

A nyelvgesztusok ultrahangos vizsgálata a magyar beszédben

Markó Alexandra¹, Deme Andrea^{2,1,3}, Juhász Kornélia^{3,1,2}, Grácz Tekla Etelka^{3,1}

¹MTA–HUN–REN NYTK Lendület Neurofonetikai Kutatócsoport

²Eötvös Loránd Tudományegyetem

³HUN–REN Nyelvtudományi Kutatóközpont

Abstract

Tamás Gábor Csapó, as one of the founding members of MTA–ELTE “Lendület” Lingual Articulation Research Group, introduced the technique of ultrasound tongue imaging (UTI) into articulatory research in Hungary. He mastered this technique during a Fulbright Scholarship at Indiana University, Bloomington. He did pioneering work in the development of UTI data analysis methodology, silent speech interfaces, and other digital applications of UTI. His achievements are acknowledged and his work is referenced worldwide. In the present paper, after a brief outline of the technique of 2D ultrasound tongue imaging, we review the research, results, and application areas in which Tamás Gábor Csapó (with his co-authors) made his marks within the fields of articulatory research and articulatory-to-acoustic conversion.

1. Bevezetés

Csapó Tamás Gábor (1985–2024) honosította meg Magyarországon az ultrahangos képalkotó technikát a beszéd artikulációs működéseinek vizsgálatára, amelyet Fulbright-ösztöndíjasként az Indianai Egyetemen, Bloomingtonban sajátított el, Steven Lulich vezetése alatt. Csapó Tamás Gábor úttörő munkát végzett a nyelvultrahangos elemzési módszertan, a némabeszéd-interfész és egyéb beszédtechnológiai alapú digitális alkalmazások fejlesztése terén, nemzetközi viszonylatban is. Tanulmányunkban a nyelvultrahang módszertanának vázlatos ismertetése után áttekintjük azokat a kutatási és alkalmazási területeket, ame-

Email addresses: marko.alexandra.phd@gmail.com (Markó Alexandra), deme.andrea@btk.elte.hu (Deme Andrea), juhasz.kornelia@nytud.hun-ren.hu (Juhász Kornélia), graczi.tekla.etelka@nytud.hu-ren.hu (Grácz Tekla Etelka)

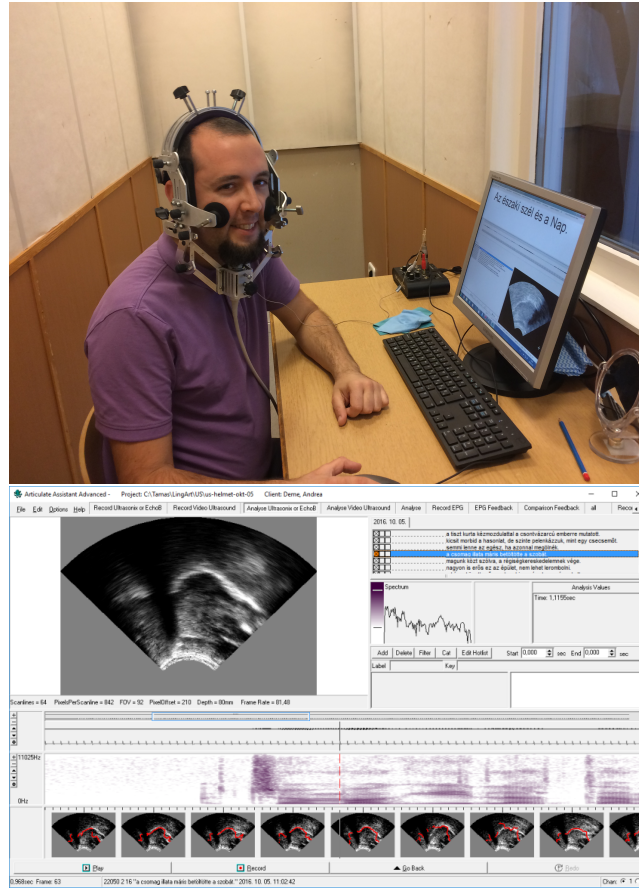
lyeken Csapó Tamás Gábor – mindazon kutatócsoportokkal, amelyekben dolgozott – maradandót alkotott.

2. A nyelvultrahang mint artikulációs vizsgálati eszköz

Az akusztikai elemzés holisztikusságával ellentétben az artikulációs vizsgálati módszerek általában csak részleges információt nyújtanak: szinte mindegyik az artikulációs gesztusoknak csak egy kis részhalmazához fér hozzá, és gyakran csak korlátozott térbeli és/vagy időbeli felbontású információt szolgáltatnak. A teljes artikulációs csatorna vizsgálata még mindig technológiai kihívást jelent, költséges, és a legtöbb műszer bizonyos mértékig beavatkozik a beszédképzés folyamatába.

A nyelvultrahang (ultrasound tongue imaging: UTI, magyarul jellemzően UH) egyetlen aktív beszéd szerv vizsgálatára szolgál: a nyelv felszínét képezi le (bár bizonyos feltételekkel – lásd alább – a szájpad is kirajzoltatható a segítségével, illetve esetenként a gége helyzetének követésére is használják, vö. Kochetov, 2020). Az ultrahang olyan képalkotó technika, amely ultramagas frekvenciájú hanghullámot bocsát ki egy piezoelektromos kristály gerjesztése révén. Ez a hanghullám áthalad a szöveten és visszaverődik annak felületéről (Stone, 2005). Az áll alatt elhelyezkedő ultrahangos jelátalakító így készít képet a nyelvről. A (kétdimenziós) ultrahangfelvételek tipikus eredménye szürkeárnyaltos képek sorozata, amelyeken a nyelv felszíni kontúrja a környező szöveteknél és a levegőnél nagyobb fényerősséggel látszik (1. ábra alsó panel).

Az ultrahangos technika alkalmazását a beszéd szervek működésének vizsgálatára az 1960-as évek végén Kelsey, Minifie és Hixon (1969) javasolták, elsősorban a röntgennel végzett vizsgálatok alternatívájaként, amelyet egészséges emberek esetében a sugárterhelés miatt indokolatlanul kockázatosnak tartottak. Az ultrahang ezzel szemben – ahogy írják (564) – biztonságos, gyors, és nem okoz kényelmetlenséget a vizsgálati személy számára. Ennek megfelelően egyre gyakrabban alkalmazzák a fonetikai kutatásban, különösen az elmúlt húsz évben (Kochetov, 2020). Az artikulációs kutatások módszertanának nemzetközi nép-



1. ábra. Csapó Tamás Gábor az ultrahangfejet rögzítő sisakban (fent, a sisak az Articulate Instruments Ltd. terméke) és az Articulate Assistant Advanced (AAA) szoftver képernyőképe (lent) (Csapó et al., 2017a: 343; Csapó & Xu, 2020: 3735).

szerűségi listáján a második helyet a nyelvultrahang foglalja el az elektromágneses artikulográfia (EMA) után: a nemzetközi fonetikai folyóiratokban 2000 és 2019 között megjelent publikációkban az EMA a tanulmányok 29%-ában, az ultrahang 17%-ukban szerepelt (vö. Kochetov, 2020).

A nyelvultrahang előnye, hogy a teljes nyelvfelszín működéséről képet ad, nem csak a nyelv (és más beszédszervek) néhány meghatározott pontjáról (mint az EMA). Egyszerűen használható, elérhető árú, valamint nagy felbontású (akár 800×600 pixel) és nagy sebességű (akár 100 képkocka/s) felvétel készíthető a

segítségével (Csapó et al., 2017a). A jó térbeli felbontás azért fontos, hogy a nyelv alakjáról minél pontosabb képet kapjunk, míg a jó időbeli felbontás ahhoz szükséges, hogy a beszédhangok képzésének gyors változását (pl. zárfelpattanás, koartikuláció) is vizsgálni lehessen. Előnye még, hogy a beszélők döntő többségénél jól alkalmazható, nem okoz különösebb nehézséget a használata, mivel jól tolerálható. Az ultrahang további előnyei közé tartozik az is, hogy mérete és súlya miatt hordozható, így például a gyermekkorú vagy a más sérülékeny populációba tartozó beszélőket nem szükséges fonetikai laboratóriumban vizsgálni, a természetes közegükben is felvehetők a kísérletek. Mindezen túlmenően más műszerekkel, elsősorban elektroglossográf (EGG, más néven laringográf) is kombinálható, így egyszerre rögzíthető és vizsgálható a nyelv artikulációs tevékenysége és a hangszalagműködés.

