric identification engines. Both versions' output is LR data, and their inputs are the voice samples to be compared. Version 3.1 is based on GMM-UBM (Gaussian Mixture Models – Universal Background Model) (Zhang. & Tang, 2018), while version 4.1 uses the PLDA (Probabilistic Linear Discriminant Analysis) method with i-vector extraction (Van der Vloed, 2016).

In the detection and proof of criminal offenses, it is common for the audio of an unknown person to be compared with the speech sample of a known speaker recorded by an expert working on the case. During sampling, the expert records several speech samples of different types of the known person and compares each with the unknown speaker's voice recording (Fejes, 2022). The type of the speech sample influences the result of the identification; however, the extent of this can be determined by implementing performance tests. The probability value is also

affected by the duration of the research material, for which different biometric methods set dissimilar threshold levels (Meuwly, 2009).

In the present study, we aimed to show the effect of changing the type and duration of the speech sample on the biometric identification results. These factors are important because in forensic science, the expert often only has short recordings at his disposal, so we need to understand the relationship between duration and performance for a given method. On the other hand, in forensic audio sampling, the expert working on a given case uses several types of samples, so we need to know the characteristics of different types of samples in terms of identification results. In our study, we focused on speech duration and speech type. For the comparison, we created the same test conditions for both versions, and in this way, we produced 36 identification matrices per software version using samples from speakers of both genders. Data were converted into LLR format for evaluation, and performance metrics and other data were evaluated using the Bio-Metrics 1.8 software (Kelly et al., 2019).

1.1. Methodology of Forensic Voice Comparison in Hungary

The purpose of Forensic Voice Comparison (FVC) is to determine the probability of the compared speaker's identity. In typical cases, there is a sample of a recorded unknown person via wiretapping and a suspect speaker whose identity is known. In other cases, samples of unknown speakers need to be compared to determine the probability of identity.

The FVC methodology includes auditory phonetic-linguistic analysis, acoustic measurements, and the use of voice biometrics technology. In the auditory analysis, the expert examines the features of articulation, language and speech, dialect, idiolect, hesitation phenomena, speech pathology, etc. With special expert software, one can measure, for example, the similarity of formants in matching sounds, fundamental frequency (f0), and formant frequencies (European Network of Forensic Science Institutes, 2022). After the phonetic-linguistic analysis and acoustic measurements, the audio forensics expert assesses the

similarities and differences in these features and determines the probability of identity utilizing available scientific background information.

Voice biometric measurements are the final stage of the speech analysis because the results can influence the expert, as they may cause cognitive bias (Kovács, 2017). The auditory phonetic-linguistic analysis and the acoustic measurements depend on the expert's judgement. In contrast, voice biometric measurements are objective and reproducible, and the results are independent of the person conducting the analysis. Automatic speaker recognition systems are the state-of-the-art technologies in voice comparison nowadays, and an efficient and powerful tool (Ramos, 2007:9–10). The voice biometric methodology is used in Forensic Automatic Speaker Recognition (FASR) systems, which are used by audio forensic experts. Of these, Batvox is not the latest technology, but it is still in use in Hungarian forensic science.

2. Methods

2.1. Measurements

Speech samples were selected from the FORENSICSpeech (Beke et al., 2020; Sztahó et al., 2021) project database. We used Spontaneous, Read, and Narration speech samples. Recordings were conducted in a quiet office room, similar to those found in expert sampling, using a laptop, an external sound card, and a condenser microphone. There was no background noise at the recording site, and the recording was made at a sampling rate of 44,100 Hz with a depth of 16 bits. The age of the speakers ranged from 16 to 48 years, and 39 female and 39 male speech samples were measured. Samples were recorded during two separate sessions on different days, which will be referred to as sessions 1 and 2. Spontaneous conversations recorded at the first session served as models (i.e., the known speaker's speech sample). These were compared with the Spontaneous, Read, and Narration-like monologue (the recollection of the events of the speaker's previous day) audio recordings, comparing the first session with the second session of audios. Samples were cut into clips of six different dura-

tions with 20-second increments, ranging from 20 seconds to 120 seconds. The methods of audio expert sampling were applied to the recordings, similarly to a real forensic condition. All recordings were automatically edited. First, pauses longer than 500 milliseconds were removed, and then the samples were divided into chunks with the desired durations. No phonological boundaries were considered during the chunking. Each chunk has the exact target duration. Accordingly, during the measurements, we created 36 identification matrices per software version with speech samples of male and female speakers, as all three different styles were compared for each of the six different durations. A typical matrix contains 39 x 39 = 1,521 probability values, as the software compares all speakers against each other. There are LLR data of 39 same (SS-Same Source <same speakers>) and 1,482 different (DS-Different Source <different speakers>) speakers. On the x and y axes are voice samples from the same speakers, recorded at different times and speech styles. The samples from the first recordings are plotted on the y-axis, and the samples from the second recordings are plotted on the x-axis. The samples from the first recordings were always the same, while in each matrix, the length and speech style of the samples from the second recordings were adjusted.