Mindezzel szemben hátránya az, hogy kizárólag egyetlen beszédszerv, a nyelv mozgását teszi megfigyelhetővé, és meglehetősen érzékeny a beszélő szöveteinek állapotára (pl. zsírosság, hidratáltság), továbbá bizonyos mértékben a testméretekre is (mert a nagyobb üregekben kevésbé ad jó képet, mint a kisebbekben) (vö. pl. Stone, 2005; Csapó et al., 2017a). További nehézséget jelent az ultrahangfej rögzítése a mozgó artikulátoron (állkapocs), valamint az adatkinyerés (a nyelvkontúrok dokumentálása és elemzése).

Artikuláció közben a szájpád nem látható az ultrahangképen, mert az ultrahangsugár visszaverődik a szájúregbeli levegőről, ugyanakkor a nyelvhelyzetet illetően a nyelv és a szájpád távolsága alapvető fontosságú információ lehet. Nyelés során azonban a nyelv szinte teljes mértékben érintkezik a szájpáddal, így a szájpád kontúrja megfigyelhetővé válik (Epstein & Stone, 2005). A szájpád vonala így módon rögzíthető az ultrahangképen, illetve referenciaként használható a nyelvhelyzet vizsgálatában.

További viszonyítási alapot jelent a harapási sík felvétele (Scobbie et al., 2011). A harapási síkot úgy rögzíthetjük, hogy a két fogsor között egy merev lemezt helyezünk el, amelyre a kísérleti személy ráharap. A nyelvet a lemez aljához támasztva a nyelvkontúr kirajzolja a harapási síkot.

A kétdimenziós nyelvultrahang (amelyet a magyar beszéd vizsgálatára eddig használtak) a nyelv szagittális (oldalnézeti keresztmetszeti) középvonalának mozgását rögzíti (Stone, 2010). Az ultrahangfejnek az áll alatt történő pozicionálására sisak használatos (erre látható példa az 1. ábra felső képén). Az ultrahangfelvétellel egyidejűleg a beszédet általában a sisakra csíptetett mikrofonnal veszik fel. A videó és a hang szinkronizált felvételét, a felvételek megjelenítését és az elemzést valamilyen szoftver teszi lehetővé, illetve támogatja. Az 1. ábra alsó képén erre látunk példát: az Articulate Assistant Advanced (AAA) szoftver bizonyos ablakaiban egyszerre látszik az ultrahangkép, a beszéd hullámformája, FFT-spektruma és spektrogramja. Emellett az AAA szoftver az ábra alján látható módon automatikus nyelvkontúrkövetésre is alkalmas, tehát (valamennyi hibával) képes az ultrahangképeken automatikusan letapogatni a nyelv felszínének körvonalát.

Ugyan az ultrahang láthatóvá teszi a nyelv felszíni kontúráját (1. ábra alsó panel), de ez a kontúr közvetlenül még nem elemezhető. Ahhoz, hogy a nyelv alakot és annak változásait vizsgálni lehessen, a rögzített képsorozatból valamilyen módszerrel ki kell nyerni a nyelv körvonalát (illetve ennek adatait). Ez kivitelezhető manuálisan: ennek során a képernyőn megjelenő ultrahangképre egy célszoftver segítségével az egér mozgásával berajzoljuk a nyelv kontúráját. Ez azonban rendkívül időigényes. A másik lehetőség, hogy automatikus módszerekkel kíséreljük meg a nyelv körvonalának azonosítását (vö. az 1. ábra alsó képsorán a piros vonalakat). Ezek pontossága azonban még nem éri el a manuális kontúrkinyerését, ezért mindenképpen szükség van emberi ellenőrzésre és javításra, ha automatikus módszereket alkalmazunk (Whalen et al., 2019).

3. A magyar beszéd vizsgálata ultrahangos technikával: rövid történeti áttekintés

A magyar nyelvre vonatkozóan az ultrahang hazai megjelenését megelőzően kevés olyan vizsgálat született a lingvális artikulációról, amely dinamikus adatokon (azaz nem csak statikus állóképeken) alapult, és amely valós időben rögzí-

tette a beszédszervek mozgását (nem pedig utólag dokumentálta a lenyomatot, mint például a fotópalatográfia). Lotz János az 1960-as években (1966), Szentde Tamás az 1970-es években (1974), majd Bolla Kálmán az 1980-as években (1981a; 1981b) röntgenfilm (ún. röntgenogram/kinoröntgenografikus vizsgálat) technológiával vizsgálta a nyelvgesztusokat. A kétezres évek elején elektromágneses artikulográfiával készültek különféle vizsgálatok külföldi laboratóriumokban (Mády, 2008; Deme et al., 2016, 2017).

Habár a nemzetközi szakirodalomban a nyelvultrahang alkalmazása az artikulációs kutatásokban az 1960-as évektől dokumentálható (Kelsey et al., 1969), az általunk ismert első olyan ultrahangos vizsgálat, amelyet a magyar beszédre vonatkozóan végeztek el, Stefan Beňuš és Adamantios I. Gafos 2007-ben publikált kísérlete, amelyet a Haskins Laboratóriumban elektromágneses artikulográf (3 beszélő) és nyelvultrahang (1 beszélő) segítségével rögzítettek. A kutatók a magyar nyelv egy sajátos jelenségét, a magánhangzó-harmóniát, illetve azon belül az áttetsző magánhangzók artikulációs jellemzőit tanulmányozták. Noha a kísérlet anyagával kapcsolatban felmerülnek kérdések (vö. Markó et al., 2019a), Beňuš és Gafos (2007) arról számolnak be, hogy az eredményeik szerint az anti-harmonikusan viselkedő magánhangzók esetén hátulsóbb nyelvhelyzet volt mérhető, mint a harmonikusan viselkedők esetében.

A hazai tudományban a nyelvultrahangos vizsgálatok 2016-tól kezdődtek meg, amikor az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport pályázati támogatása lehetővé tette nagyobb értékű artikulációs műszerek beszerzését. Csapó Tamás Gábor a kutatócsoport alapító tagjaként jelentős részt vállalt a nyelvultrahangos technika itthoni implementálásában. Az első kísérletek nagy része módszertani-feltáró jellegű volt (ezek többségét nem publikálták, de később lásd pl. Csapó & Xu, 2020; Csapó et al., 2021), majd a folyamatos magyar beszéd olyan jelenségeire összpontosítottak (a teljesség igénye nélkül), mint például a gemináta és szingleton zár- és réshangok artikulációs eltérései a nyelvhelyzet tekintetében (Percival et al., 2020a,b,c); a zöngétlen szibilánsok artikulációs mintázatának longitudinális változása gyermekeknél (Grácsi et al.,

2021a); a zöngés és zöngétlen obstruensek ejtési sajátosságainak eltérései (Grácsi et al., 2020a,b, 2021b).

A kezdetektől fogva zajlottak kísérletek az UH-adatok alapján beszédtechnológiai alkalmazások fejlesztésére is (pl. Csapó et al., 2020), majd Csapó Tamás Gábor több egyéni és csoportos pályázati projektje is ezt tűzte ki célul, elsősorban az ún. némabeszéd-interfész témakörében: 2017–2021 NKFIH FK-17 (124584) „Artikulációs mozgás alapú beszédgenerálás”; 2018–2021 NKFIH PD-18 (127915) „Némabeszéd-interfészek kidolgozása és a beszélőfüggőség vizsgálata nyelvultrahang és elektromágneses artikulográf eszközökkel”; 2022–2026 NKFIH FK-22 (142163) „Az artikuláció és az agyi jelek elemzése beszéd alapú agy-gép interfészhez”.

A Csapó Tamás Gábor által az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport laboratóriumában rögzített ultrahangfelvételek alapján a Miskolci Egyetem és a Debreceni Egyetem kutatói is végeztek méréseket, összehasonlításokat az MRI- és az ultrahangképek harmonizációjára, összekapcsolására (lásd pl. Trencsényi, 2020b,a; Trencsényi & Czap, 2021).

4. Módszertani kérdések a nyelvultrahang használata és az adatfeldolgozás kapcsán

4.1. A nyelvultrahang alkalmazása kísérleti helyzetben

Tekintettel arra, hogy az ultrahangkészülék beszerzésekor az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport tagjainak nem volt még tapasztalata a technika használatával kapcsolatban, és korábban Csapó Tamás Gábor is csak tipikus felnőtt beszélőkkel és az ultrahanggal magában (más műszerrel való összekapcsolás nélkül) végzett kísérleteket, kezdetben számos megvalósíthatósági tanulmány elkészítésére volt szükség. Csapó Tamás Gábor vezetésével zajlottak próbakísérletek a gyermeki artikuláció vizsgálatára vonatkozóan (ennek konklúzióit lásd Markó et al., 2017); az ultrahang és az elektrolottográf (EGG) összekapcsolásával (ennek tapasztalatait a későbbi publikációkban használták fel, pl. Percival et al., 2020a); illetve az ultrahang és az EGG kapcsolt

alkalmazásával egy fiatalkori Parkinson-kórban szenvedő beteg beszédének elemzésére is sor került (Grácz et al., 2018).

4.2. A nyelvkontúr kinyerése

A kétdimenziós ultrahangfelvételek szürkeárnyaltos képsorozatok, amelyeken a nyelv felszíni kontúrja a környezetéhez képest nagyobb fényerősségű vonalként jelenik meg. Ez azonban önmagában nem alkalmas a kvantitatív elemzésre. Ahhoz, hogy a nyelv kontúrján elemzéseket lehessen végezni (például összehasonlítani a nyelvformákat, méréseket végezni a nyelv görbületére vonatkozóan), a nyelvkontúrt (pontosabban annak számszerű adatait, illetve az egyes pontok koordinátáit) célszoftver segítségével ki kell nyerni. Mint említettük, a manuális kontúrrajzolás, bár pontos, rendkívül időigényes, az automatikus kontúrkiyérés ugyanakkor gyors, azonban hibák torzítják. A kézi kontúrrajzolás és a különféle automatikus módszertanok összehasonlítása, hibaelemzése és ennek révén a gépi technikák továbbfejlesztése a kezdetektől foglalkoztatja a kutatókat. Csapó Tamás Gábor egyik első, az ultrahangos technikára vonatkozó, Steven Lulichsal végzett kutatása is erre a kérdéskörre irányult (Csapó & Lulich, 2015). A vizsgálat célja a kétdimenziós ultrahangos képsorozatokból történő midszagittális (középmetszeti) nyelvkontúrok meghatározásakor fellépő hibák jellemzése volt. A kísérletben hét személy – két szakértő és öt egyetemi hallgató – végezte el négy beszélő ejtéséből származó, összesen 1145 nyelvultrahangkép kézi berajzolását, és a kontúrokhoz kapcsolódó variabilitást számszerűsítették a kutatók. Ugyanezekből a felvételekből automatikusan is nyertek ki nyelvkontúrokat az EdgeTrak (Li et al., 2005), a TongueTrack (Tang et al., 2012) és az AutoTrace (Fasel & Berry, 2010) algoritmusok segítségével, majd ezeket kvantitatív módszerrel összehasonlították a kézíleg berajzolt nyelvkontúrokkal. Az eredmények szerint az ultrahangkép minősége nagyobb mértékben befolyásolta a kézi kontúrrajzolás pontosságát, mint a kontúrt berajzoló személy gyakorlottsága, szakértelme. Emellett a kézi kontúrok pontosabbak voltak a gépi kontúroknál, illetve az algoritmusok közül az AutoTrace 3.5 bizonyult a legpontosabbnak. A szerzők három alapvető hibafajtát azonosítottak: 1. a program képtelen nyo-

mon követni a gyorsan mozgó nyelv kontúrját; 2. a zajból adódó hibák, amelyek a kép rossz minőségére vezethetők vissza; valamint 3. amikor a nyelv egyes részeit nem követi le a program. Ezek a hibatípusok jelzik a javításra szoruló területeket a jövőbeli fejlesztések során. A hibamértékek azonosítására további tesztek is készültek (vö. pl. Csapó & Csopor, 2015).

4.3. A nyelvkontúr adatainak feldolgozása

Csapó Tamás Gábor többféle módszerrel kísérletezett a nyelvkontúradatok számszerűsítése érdekében. A legközelebbi szomszéd távolsága (nearest neighbour distance, NND) módszerrel a nyelvkontúrok közötti távolság számszerűsíthető (Zharkova & Hewlett, 2009) a következőképpen: U -val és V -vel jelöljük a két nyelvkontúrt, illetve $U = u_1, u_2, \dots, u_n$ és $V = v_1, v_2, \dots, v_m$ a nyelvkontúrok pontjai. Első lépésként sorban meghatározzuk mindegyik u_i ponthoz a hozzá legközelebbi v_j pont távolságát; majd sorban meghatározzuk mindegyik v_i ponthoz a hozzá legközelebbi u_j pont távolságát; végül kiszámítjuk az összes fenti módon számított távolság átlagértékét. A legközelebbi szomszéd távolság számítási módjából következik, hogy amennyiben a két nyelvkontúr látszatra közel van egymáshoz, és hasonló hosszúságúak, akkor a köztük lévő távolság kicsi lesz; illetve amennyiben a két nyelvkontúr látszatra távol van egymástól, és hasonló hosszúságúak, akkor a távolságuk nagy lesz. Ugyanakkor ha a két nyelvkontúr látszatra ugyan közel van egymáshoz, de nem egyforma hosszúságú, akkor ott a számított távolság ugyancsak viszonylag nagy lesz. Csapó Tamás Gábor ezt a módszertant alkalmazta például a gyermeki artikuláció vizsgálatában (pl. Markó et al., 2017, 2019b).

Más kutatásokban (pl. Grácsi et al., 2021a) a nyelvkontúrok összehasonlításának alapjául Csapó Tamás Gábor a LOC_{a-i} indexet (Zharkova et al., 2015) alkalmazta. Ebben a módszerben a nyelvkontúr két szélső pontját összekötik egy egyenessel, majd az egyenes harmadolópontjaira merőlegest illesztenek, és kiszámítják a nyelvkontúr és az egyenes közötti távolságot ennél a két harmadolópontnál. Az index értéke ezeknek a távolságoknak az aránya, és ezzel számszerűen jellemezhető a nyelvkontúr.

Egy másik fajta, a nyelvkontúrok összehasonlítására alkalmas módszertant fejlesztett ki Csapó Tamás Gábor és Maeda Percival a magyar szingleton és gemináta mássalhangzók összehasonlítására (lásd 5.1; Percival et al., 2020a,b,c). Ebben a módszerben a nyelvkontúrokat a beszélő harapási síkja által definiált koordináta-rendszerre forgatták, majd koronális, veláris és faringális régiókra osztották őket (Recasens & Rodríguez, 2016 alapján). Minden régió esetében a poláris koordinátákat lineáris kevert regressziós modellekkel elemezték annak meghatározására, hogy az ultrahangfej és a nyelv felszíne közötti radiális (sugárirányú) távolság mely régióban tér el a mássalhangzók között (tükrözve a nyelv alakjának és helyzetének különbségeit).