2.2. The Bayesian framework and the method of evaluating the results

In the methodology of speech-based personal identification, we analyze (through perceptual and acoustic-phonetic studies) and measure (using voice biometrics) various sound parameters (Drygajlo et al., 2015). Depending on the methodology, the characteristics are evaluated in text, measured manually, or the biometric software calculates the probability of identity from the data using mathematical and statistical methods (Craig, 2010). Since we do not know the characteristics of speech that uniquely represent the speaker, we do not look for matching data; instead, we compare and infer probability. In addition to that, the even greater difficulty is that there is within-speaker variation in voice characteristics. Thus, in a speaker identification study, two hypotheses must be considered and calculated by the voice biometrics software: the probability of evidence

(Morrison, 2009). Accordingly, the null hypothesis (H0) is that the speakers on the two audio recordings being compared are the same, and the alternative hypothesis (H1) is that the speakers on the two recordings are different. The software then provides the probabilities for each of these hypotheses to be true. The Bayesian framework (Meester & Slooten, 2021) is suitable for determining the strength of the evidence and the probability of identity by considering and calculating both probabilities. In the Bayesian approach, the competence of the audio engineering expert conducting the study is to determine and evaluate LR. The relationship between probabilities is shown by the formula in Figure 1 below.

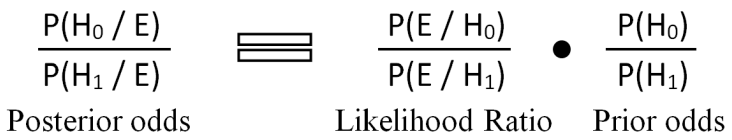


Figure 1: Bayesian framework formula.

H0 is the hypothesis that the speaker on the two recordings is the same, and $P(E|H0)$ is the probability of observing the evidence given that H0 is true. H1 is the hypothesis that the speaker on the two recordings is different, and $P(E|H1)$ is the probability of observing the evidence given that H1 is true. E (Evidence) denotes the sound recording as evidence, and P (Probability) expresses probability. It can be seen that a priori knowledge is weighted by the strength of the evidence as determined by the expert, taking into account both hypotheses. The statement about the strength of evidence is the Likelihood Ratio (LR), which is the ratio of the probability of the two hypotheses, denoted by H0 and H1 in the formula.

Biometric measurements were performed with Batvox 3.1 and 4.1. For both versions, the user must create a reference population database, which should consist of audio files with the same characteristics (gender, language, channel, and speech type) as the audio recording group used as a model. The same

reference population database, comprising 80 female and male audio files, was used for both software versions in our study. The audio files of the reference population databases were randomly selected from the Hungarian Spontaneous Speech Database (Gósy et al., 2012). The first step in biometric identification is feature extraction in the case of both software versions. Due to the uniqueness of the vocal tract and other organs involved in speech, the forms of speech sounds are unique characteristics of the speaker, so their acoustic parameters can be measured and subtracted by the software as described below.

Voice Biometrics technology, as applied in our research, utilizes the envelope of the audio signal spectrum to extract its characteristics. To do this, both Batvox versions split the voice into 20-millisecond windows with 50% overlap. Then, they extract the individual characteristics from the spectrum using the Mel-Frequency Cepstrum Coefficient (MFCC) method. In addition to spectral characteristics, the subtraction of phonetic and prosodic characteristics, fundamental frequency, and energy conditions of voice provide additional data on the speaker's speech. After feature extraction, the system sets up feature vectors and performs speech modeling. The two Batvox versions used in the study have the same interface, and the measurement sessions are also configured in the same way, with one exception: in the more modern version (4.1), the known speaker's audio recording has a minimum duration of 30 seconds, while version 3.1 defines a minimum of 40 seconds as an input requirement.

The results were obtained in LR format, which were converted to a decimal-based logarithmic format (LLR) for full-scale evaluation, as only a Tippet Plot graph can be created for LR format data. "The Tippett plot is a cumulative probability distribution plot expressing the proportion of likelihood ratios (LRs) greater than a given value, i.e., $P(LR(H) > LR)$, for cases corresponding to the H0 hypothesis (biometric samples are from the same source) and the H1 hypothesis (biometric samples are from different sources)" (Oxford Wave Research, 2025). In this study, H0 and H1 in the above Bayesian approach formula are not the same as the H0 and H1 denoted below. Outside the formula, H0 data is equal to the probability values of same speakers, and H1 data is equal to the

probability values of different speakers. The evaluation was performed using Oxford Wave Research Bio-Metrics 1.8 (Oxford Wave Research, 2025) using the following outputs to analyze the data:

- Mean of H0 and H1: arithmetic mean of LLR data from hypotheses H0 and H1;
- Standard Deviation of H0 and H1: standard deviation of LLR data from hypotheses H0 and H1;
- Log likelihood ratio cost (Cllr): the degree of calibration, the accuracy of the system;
- False acceptance rate (FAR) is the rate at which the comparison between two different individuals' samples is erroneously accepted by the system as a true match. In other words, FAR is the percentage of impostor scores that are higher than the decision threshold;
- False rejection rate (FRR) is the percentage of times when an individual is not matched to his/her own existing reference templates. In other words, FRR is the percentage of the genuine scores that are lower than the decision threshold;
- Equal error rate (EER) is the rate at which both acceptance and rejection errors are equal (i.e., FAR=FRR). Generally, the lower the EER value, the higher the accuracy of the software;
- Detection Error Trade-off (DET) plot: represents FAR and FRR values.

3. Results

The values of Same Source (SS) LLR data, which indicate the identity of speakers, tend to increase with longer, higher-quality audio materials, and the variance also decreases as the system's robustness improves. The values of Different Source (DS) LLR data, which indicate the likelihood that the compared

audio samples come from different individuals, typically decrease as the signal-to-noise ratio (SNR) and duration increase. This is because a low SNR reduces the performance of the analysis, while a longer duration enhances it. In general, the effectiveness of feature extraction is reduced by noise and shorter audio length. The degree of standard deviation of SS and DS LLR data, interpreted separately, suggests discriminatory power: a high-performance voice biometrics software identifies matching speakers with low standard deviation and high probability results, while also distinguishing between different samplers with low standard deviation and values closer to the minimum.

The smaller the Cllr value, the more accurate the system is, and the $0 <$ Cllr < 1 relationship is a feature of well-calibrated (Sztahó & Fejes, 2023) applications.

The speech style notations used in Tables 1 through 6 and Figures 2 through 6 are as follows:

- 2.1: spontaneous (sess2_task1) audio sample,
- 2.2 : read (sess2_task2) audio sample,
- 2.3 : narration (sess2_task3) audio sample.

A decimal logarithmic transformation was applied to the LR values. The primary advantage of the LLR data obtained in this manner is the symmetric scale. LLR values less than 0 suggest different speakers (H1) and ones greater than 0 suggest identical ones (H0). The evaluation was based on the trends observed in Tables 1 through 6 and Figures 2 through 6, as well as on the DET curves generated using Bio-Metrics software.

3.1. Mean values of H0 and H1

In the case of the two different biometric identification software versions, as the speech duration increases, SS values also show an increase in both genders. However, in certain cases, different phenomena can be observed. The LLR results are shown in Tables 1 and 2 and Figure 2.

Table 1: H0 mean values of LLR data for audio samples of female speakers.

mean of H0, female speakers						
duration (s)	**Batvox 3**			**Batvox 4**		
	Spo	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	9.28753	7.20152	9.06353	5.5569	3.15539	4.52925
40	9.51955	8.53179	9.30864	6.45635	4.13735	5.06021
60	9.62084	9.02421	9.40291	7.11815	4.37704	5.77591
80	9.82359	9.0712	9.50403	7.4881	4.81714	6.27967
100	9.86962	9.49392	9.64835	7.75766	4.70182	6.51493
120	9.92827	9.33723	9.66373	7.73667	4.99168	6.47422

Table 2: H0 mean values of LLR data for audio samples of male speakers.

mean of H0, male speakers						
duration (s)	**Batvox 3**			**Batvox 4**		
	Spo	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	9.45068	7.45508	9.32333	5.74415	3.73451	5.08057
40	9.81685	8.27898	9.95291	6.60618	3.93672	6.45817
60	9.98095	8.60391	9.97565	7.79713	4.2863	6.99744
80	9.99395	9.02827	9.96294	7.62297	4.40198	7.36554
100	9.99173	8.90239	9.58461	7.90105	4.38097	6.35662
120	9.99994	9.17295	9.99726	8.20106	4.38394	7.57775

In the case of the Narration (Narr) type audio samples of male speakers, a slight decrease is observed for both Batvox versions at a duration of 100 seconds, after which it jumps to the local maximum of the average SS for the samples with a duration of 120 seconds. The highest LLR values were measured in Spontaneous (Spo) speech, followed by Narration, then Read speech. Note that data should be interpreted separately for each system, as the different biometric methods of the software versions affect the order of magnitude of the LLR values. For samples 2.1 and 2.3, it can be seen that even at 20 seconds, the LLR is above