Felmerül a kérdés, hogy a nyelvkontúrok összehasonlítására mely módszertanok a legalkalmasabbak, illetve melyik módszer melyik kutatási kérdés(ek) megválaszolására a legalkalmasabb. A különböző szakirodalmi források különféle mérési módszertanokat alkalmaztak, így ha a saját eredményeinket egy adott tanulmány eredményeivel tervezzük összevetni, hasonló módszert is kell alkalmaznunk. A legközelebbi szomszédok közötti távolság becslésénél, mint említettük, fontos szempont, hogy a két nyelvkontúr hasonló hosszúságú-e. Lényeges az is, hogy a teljes nyelvkontúr elemezhető legyen, amit azonban az alsó állkapocs és a nyelvcsont árnyéka befolyásolhat. A szibilánsok elsajátításának vizsgálatában Zharkova és munkatársai (2015) a nyelvgörbületet számszerűsítették: a LOC_{a-i} és más hasonló görbületmérőszámok a nyelvformáról közvetetten, egy számadattal adnak becslést. Mivel ebben az eljárásban a nyelv legelülső és leghátsó (látható) pontját kötjük össze egy egyenessel, és az erre az egyenesre állított merőlegesekkel dolgozunk, az eredmény ismét nagymértékben függ attól, hogy mekkora rész látszik a nyelvből. A Lingvális Artikuláció Kutatócsoport mindezek miatt azokban az esetekben, amikor nem egy-egy tanulmánnyal/szerző eredményével összevetve, hanem általánosságban kíván vizsgálni egy jelenséget a teljes nyelvkontúrok összehasonlításával, akkor az SS-ANOVA vagy a GAMM (általánosított kevert lineáris modellek) statisztikai elemzési módszertanokat alkalmazza, mivel az ezek által adott eredmények pontosabban képezik le a valószínűségeket. Ezekben a módszerekben több nyelvkontúrra statisztikai eljárással gör-

bét illesztünk, amely egy becsült nyelvkontúrt eredményez a kérdéses jelenség leírására. Ezek esetében is célszerű a valószínűtlenül rövid nyelvkontúrokat elhagyni, ugyanakkor a statisztikai modellezés során a kisebb bizonytalanságok kellő számú adat esetében kisebb eséllyel okoznak téves eredményt. Mindezekon túl azonban további módszerek is léteznek, valamint hangsúlyozzuk, hogy ezek az általunk vizsgált kérdésekre legjobb módszertanok, nem végeztünk általános összehasonlító kutatást.

4.4. A beszélők és a felvételek eltéréséből adódó variancia kezelése az adatokban

Mint láttuk, az ultrahangfelvételek számos okból variábilisak, és ezek közül csak egy az artikulációra inherensen jellemző változatosság. Magából az ultrahangos módszertanból is következik, hogy az ultrahangfej helyzete még rögzítők használata esetén is eltér két felvétel között, de egy felvétel közben is változhat egyazon beszélőn belül is. Az ultrahangból adódó további variabilitási tényező a felvételek minősége és ezáltal a nyelvkontúrok minősége (például hossza). A beszélők közötti változatosság szintén közhelynek számít a fonetikában. Mindezek a tényezők korrumpálhatják a nyelvkontúrok számszerűsítésére használatos mérőszámok (lásd például 4.3) érvényességét is.

Máig nem áll rendelkezésre olyan módszer, amely megoldaná ezeket a problémákat, azaz a különböző alkalmakkor és/vagy különböző beszélőkkel rögzített ultrahangfelvételek tekintetében lehetővé tenné a normalizálást, az illesztést vagy az adaptálást. Csapó Tamás Gábor Kele Xuval dolgozott olyan módszereken, amelyek segíthetik e problémák megoldását. 2020-ban publikált tanulmányukban mindenekelőtt az ultrahangfej pozíciójának eltéréséből adódó hibák kvantifikálására tettek kísérletet, és ennek kapcsán a harapási sík felvételét látták megoldásnak. Meglátásuk szerint a harapási sík alapján egy referencia koordináta-rendszert kell felvenni (vö. Scobbie et al., 2011), amelyhez a nyelvkontúrok a megfelelő szögben illeszthetők (beforgathatók), és így összehasonlíthatóvá válnak. Ennek a módszernek kétségkívül hátránya az, hogy a nyelvkontúrok kinyerése mellett további, nagyrészt csak manuálisan elvégezhető munkával jár, a tanulmány írásakor ugyanis erre nem állt rendelkezésre automa-

tikus módszertan. További nehézséget jelent véleményünk szerint az, hogy az ultrahangfej felvétel közbeni elmozdulására nem jelent megoldást, ha a felvétel készítője nem észleli ezt azonnal, és nem rögzíti újra a harapási síkot, amely alapján a referencia koordináta-rendszer újból létrehozható.

A témában további elemzések is készültek (Csapó et al., 2021), és Csapó Tamás Gábor tervezte a jövőben egy automatikus osztályozó kifejlesztését, amely a nyelvkontúr elemzése során figyelmeztet, ha az ultrahangfej helyzete eltérő egy felvételen belül; illetve amely konfidenciaintervallumokat rendel az adatok megbízhatóságához. Úgy gondolta, hogy a nyelv semleges pozícióban való felvétele (az egyes megnyilatkozások elején vagy végén), és e referenciakép változásának a regisztrálása segítség lenne a felvétel közben adódó problémák automatikus észleléséhez is.

5. Alap kutatások: a nyelv artikulációs működésének tanulmányozása

Az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport Csapó Tamás Gábornak a nyelvultrahanggal kapcsolatos szakértelme révén több olyan alap kutatást is végzett, illetve kezdett meg, amelyek hozzáférhetővé tettek a magyar nyelvvel és beszéddel kapcsolatban korábban nem ismert adatokat, illetve bővítették a tudásunkat. Ezek a kutatások magánhangzók és mássalhangzók artikulációjára, valamint az ezek kapcsolatában jelentkező koartikulációs jelenségekre irányultak, nemcsak felnőtt beszélők, hanem gyermekek esetében is.

Az ultrahangfelvételeken alapuló alább bemutatandó vizsgálatokban Csapó Tamás Gábornak meghatározó szerepe volt számos tekintetben. Egyfelől fontos kiemelni, hogy ezek a kutatások a Csapó Tamás Gábor által kiépített szakmai kapcsolatokon (és erre épülő kutatói mobilitáson) alapultak. Továbbá az is hangsúlyozandó, hogy ezeknek a kutatásoknak az ultrahangos módszertanában is újszerű megoldásokat eredményezett Maeda Percival és Csapó Tamás Gábor közös munkája (5.1): közösen dolgozták ki az elemzés technikáját, és Magyarországon, illetve a magyar nyelvre irányuló elemzésben elsőként alkalmaztak EGG és UH eszközöket összekapcsoltan adatfelvétellel. A további kutatásokban (5.2,

5.3) is ehhez hasonlóan a Csapó Tamás Gáborral közös munka eredménye a felvételi és elemzési módszertanok tesztelése, összevetése, majd az adekvát eljárás kiválasztása és alkalmazása (vö. még 4.3).

5.1. *A nyelv működése a gemináta és szingleton zár- és réshangok artikulációjában*

A magyar nyelvben a szingleton-gemináta oppozíció, azaz a nyelvileg rövid és hosszú mássalhangzók kontrasztja fonológiai, azaz jelalak-megkülönböztető szerepű szembenállás, pl. *ép* vs. *épp*, *tol* vs. *toll*, *ara* vs. *arra*. A leírások szerint a magyar mássalhangzórendszerben a szingleton-gemináta mássalhangzó-oppozíció legfőbb akusztikus kulcsa az ejtés időtartama. Emellett hagyományosan azt is feltételezik, hogy gemináták csak bizonyos pozíciókban, csak intervokálisan, illetve megnyilatkozás végén (magánhangzó után és szünet előtt) fordulhatnak elő a magyarban. Így nem állhatnak például szó elején, illetve nem fordulnak elő egy további mássalhangzóval szomszédosan. Amennyiben (például toldalékolás vagy szóképzés miatt) mégis előállna az utóbbi eset, a gemináta mássalhangzó (a legtöbb esetben) rövidül, azaz degeminálódik (Siptár & Törkenczy, 2000/2007).