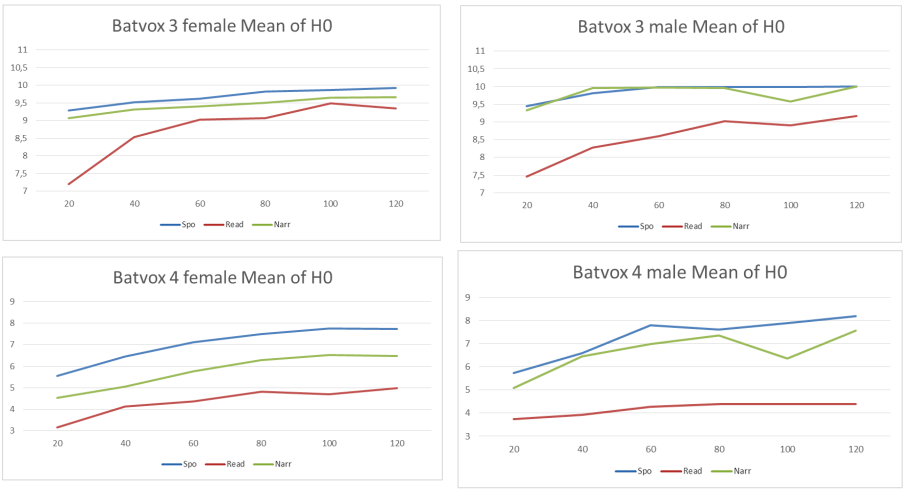


Figure 2: Graphs of the mean of LLR data for Hypothesis H0.

9 for Batvox 3.1, while Batvox 4.1 is more sensitive to speech duration based on the SS averages.

Contradictory results were obtained for the mean of the H1 DS data (see Tables 3 and 4), and our hypothesis – the mean decreases with increasing speech duration – cannot be supported in either case. For all three speech styles and both software versions, the mean of the DS data tends to increase, indicating that the longer the speech duration, the less likely the system is to differentiate speakers. However, it does not indicate a malfunction of the system, but rather reveals that the average of DS values cannot be used as a measure of performance. This statement is supported by the relations discussed in the following subsections.

3.2. Standard Deviation of SS and DS values

For Standard Deviation (SD), we assumed that longer speech duration results in lower SD, thus increasing the discriminating power of the applied method, but this was only partially confirmed by the data shown in Tables 55 and **??** and Figure 3.

Table 3: DS data mean values of LLR data in the speech samples of female speakers.

mean of H1, female speakers						
duration	**Batvox 3**			**Batvox 4**		
(s)	**Spo**	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	−0.4067	−0.4665	−0.5536	−2.2132	−2.3277	−2.3132
40	−0.235	−0.3643	−0.4232	−2.1648	−2.2746	−2.3075
60	−0.1089	−0.2745	−0.3603	−2.1117	−2.2828	−2.2945
80	−0.0763	−0.251	−0.2648	−2.1115	−2.2485	−2.2681
100	0.01711	−0.1147	−0.203	−2.054	−2.2508	−2.2709
120	0.05172	−0.0751	−0.2005	−2.0531	−2.2432	−2.2514

Table 4: DS data mean values of LLR data in the speech samples of male speakers.

mean of H1, male speakers						
duration	**Batvox 3**			**Batvox 4**		
(s)	**Spo**	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	−0.3803	−0.6292	−0.2814	−1.7706	−1.8261	−1.8127
40	−0.3156	−0.5153	−0.2109	−1.7646	−1.8184	−1.762
60	−0.2259	−0.4623	−0.1506	−1.6615	−1.8013	−1.7183
80	−0.2729	−0.4894	−0.0934	−1.6659	−1.8268	−1.6934
100	−0.1643	−0.4638	−0.1912	−1.6305	−1.7941	−1.7092
120	−0.1663	−0.4585	−0.0756	−1.5791	−1.8282	−1.664

From the above data, it can be seen that in Batvox 3.1, the Standard Deviation of the SS data values decreases with increasing speech duration in all three speech styles; however, a notable jump can be observed in the Narration and Read style samples. The SD of DS data shows an increasing or fluctuating trend depending on the software version and the gender of the speakers. Nevertheless, it should be noted that the absolute range of the values is smaller than in the case of SS data. We also found that the two software versions are characterized by a lower Equal Error Rate for the same speakers and a higher Equal Error Rate for different speakers. This statement is supported by the histograms

Table 5: H0 SD of LLR data in the speech samples of female speakers.

Standard Deviation of H0, female speakers						
duration	**Batvox 3**			**Batvox 4**		
(s)	**Spo**	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	1.47925	2.87787	1.76351	2.77122	2.27976	2.83024
40	1.47628	2.18109	1.54118	2.72329	2.98783	2.80453
60	1.12278	1.91644	1.38321	2.79839	3.03602	2.60051
80	0.56944	1.78398	1.2422	2.63205	3.07059	2.77972
100	0.55016	1.29549	1.21685	2.56946	3.04013	2.83464
120	0.44792	1.52514	1.2573	2.43123	2.96467	2.77677

Table 6: H0 SD of LLR data in the speech samples of male speakers.