A gemináták és a degemináció fonetikai vonatkozásait korábban már több idegen nyelvben is vizsgálták, és a magyarra vonatkozóan is napvilágot látott néhány, az akusztikai szerkezetre fókuszáló elemzés. Ezek fő kérdése jellemzően az volt, hogy mi különbözteti meg a gemináta és szingleton mássalhangzókat a hangok akusztikai megvalósításában. A kérdést felpattanó zárhangok és zár-rés hangok körében elemezték, és azt találták, hogy csakúgy, mint más nyelvekben (ehhez vö. pl. Ridouane, 2007), a magyarban is az időtartam, a felpattanó zárhangok esetében pedig különösen a zárképzés időtartama a fonológiai szingleton-gemináta oppozíció legfontosabb akusztikai korrelátuma (Olaszy, 2006; Pycha, 2009, 2010; Neuberger, 2015; Siptár & Grácsi, 2014).

Az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport megalakulásával és az új artikulációs eszközök beszerzésével újabb kérdések nyíltak, amelyekre a csoport vizsgálatokat tervezett. Egyrészt elemezték azt, hogy a gemi-

náta felpattanó zárhangok rövidüléséből (degeminálódásából) előálló rövid hang ejtésében megegyezik-e a további (elvben azonos minőségű) szingleton hangokkal, különös tekintettel az artikulációs mozdulatok időbeli szerveződésére, időzítésére (ennek elemzésére EMA-t alkalmaztak az akusztikai elemzés mellett, vö. Deme et al., 2019b,a). Másrészt vizsgálták, hogy a gemináta felpattanó zárhangok és réshangok a szingletonokhoz képest nagyobb artikulációs erőfeszítéssel képzettek-e, amelyekben a nyelv inkább eléri az artikulációs célt – ez a vizsgálat ultrahanggal készült (Percival et al., 2020b,c). Harmadrészt pedig elemezték azt is, hogy hogyan alakul a gemináta felpattanó zárhangok és réshangok zöngéssége, tekintettel arra, hogy a zöngé fenntartása nagyobb időtartamban artikulációsan nehéz – ennek vizsgálatára ultrahangot és elektroglottográfiát is használtak összekapcsolva (Percival et al., 2020a).

A hosszú, degeminálódó hosszú, intervokális rövid és más rövid mássalhangzó mellett álló mássalhangzók EMA-n alapuló vizsgálata azt mutatta, hogy a rövidülő gemináták és a rövid mássalhangzók artikulációs és akusztikai tekintetben is eltérnek: a degemináció folyamata nem a megfelelő intervokális szingleton időtartamára redukálta a rövidülő gemináta mássalhangzókat, hanem sokkal inkább azokhoz a szingletonokhoz tette őket analóggá, amelyek a degeminációs pozícióhoz hasonlóan jobbról egy eltérő képzési helyű mássalhangzóval szegélyezettek, azaz kétféle kapcsolat tagjai. Az artikulációs adatok továbbá azt is felfedték, hogy a rövidülő gemináták és a két rövid mássalhangzóból álló kapcsolatok az említett időtartam-paraméterek tekintetében egy, a szingletonok és a gemináták közti átmeneti kategóriát képeznek. A zárképzés artikulációs korrelátumaként értelmezett artikulációsgesztus-platók átfedése a hangkapcsolatokban ezeken túlmenően azt is megmutatta, hogy a rövidülő gemináták és a szingletonkapcsolatok csak a lingvális-labiális képzéshelyi sorrend esetén térnek el időzítésüket tekintve (azaz a /tp/ gesztusok időzítése nem volt azonos a /ttp/ gesztusokéval, mégpedig olyan módon, hogy a /ttp/-ben megjelenő zárképzési szakaszok közti késleltetés nagyobb volt, mint a /tp/ esetében). A labiális-lingvális (azaz a /pt/ és /ppt/) kapcsolat gesztusidőzítése ezzel szemben nagyon hasonló volt, és átfedést mutatott a zárképzési szakaszokat mutató

artikulációs „platók” között. Ezt úgy értelmezhetjük, hogy a gemináták rövidülésének, azaz a degeminációnak a folyamata a szomszédos mássalhangzók képzési helyének függvényében rövidíti a hosszú mássalhangzókat kapcsolatbeli rövid mássalhangzókká – „hatékonyabban” a „könnyebb” ejtésű, eleve nagyobb artikulációs átfedést mutató /ttp/, és kevésbé hatékonyan a „nehezebben” ejtendő /ppt/ esetében. A fentebbi összefüggéseknek a leírása nem csak a magyar nyelv vonatkozásában számít újszerűnek. Végül azt is megállapították, hogy a gemináták és a két szingletonból álló mássalhangzó-kapcsolatok előtt hosszabb a magánhangzók időtartama, ami abból következik, hogy ezekben az esetekben a nyelv lassabban emelkedik a mássalhangzó zárképzéséhez, mint a szingletonok esetében (Deme et al., 2019b,a). Ez az eredmény megcáfolta a más nyelvekre, kisebb beszélőszám elemzésével kapott adatok alapján megfogalmazott feltételezéseket.

Ultrahanggal kerestek választ arra a kérdésre, hogy hogyan alakul az artikulációs célok megközelítése a magyar szingleton és gemináta mássalhangzóknál. Más nyelvekre, így például a japánra, a koreaira, az olaszra és a (keleti) oromóra kapott korábbi eredmények ugyanis azt mutatták, hogy a geminátákban nagyobb a kontaktusfelszín, magasabb és laposabb a nyelv, és előrébb van a nyelvtest, tehát összességében a nyelv jobban eléri a feltételezett artikulációs célját, mint a szingletonokban (Payne, 2006; Kochetov & Kang, 2017; Percival et al., 2018). A magyar nyelvre kapott adatok jelentősen árnyalták az addig kialakított képet: a zöngés és zöngétlen bilabiális, alveoláris és veláris felpattanó és réses képzésű szingletonok és gemináták elemzése azt mutatta, hogy nem egyértelmű a nyelv előrébb tolulása, tehát az artikulációs cél jobb megközelítése a geminátákban. Ez pedig arra utal, hogy nyelvfüggő lehet az, hogy a gemináták létrehozása fortizációként, azaz erősítésként jelentkezik a szingleton megfelelőikhez képest, hiszen kétféle következtetés vonható le mindebből. Vagy arról van szó, hogy a korábban elemzett nyelvekben erősebb a fortizáció a geminátákban és/vagy a lenizáció (az artikuláció lazulása) a szingletonokban, mint a magyarban, vagy pedig arról, hogy a gemináták fortizációja a magyarban más

természetű, mint a további elemzett nyelvekben, és ez más artikulációs mérőeszközökkel lehet inkább feltárható (Percival et al., 2020b,c).

A világ nyelveiben viszonylag ritkák a zöngés gemináta mássalhangzók, valószínűleg azért, mert a zöngésség fenntartása hosszabb időtartamban nehezebb. Ennek megfelelően egyes eredmények a zöngés gemináták részleges zöngétlenedésére utalnak olyan nyelvekben (pl. a tokiói japánban), ahol ezek a ritkább hangok egyáltalán előfordulnak (Kawahara, 2015). A magyarban léteznek zöngés gemináták, amelyek közül a bilabiális, alveoláris és veláris felpattanó és réses képzősűket elemezték, illetve ezek szingleton párjait, valamint mindezek zöngétlen megfelelőit. Az EGG-adatok elemzése szerint nem volt statisztikailag kimutatható hatása a zöngésség időarányára annak, hogy az adott hang gemináta vagy szingleton-e, tehát a vizsgálat nem támasztotta alá azt a feltevést, hogy a zöngés gemináták részben vagy egészen zöngésednének vagy teljes mértékben zöngések lennének (fonetikai tekintetben). Ugyanakkor a zöngés hangok jelentős (és a zöngétlenekben látottnál nagyobb) változatossággal valósultak meg ebben a tekintetben. A nyelv helyzetét leképező UH-adatok pedig azt mutatták, hogy a beszédhangok fonológiai zöngéssége nem jelzi előre a nyelv előre mozdulását, sokkal inkább a fonetikai zöngésség. Ebből a szerzők arra következtettek, hogy a magyarban nem valószínűsíthető, hogy a nyelvgyök előre tolása a zöngésség fenntartásának előre tervezett mozzanata lenne – de ennek a feltevésnek az alátámasztása még további vizsgálatokat igényel (ehhez vö. 5.3).