Standard Deviation of H0, male speakers						
duration	**Batvox 3**			**Batvox 4**		
(s)	**Spo**	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	1.17316	2.84889	1.43012	2.37577	2.57552	2.61177
40	0.82076	2.53241	0.1818	2.34037	2.64598	2.55186
60	0.10194	2.32753	0.1301	2.18984	2.86471	2.41174
80	0.03495	2.29878	0.21188	2.34126	2.79255	2.47985
100	0.02988	2.25904	1.34614	2.0958	2.851	2.69386
120	0.00035	1.90806	0.01453	1.9704	2.80516	2.38392

shown in Figure 4, where you can see the distribution of the values generated during the comparison of spontaneous speech samples of female speakers with Batvox 3.1. The y-axis shows the distribution rate, and the x-axis shows the LLR probability values.

3.3. EER and Cllr values

EER and Cllr values are measures of the performance of the biometric speaker identification software. EER is a measure of discrimination that shows how well a software can distinguish between same and different (SS-DS) speakers. In forensics, we typically compare the speech samples of two speakers to

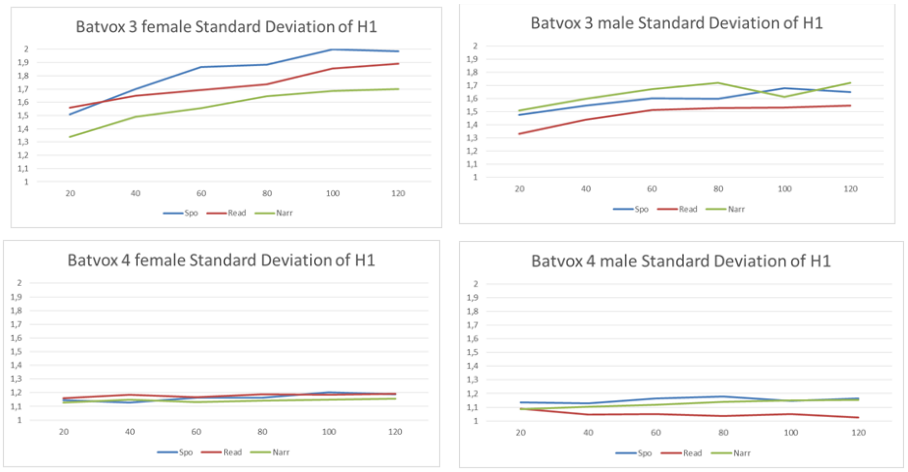


Figure 3: Graphs of the Standard Deviation data for DS.

determine the likelihood of identity, making it particularly important for the results to be robust enough to identify the same speaker and differentiate between different individuals. Another important criterion is to keep the error rates as low as possible (FAR, FRR, EER) in order to prevent erroneous expert reports. When determining performance, it is essential to conduct measurements in the same test environment (speech sample database), thereby comparing different software versions and performing tests with the same speech samples. Tables 7 and 8 and Figure 5 below show the EER data for the two Batvox systems for speakers of both genders.

For both systems, we obtained low EER results in the vast majority of both female and male samples, indicating that the system reliably separates the same and different individuals based on their voice patterns, even at short speech durations. More so in the case of female speakers, and to a lesser extent for male speech samples, it can be seen that the newer, more modern Batvox 4.1 software achieves a lower EER value. However, for both genders, it can be observed that the values for the 20-second recordings contradict the trend: they exhibit smaller or larger values compared to the subsequent 40-second measurement runs in several cases.

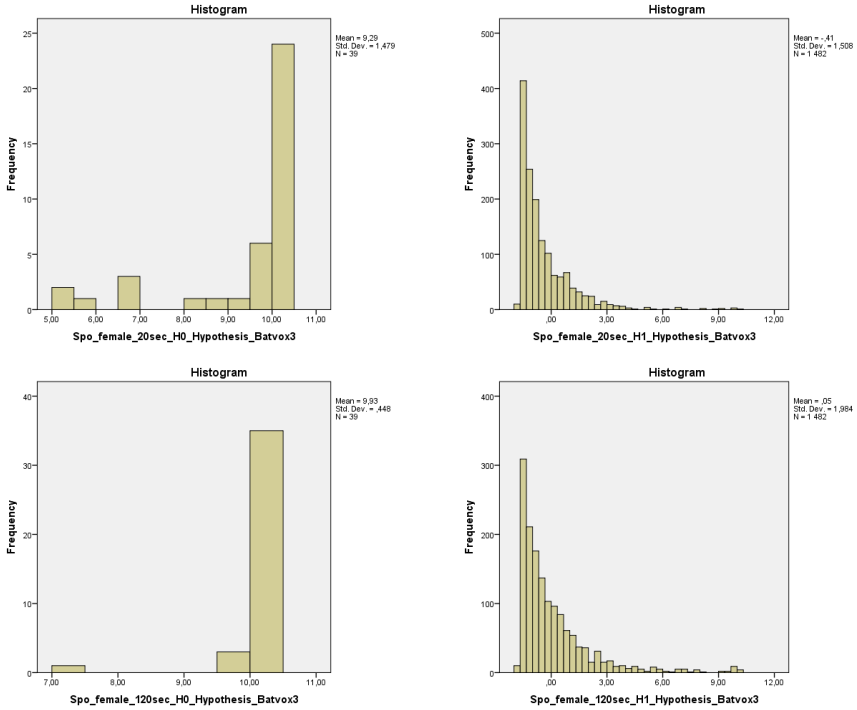


Figure 4: Histograms of SS and DS data for 20- and 120-second audio samples of the same speech style (The y-axis shows the distribution rate, the x-axis shows the LLR probability values).

Table 7: EER values for speech samples of female speakers.

EER, female speakers						
duration	**Batvox 3**			**Batvox 4**		
(s)	**Spo**	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	0.5735	5.1282	2.193	0.4386	5.0607	2.2942
40	2.5641	2.8677	2.5641	0.5735	4.9595	2.0243
60	2.5641	2.6653	2.5641	0.3036	5.1619	0.1687
80	2.0429	2.5978	2.5304	0.5398	2.5641	0.2024
100	2.1592	2.5641	2.5641	0.2699	0.6073	0.2362
120	2.1255	2.5978	2.5641	0.1012	1.9568	0.1687

Table 8: EER values for speech samples of male speakers.

EER, male speakers						
duration	**Batvox 3**			**Batvox 4**		
(s)	**Spo**	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	0.4723	4.5547	2.2605	2.5641	2.8003	2.5978
40	0.5398	2.5978	0.135	0.5735	7.4224	0.3374
60	0.135	2.8003	0.2024	0.1687	5.1282	0.1012
80	0.1012	3.0702	0.2699	0.1687	5.1282	0.135
100	0.0375	5.1282	2.5641	0.0337	7.6586	2.5978
120	0.1012	2.5641	0.1687	0.135	4.892	0.2362

Figure 5: Graphs of EER values.

Cllr refers to the accuracy of a biometric speaker identification software, with lower values being more favorable. Cllr measures the discrimination error (how much overlap between H0 and H1 LRs there is) and calibration error (whether the LRs are too large or too small). A Cllr of 0 is a perfect system, and a Cllr of 1 is a system that is completely worthless (performs at chance level). Tables 9 and 10 show the Cllr data for the two systems and speakers of both genders.

Table 9: Cllr values for speech samples of female speakers.

Cllr, female speakers						
duration	**Batvox 3**			**Batvox 4**		
(s)	**Spo**	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	0.5136	0.5202	0.4496	0.1523	0.2365	0.1722
40	0.5952	0.5489	0.512	0.1462	0.2228	0.1591
60	0.6604	0.584	0.5401	0.1476	0.2114	0.1383
80	0.6744	0.595	0.5814	0.1472	0.1846	0.1377
100	0.7264	0.659	0.6064	0.157	0.1752	0.1352
120	0.7347	0.677	0.6087	0.1524	0.1716	0.1368

Table 10: Cllr values for speech samples of male speakers

Cllr, male speakers						
duration	**Batvox 3**			**Batvox 4**		
(s)	**Spo**	**Read**	**Narr**	**Spo**	**Read**	**Narr**
20	0.52341	0.44963	0.56021	0.19851	0.25501	0.21117
40	0.55134	0.49186	0.59174	0.18901	0.25449	0.18765
60	0.55134	0.51943	0.62482	0.20237	0.25218	0.19076
80	0.56838	0.50988	0.65201	0.20331	0.23497	0.19612
100	0.61989	0.51981	0.60192	0.20291	0.26457	0.20612
120	0.61216	0.51705	0.6584	0.21286	0.24123	0.2006

During our measurements, we also observed fluctuating trends in Cllr values in the case of Batvox 4.1. In the case of Batvox 3.1, Cllr increases with increasing speech duration. To obtain an accurate picture of the characteristics of the operation of both software versions, LLR values were plotted on histograms. Paired t-tests were performed between the trials of same and different speakers to show the separation power between the two hypotheses.

The histograms in Figure 6 show that the results of the same and different speakers are well separated by both software versions. It can be seen that the more advanced Batvox 4.1 is more likely to identify the match and differentiate

between different speakers. It is more sensitive to speech duration compared to version 3.1, yet it identifies matching speakers with a high LLR at 60 seconds. The broader histogram of the same speaker LLR values suggests that this system is more sensitive to similarities/differences in speaker voice characteristics and can measure this similarity better.