5.2. *Gyermekek lingvális artikulációjának vizsgálata*

Az anyanyelv-elsajátítás vizsgálata a magyar gyermekek körében régóta népszerű kutatási téma, ugyanakkor megfelelő műszerek híján az artikulációs elemzések az említett kutatócsoport megalakulása előtt teljesen hiányoztak. Mivel a nyelvultrahang a nemzetközi kutatások alapján jól alkalmazható a gyermeki artikuláció regisztrálására is (vö. pl. Zharkova et al., 2015, 2018), a magyar gyermekek ejtéséről is készültek elemzések ezzel a módszerrel: vizsgálták a magánhangzók koartikulációs jellemzőit és a zöngétlen réshangok ejtésének longitudinális változását egyazon beszélőn belül.

Az artikulációs csatorna születéstől felnőttkorig jelentős változásokon megy keresztül, mind méreteit, mind arányait, mind pedig az egyes beszédképző szervek működési precizitását és összehangoltságát tekintve. A különböző életkori szakaszokban az érintett szervek méretbeli változása eltérő ütemű (Seikel et al., 2010), és mindezek a változások természetesen hatnak a toldalékcso rezonatorsajátosságaira is. Az egyes képzőszervek koordinációjának fejlődése sem egyenletes: például az állkapocsgesztusok korábban érnek (azaz válnak hasonlóbbá a felnőtt működéshez, a gesztus téri-idői kivitelezéséhez), mint az ajakmozgás, és még később rögzülnek a nyelvgesztusok. Ez utóbbinak egyrészt az az oka, hogy a nyelvnek mint beszéd szervnek a beszédbeli kontrollálása komplexebb feladat; másrészt pedig az, hogy a nyelv mozgása beszéd közben nem látható, így a nyelvmozgás felnőtt mintájának imitálása is nehezebb a gyermekek számára (Goffman, 2015). Habár a gyermekek úgynevezett fonetikai jártassága, a beszédhangoknak az elvárásoknak megfelelő ejtésére való képesség 8 éves korra kialakul, a finommotorikus ügyesség még később is jelentősen fejlődik, és ez a folyamat tizenéves korra is kitolódik (Vorperian & Kent, 2007). A beszéd motoros rendszerének érésével az orofaciális struktúrák közötti dinamikus koordináció egyre konzisztensebbé válik, ennél fogva az artikulációs gesztusok egyre kevésbé variábilisak az életkor előrehaladtával (egészséges fejlődés esetén) (vö. Terband et al., 2011).

Tekintettel arra, hogy a magánhangzók nem akadályhangok, beszéd közben nem kapunk jól megfigyelhető taktilis visszajelzéseket a magánhangzók létrehozását eredményező közelítés vagy szűkület létrejöttének helyéről, azaz a magánhangzó képzéshelyéről. Egyes magánhangzók képzésekor a nyelv érintkezhet a nem mozgatható artikulátorokkal – például a nyelvperem a felső nyelvállású magánhangzók esetében hozzáér a felső fogsorhoz (vö. Stevens, 1998) – és ezek a szenzoros visszajelzések feltehetőleg hozzájárulnak ahhoz, hogy a beszélő az egyes magánhangzók ejtéséhez szükséges artikulációs beállításokat pontosan eltalálja, ez azonban magánhangzófüggő. Emellett a magánhangzók izolált ejtése statikus, így kinezetikus információk sem segítik a magánhangzók képzéshelyének azonosítását, miközben a nyelv igen nagy variabilitással mozgat-

ható és alakítható (Ashby, 2011). A magánhangzók ejtése rendkívül variábilis is (egyes magánhangzóké nagyobb, másoké kisebb mértékben, bizonyos képzési és akár nyelvfüggő sajátosságokkal összefüggésben, vö. Deme, megj.), hiszen a nyelv alakjának kontrollálása (a nyelv izmainak komplex szerkezete és a motoros ügyesség szükséges foka miatt) igen nehéz feladat.

A gyermekek koartikulációjával számos nemzetközi tanulmány foglalkozik, magyarra azonban kevés adatunk van. Artikulációs szempontú elemzés pedig kettő, Markó és munkatársai (2019b; 2019c) vizsgálatából. A gyermekek beszédelsajátításában az egyik kulcsfontosságú feladatnak egyes szerzők annak a megtanulását tartják, hogy egy adott szegmentum létrehozásához szükséges artikulációs gesztusokat együttesen, egyre inkább szinkronban hozzák létre a gyermekek – a téri-idői átfedés pontos mértéke teszi az artikulációt felnőttsterrűvé (Nitttrouer et al., 1996). Általánosságban a gyermekek ejtésében erőteljesebb koartikulációs hatásokat adatoltak, mint a felnőtteknél (Terband et al., 2011). A magyar vizsgálatokban a 9 magyar magánhangzó-minőséget elemezték VCVCVC szerkezetben; a résztvevő gyermekek életkora 7 év 10 hónap és 14 év 8 hónap között szóródott. A két vizsgálatban kétféle elemzés történt, az elsőben az NND-t (a legkisebb szomszéd távolságot) határozták meg, azaz az azonos logatomok ismétlései között a nyelvkontúrok hasonlóságát elemezték. A második elemzésben pedig a nyelv és a szájpád közötti legkisebb távolságot számították ki. Az eredmények alapján általános tendenciaként az életkor előrehaladtával csökkent az ejtési variabilitás, ugyanakkor jelentősek voltak az egyéni különbségek. A nyelvkontúrbeli variabilitás nem függött össze az időtartambeli variabilitással a gyermekeknél.

A nemzetközi irodalomban két zöngétlen szibiláns, az /s/ és az /ʃ/ gyermeki artikulációját vizsgálta Zharkova több publikációban. Bár hasonló nyelvmamintázatot kapott 10–12 éves gyermekek ejtésében, mint a felnőtteknél (Zharkova, 2016), mégis még serdülőkorban is talált artikulációs eltéréseket a felnőtt beszélők produkciójától (Zharkova et al., 2018). Emellett a koartikulációs rezisztencia (a hangkörnyezet hatásának való ellenállás) és az akusztikai eredmények eltérőek voltak a felnőtt ejtéstől (Zharkova et al., 2011, 2012). Ezen

eredményekre alapozva egy longitudinális esettanulmányban elemezték két magyar anyanyelvű testvér /s/- és /f/-ejtését nyelvultrahang segítségével. A felvételek négy időpontban készültek, a lány 7;5, 7;11, 8;5, 8;11 éves és a fiú 11;0, 11;6, 12;0, 12;5 éves korában. A gyermekek nyelvkontúradatait az édesanyjuk (43 év) ejtésével vetették össze (Grácz et al., 2021a). Az artikuláció számszerűsítésére az LOC_{a-i} módszert választották Zharkova és munkatársai (2015) alapján (lásd 4.3). A felnőtt nő ejtésében a LOC_{a-i} eltért a két mássalhangzó között. A lánygyermek ejtésében mind a négy időpontban, míg a fiúgyermek ejtésében az utolsó felvétel idejében találtak szignifikáns eltérést a két mássalhangzó között. Mivel a lánygyermek fiatalabb volt a felvételek készültkor, így fordított mintázatot várnánk. Ugyanakkor a beszédképző szervek nem egyenletes és egymással sem azonos ütemű növekedése (Vorperian et al., 2009), illetve az ehhez való folyamatos artikulációs alkalmazkodás miatt is előfordulhatott, hogy az idősebb gyermeknél egy ideig kevésbé jelent meg a szembenállás a nyelvkontúr (pontosabban annak az LOC_{a-i} indexének a) vonatkozásában, holott a két mássalhangzó megkülönböztetése az ő esetében is stabilan és konzekvensen észlelhető volt. A nyelvkontúrokra kapott eredmények nem mutattak jelentős eltérést a mássalhangzók kezdete, közepe és vége között, tehát végeredményben a mássalhangzó teljes időtartamán hasonló tendenciákat találtak a mássalhangzók elkülönítését illetően.

A gyermekek artikulációjának kapcsán nyelvultrahanggal végzett vizsgálatok eredményei anyanyelvi tantárgymódszertani nézőpontból is szolgáltak hozadékokkal. A közoktatásban 5. és 9. osztályban tananyag a magánhangzók képzése, amelyet részben percepció (vö. magas, mély), részben artikulációs (vö. nyelvállású, ajakkarekítéses, ajakréses) alapú terminusokkal jellemeznek a tankönyvek. Az absztrakt képzési jegyek tanítása – különösen, ha ezeket olyan jegyeknek tartjuk, amelyek külön-külön egy az egyben megfeleltethetők artikulációs mozdulatoknak, képzőszervi helyzeteknek – egészen bizonyosan nehézséget okoz, hiszen az artikulációs és vizuális tapasztalat legalábbis részben ellentmond az absztrakt képzési jegyeknek (Markó et al., 2020). Egy 5. osztályos fiú (a felvétel időpontjában 11;5 éves volt) és egy 9. osztályos fiú (a felvétel

időpontjában 14;8 éves volt) ultrahang-felvételeiről a 9 magyar sztenderd nyelvi magánhangzó-minőséget (VpVpV szekvenciák közepéről) reprezentáló kontúrok alapján az volt leolvasható, hogy a különböző korú beszélők között igen nagy az eltérés a szájpád és a nyelv közötti távolságot tekintve a szájtér és a nyelv méretkülönbségeiből adódóan. A felső nyelvállású magánhangzók esetében jól elkülönült a hátul képzett /u/ és az elől képzett /i/ és /y/ mindkét beszélő esetében, az elől képzettek kontúrja viszonylag fedte egymást. A középső nyelvállásúak esetében mind a nyelv vízszintes, mind a nyelv függőleges helyzete nagyobb variabilitást mutatott, mint ami a tankönyvi kategóriák alapján várható lett volna. Mindkét beszélőnél az látszott, hogy az /o/ és az /e:/ magasabb nyelvhelyzettel képződött, mint az /ø/, valamint hogy a nyelv vízszintes helyzetét tekintve az /ø/ nyelvkontúrja köztes pozíciót foglalt el az /o/ és az /e:/ nyelvkontúrja között. A vízszintes nyelvhelyzetet tekintve az 5. osztályos beszélő esetében az /a:/ nyelvkontúrja köztes helyzetet foglalt el az /ɒ/ és az /ɛ/ nyelvkontúrja között; a 9. osztályos beszélő esetében ez kevésbé látszott jól, mivel a nyelv hátsó része szinte egymást átfedve jelent meg az /ɒ/ és az /a:/ nyelvkontúrján, de a nyelv formáját tekintve az /a:/ esetében az ív a szájtér közepén is magasan volt még, amiből következtetni lehet az /a:/ előrébb képzett voltára. Mindezek alapján artikulációs alapon mind a nyelvállásfok, mind az elől-hátul képzettség kategorizálása bizonytalan, a határok elmosódtak, és kérdéseket vetnek fel. Ha feltételeznénk is, hogy a beszélő pontos fiziológiai visszacsatolással rendelkezik a nyelvhelyzetről a különböző magánhangzók kiejtése közben, és ezeket össze tudja hasonlítani egymással, majd ezen összevetések alapján képes kategóriákat elkülöníteni, ezek a kategóriák a lingvális képzési jegyek esetében korántsem mutatnák az elméleti kategóriák alapján csoportosítható magánhangzó-éjtési mintázatokat. Ehhez hozzáadódik még az életkorral összefüggésben a szájüreg méretbeli (és egyéb morfológiai) eltérése (például a szájpád alakja), valamint az egyének közötti és az egyénen belüli éjtési variancia is, amely ezeknek a kategóriáknak a határait igencsak elmosódottá teszi.

5.3. A zöngés és zöngétlen obstruensek, valamint a megelőző magánhangzók ejtési sajátosságai

A magyar nyelvre vonatkozó ultrahangos vizsgálatok további vonulata a zöngés és zöngétlen obstruensek ejtési sajátosságaira irányul. A fonetikában közismert tény, hogy a fonáció fenntartása szájüregi akadály jelenlétében különféle artikulációs manővereket igényel, mivel a szupralaringális és szájüregi nyomásnövekedés a hangszalagrezgés leállítását okozhatja (vö. pl. Stevens, 1998). Az egyik ilyen artikulációs manőver a nyelvgyök előrébb mozdítása a zöngés szegmentumok ejtésekor a zöngétlen párhoz képest – főként lingvális obstruensek ejtésekor – a garatüregi régió megnagyobbítása érdekében (Narayanan et al., 1995). Egyes tanulmányok szerint a zöngés obstruensekben (különösen a részhangokban) előrébb tolódott a nyelvgyök e mássalhangzók zöngétlen párjához képest, de olyan mássalhangzópárt is találtak, amelyre ez nem volt jellemző (Ahn & Davidson, 2016). A magyarra vonatkozóan eddig a /z/ és az /s/ ejtésének összehasonlítása történt meg, azt a kérdést vetve fel, hogy van-e összefüggés a nyelvgyökhelyzet variabilitása és i) a fonációs arány, illetve ii) a nyelvhegy mozgása között (Grácz et al., 2020a,b, 2021a), illetve, hogy a képzési mód és hely, valamint a kontextus milyen hatással van erre a jelenségre. Az első eredmények (a szibilánsokra vonatkozóan) összefüggést mutattak a mássalhangzók képzési helye, a fonáció beszélőspecifikus aránya és a nyelvgyök pozíciója között (Grácz et al., 2023). Jelenleg is zajlik a nyelvgyök pozíciójának, a mássalhangzó képzési módjának és helyének, valamint a szomszédos magánhangzók artikulációs jellemzőinek kapcsolatát vizsgáló kutatás, kiterjesztve az összes magyar zöngés és zöngétlen obstruensre.

6. Alkalmazott kutatások: a nyelvultrahang-adatok felhasználása a beszédszintézisben

A beszédszintézis, vagyis a mesterséges beszéd előállítása hagyományosan a gépi szövegfelolvasást jelenti (Csapó, 2022), de újabban már nemcsak írott szövegből, hanem artikulációs mozgásból, vagy akár agyi jelekből is kísérletez-

nek beszéd létrehozásával. Artikuláció-akusztikum becslésnek (Articulatory-to-Acoustic Mapping, AAM) nevezzük azokat az eljárásokat, amelyekben a beszéd-szintézis bemenete az emberi artikuláció valamilyen reprezentációja, és a kimenete a hangzó beszéd (Csapó, 2022). A beszéd-szintézis területén a 21. század első két évtizedében a rejtett Markov-modell (Hidden Markov Model, HMM) számított kurrens módszertannak. A 2010-es évek közepétől ennek a helyébe a mély neurális hálózat (Deep Neural Network, DNN) lépett. A DNN magukból a nyers adatokból „tanulja meg”, hogy azok milyen absztrakcióval írhatók le: a betanítás során megtanulja, hogy a bemeneti adatokhoz (a beszéd-szintézis esetén ez lehet például az írott szöveg) milyen kimeneti adatokat (a beszéd-szintézisben ezek a beszédparaméterek) rendeljen (Csapó, 2022).