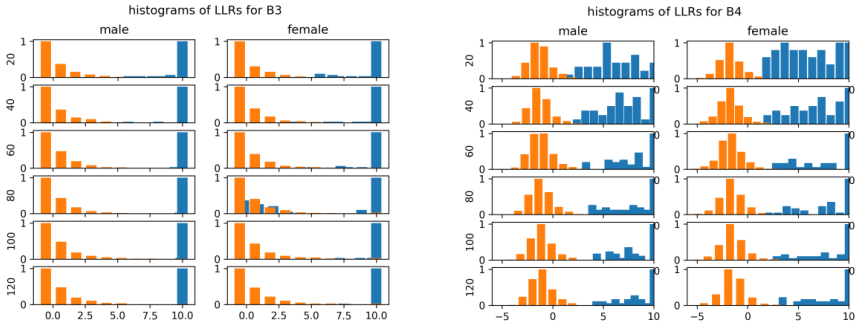


Figure 6: Histograms of LLR data. Yellow: different speaker trials, blue: same speaker trials. B3 and B4 represent Batvox 3.1 and 4.1 systems, respectively.

Although p-values do not reflect the distinctive power between the two groups (same speaker versus different speaker), the tendency in their value suggests that there is a real effect of sample duration. For better visualization, the logarithm of the p-value is shown in Figure 7. Indeed, this is not a standard way of representing the significance of differences between groups, and the p-value cannot be considered a "measure" of the difference, but it does give us a general idea of the trend in the magnitude of differences between groups as a function of recording length.

4. Conclusion

In forensic identification tests for forensic purposes, the expert often only has a short duration of speech samples available. In such cases, it is possible to determine the probability of speaker identity with high accuracy using voice biometrics. In our research, we have demonstrated that even for voice recordings with a gross duration of 20 seconds, in which the net, uninterrupted speech

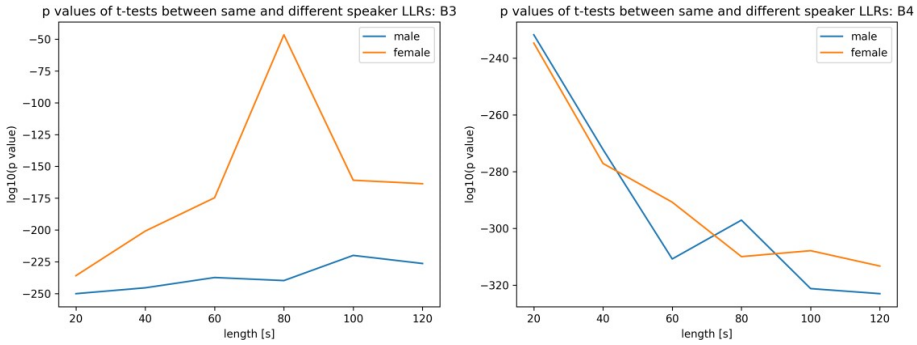


Figure 7: The p-values of the t-tests.

duration is even lower, the automatic identification software is likely to identify or distinguish between different speakers. Overall, the more advanced Batvox 4.1 performs better than the previous version, Batvox 3.1; the Cllr and EER values are mostly lower for Batvox 4.1. In general, the higher the SS value and the lower the DS data is, the better performance we can expect from the system. However, the older software version also produced good results, with a low EER error rate for shorter recordings.

Three different types of speech were used as model test recordings (spontaneous, read, narrative style) that modeled the "known speaker" speech sample of the common forensic case, and compared this with the spontaneous sound sample of the "unknown person" two weeks apart. By evaluating the measurement results, we obtained better results with the spontaneous and narrative-type speech samples compared to the samples of the "read" speech style. This is promising in terms of expert voice sampling methodology: in the future, the use of read voice sampling in biometric speech identification is of limited use; forensic voice comparison methodology should therefore adapt these results. The other conclusion is that the error rate is significantly reduced at 60 seconds, so voice biometric measurements can be made reliably at or above this audio duration. The forensic audio sample database created as part of the FORENSICSpeech research project provides an excellent research base for forensic biometric speech identification studies. A basic requirement for our research is to have more than

one speaker-style speech sample recorded at different times. Thus, in the future, the methodology of identification tests to be performed on sound recordings in Hungarian can be developed using new research results.

The Hungarian-language database of forensic speech samples will also provide significant support for speech recognition research (Kamath et al., 2019), which requires a large corpus of Hungarian-language data. A speech sample database modeling a typical forensic case is a prerequisite for both speech recognition and speaker identification research. It can be used for performance studies and to support research on speech and speaker recognition. Using the above results allows the development of systems with higher accuracy in the future.

References

Anil, K., Arun, A., & Karthik, N. (2011). *Introduction to Biometrics*. New York, Dordrecht Heidelberg London, DOI: Springer. doi:10.1007/978-0-387-77326-1.

Beke, A., Szaszák, G., & Sztahó, D. (2020). Forvoice120+ magyar nyelvű utánkövetéses adatbázis kriminalisztikai célú hangösszehasonlításra. In G. Berend, G. Gosztolya, & V. Vincze (Eds.), *XVI. Magyar Számítógépes Nyelvészeti Konferencia* (pp. 95–101). Szeged.

Craig, A. (2010). *Mathematics and Statistics for Forensic Science*. Chichester: John Wiley & Sons Ltd.

Drygajlo, A., Jessen, M., Gfroerer, S., Wagner, I., Vermeulen, J., & Niemi, T. (2015). *Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition*. European Network of Forensic Science Institutes. Agreement Number: HOME/2011/ISEC/MO/4000002384.

European Network of Forensic Science Institutes (2022). Best practice manual for the methodology of forensic speaker comparison.

Fejes, A. (2022). Hangazonosítás. In C. Fenyvesi, C. Herke, & F. Tremmel (Eds.), *Kriminalisztika*. Budapest: Ludovika Egyetemi Kiadó.

Gráczi, T., Fejes, A., Krepsz, V., & Huszár, A. (2022). Speaker recognition over the course of 10 years and across speech style. *Alkalmazott Nyelvtudomány, 22: Különszám*, 94–109.

Gósy, M., Gyarmathy, D., Horváth, V., Gráczi, T., Beke, A., Neuberger, T., & P, N. (2012). Bea: Beszélt nyelvi adatbázis. In M. Gósy (Ed.), *Beszéd, adatbázis, kutatások* (pp. 9–25). Budapest: Akadémiai Kiadó.

Jessen, M., Bortlík, J., P., S., & A, S. Y. (2019). Evaluation of phonexia automatic speaker recognition software under conditions reflecting those of a real forensic voice comparison case (forensic_eval_01. *Speech Communication*, *111*, 22–28. doi:10.1016/j.specom.2019.05.002.

Kamath, U., Liu, J., & Whitaker, J. (2019). *Deep Learning for NLP and Speech Recognition*. Switzerland AG: Springer Nature. doi:10.1007/978-3-030-14596-5.

Kelly, F., Fröhlich, A., Dellwo, V., Forth, O., Kent, S., & Alexander, A. (2019). Evaluation of vocalise under conditions reflecting those of a real forensic voice comparison case (forensic_eval_01. *Speech Communication*, *112*, 30–36. doi:10.1016/j.specom.2019.06.005.

Kovács, G. A. (2017). Egyes kognitív, emberi tényezők szerepe a szakértőivélemény-alkotásban. *Belügyi Szemle*, *65*, 89–103. doi:10.38146/BSZ.2014.10.7.

Leemann, A., Perkins, R., Buker, S., & Foulkes, P. (2025). *An Introduction to Forensic Phinetics and Forensic Linguistic*. New York: Routledge. doi:10.4324/9780367616595.

Meester, R., & Slooten, K. (2021). *Probability and Forensic Evidence*. Cambridge: Cambridge University Press. doi:10.1017/9781108596176.

Meuwly, D. (2009). Speaker recognition. In A. Jamieson, & A. Moenssens (Eds.), *Wiley Encyclopedia of Forensic Science Volume 5*. West Sussex: John Wiley & Sons Ltd.

Morrison, G. (2009). Forensic voice comparison and the paradigm shift. *Science and Justice*, *49*, 298–308. doi:10.1016/j.scijus.2009.09.002.

Morrison, G. (2013). Tutorial on logistic-regression calibration and fusion: converting a score to a likelihood ratio. *Australian Journal of Forensic Sciences*, *45*, 173–197. doi:10.1080/00450618.2012.733025.

Oxford Wave Research (2025). Bio-metrics. URL: https://oxfordwaveresearch.com/products/bio-metrics/#:~:text=The%20Tippett%20plot%20is%20a%20cumulative%20probability%20distribution,H1%20hypothesis%20%28biometric%20samples%20are%20from%20different%20sources%29.).

Ramos, D. (2007). Forensic evaluation of the evidence using automatic speaker recognition system.

Sztahó, D., Beke, A., & Szaszák, G. (2021). Forvoice 120+: Statisztikai vizsgálatok és automatikus beszélő verifikációs kísérletek időben eltérő felvételek és különböző beszéd feladatok szerint. In *XVII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, 2021. január 28–29.

Sztahó, D., & Fejes, A. (2023). Effects of language mismatch in automatic forensic voice comparison using deep learning embeddings. *Journal of Forensic Sciences*, *68*, 871–883.

Van der Vloed, D. (2016). Evaluation of batvox 3.1 under conditions reflecting those of a real forensic voice comparison case (forensic_eval_01. *Speech Communication*, *100*, 13–17. doi:10.1016/j.specom.2018.04.008.

Zhang., C., & Tang, C. (2018). Evaluation of batvox 4.1 under conditions reflecting those of a real forensic voice comparison case (forensic_eval_01. *Speech Communication*, *85*, 127–130. doi:10.1016/j.specom.2016.10.001.