Az a kérdés, hogy hogyan lehetne gépi úton akusztikumot generálni az artikulációról készült vizsgálati adatok (pl. ultrahangképek) alapján, elsősorban az ún. némabeszéd-interfész (Silent Speech Interface, SSI) rendszerek előállítása kapcsán foglalkoztatják a kutatókat és a fejlesztőket (Denby et al., 2010). Az SSI megvalósításához az artikulációs szervek hangtalan mozgását kell rögzíteni valamely erre alkalmas technikával (ez lehet ultrahang, EMA vagy más alkalmas módszertan is), majd ebből a felvételből a gépi rendszer beszédet szintetizál. Egy ilyen technológia jelentősége abban áll, hogy az eszköz használójának nem kell képesnek lennie arra, hogy hangot adjon ki. A leggyakrabban említett példa a felhasználók körére azok a beszédükben akadályozott emberek, akiknek az artikulációjuk ép, de zöngéképzésre valamilyen okból (pl. a gége teljes vagy részleges eltávolítása miatt) nem képesek. Ugyanakkor szokás hivatkozni a zajos környezetben zajló kommunikációra is, valamint olyan helyzetben történő üzenetátadásra, ahol a beszélőnek csendben kell lennie, de írásban nem tud kommunikálni (néhány hétköznapi helyzet mellett akár hadászati, titkosszolgálati akciók is említhetők itt).

Csapó Tamás Gábor saját egyéni kutatásaiban és az általa vezetett kutatócsoportokban egyaránt a beszéd-szintézis állt a fókuszban, különösen az a kérdés foglalkoztatta, hogy a beszédben akadályozott populáció számára hogyan lehetne jól használható és könnyen hozzáférhető alkalmazást fejleszteni akár az

artikulációs felvételek (UH) felhasználásával, majd később az agyi jelek (EEG) detektálása révén. Csapó Tamás Gábor alkalmazásorientált kutatásainak a teljes körű bemutatására itt nincs módunk, néhány általunk jelentősebbnek ítélt vizsgálatot és kezdeményezést foglalunk össze röviden.

Az első magyar nyelvű ultrahangalapú SSI-kísérletben (Csapó et al., 2017b) egy beszélő közel 200 szinkronizált beszéd- és nyelvultrahang-felvételét használták fel. Az akusztikai jelből kinyerték az alapfrekvencia- és a spektrális paramétereket. Ezután mélyneurálisháló-alapú gépi tanulást alkalmaztak, melynek bemenete a nyelvultrahang volt, kimenete pedig a beszéd spektrális paraméterei. Az eredeti beszédből származó f_0 -paraméterekkel és a gépi tanulás által becsült spektrális paraméterekkel a szerzők mondatokat szintetizáltak, amelyekben szavak, néhol akár teljes mondatok is érthetőek lettek.

Az ultrahangfelvételekből előállított akusztikai szerkezet értelemszerűen (mivel a nyelv a szupraglottális artikuláció egyik főszereplője) elsősorban a spektrális jellemzők becslésére alkalmas, ugyanakkor korábbi kísérletek szerint az alapfrekvencia becslése sem lehetetlen az ultrahanggal rögzített kép segítségével. Ennek háttérében az áll, hogy a zöngéesség összefügg az artikulációs jellemzőkkel (lásd az 5. pontban írtakat). Grósz és munkatársai (2018) erre vonatkozó kísérleteikben két mély neuronhálót tanítottak párhuzamosan, egyet a zöngéesség (zöngés vs. zöngétlen), egyet pedig a konkrét alapfrekvencia-érték és a spektrális paraméterek becslésére. A becsült és az eredeti alapfrekvencia-görbe között 0,74 korrelációs együtthatót kaptak, a szubjektív meghallgatási tesztek során a tesztek résztvevői az eredeti, illetve a prediktált f_0 -görbével szintetizált mondatokat közel egyformán természetesnek minősítették. Az f_0 -ra vonatkozóan további kísérleteket is folytattak, például a Wave-Glow (Prenger et al., 2019) használatával, amely az egyik legfrissebb fejlesztésű neurális vokóder, előnye, hogy relatíve egyszerű, mégis a valós idejűnél gyorsabb szintézist valósít meg. A percepció tesztek alapján négyből három beszélő esetén a Wave-Glow segítségével a korábbiaknál természetesebb hangzású szintetizált beszédet sikerült előállítani az ultrahangfelvételek alapján készült artikuláció-akusztikum konverzióban (Csapó et al., 2020).

Egy tanulmányban az SSI kapcsán azt vizsgálták, hogy a különböző felvételek közötti variabilitás (session dependency) hogyan befolyásolja az artikuláció-akusztikum konverzió sikerességét a nyelvultrahang-alapú beszédszintézisben (Gosztolya et al., 2020). A korábbi kutatási eredmények alapján a szerzők azt várták, hogy ha olyan felvételeket használnak, amelyek között az ultrahangfejet eltávolították, majd újból visszahelyezték a beszélő álla alá, az nagymértékben rontja az előre tanított rendszer pontosságát. Azt találták, hogy a szintetizált beszéd érthetetlen volt, amennyiben a rendszert az egyik felvételen tanították, és a másikon értékelték ki (adaptálás nélkül). Még egyazon felvétellel is nagy különbségek mutatkoztak a DNN-modellek performanciáját tekintve, felvételfüggő módon. Végül az adott felvételre készült DNN-modell DNN-adaptáció esetén jobban teljesített, illetve ezen túlmenően a DNN-adaptáció gyorsabb működéshez is vezetett.

A fejlesztések az ellenkező irányban is megkezdődtek, azaz az akusztikum-artikuláció inverzió (AAI) területén. Az ilyen rendszerek például számítógéppel segített nyelvoktatásban lehetnek felhasználhatóak, ahol a nyelvultrahang-felvétel vizualizációja alapján begyakorolható az egyes hangzók ejtése. Ebben a kutatásban Csapó Tamás Gábor és munkatársai szintén mély neurális hálón alapuló gépi tanulást alkalmaztak: ezúttal a beszéd spektrális paraméterei alkották a bemenetet, a kimenet pedig a generált nyelvultrahang-videó volt (pl. Porras et al., 2019).

Vitán felül áll a jelentősége Csapó Tamás Gábor azon kezdeményezésének, amely a kommunikációs agy-gép interfész (brain-computer interface, BCI) hazai fejlesztését célozta meg. A BCI-alkalmazások természetes vagy ahhoz közeli minőségű kommunikációt tehetnek lehetővé olyan személyek számára, akik fizikai vagy neurológiai károsodás miatt nem képesek hangzó beszédre. A beszéd valós idejű szintézisének alapjául a mért idegi aktivitás szolgál. Csapó Tamás Gábor és munkatársai elektroencefalográf (EEG) segítségével rögzített agyi aktivitásból kíséreltek meg beszédet szintetizálni (pl. Arthur & Csapó, 2022); illetőleg ugyancsak EEG-jelből ultrahangképet előállítani (Csapó et al., 2023). 2023-ban e téma köré Csapó Tamás Gábor egy workshopot rendezett Moonshine fantázia-

névvel (<https://smartlab.tmit.bme.hu/moonshine-2023/>), hogy szorgalmazza a nemzetközi kapcsolatokat az ekkor már kidolgozás alatt álló projekthez.

7. Összegzés

Csapó Tamás Gábor több területen katalizálta a hazai kutatásokat az által, hogy az ultrahangos képalkotó technikát megismertette a magyar kutatókkal, módszertani és alkalmazásfejlesztéseket végzett és indított el. Az említetteken túlmenően számos egyetemi hallgató és doktorandusz különböző kutatásait is motiválta és támogatta a nyelvultrahangos vizsgálatok különféle területein is, tevékenységének iskolateremtő hatása volt. A széleskörű nemzetközi ismertség és elismertség, a pályázati sikerek jelzik az általa végzett tudományos munka értékét.

Zárszó

Csapó Tamás Gábor mindig érdeklődő, minden új ismeretre nyitott és rendkívül pozitív szemléletű és személyiségű kutató volt, aki nem félt kérdezni, és mindezzel nem csak hallgatói, hanem kollégái előtt is követendő példával járt. Inspiráló volt vele együtt dolgozni. Tamás mindenben támogató volt: akár szakmai kérdések, akár konfliktusok, akár szervezési problémák merültek fel, ott volt, hogy segítsen, megoldja, együtt gondolkodjon, vagy csak jó szóval támogassa társait. Az általa szervezett Moonshine workshopon olyan légkört teremtett, amely hosszútávú nemzetközi együttműködések alapozott meg a BCI területén, és ezek Tamás váratlan távozása nélkül talán a legrangosabb pályázatok elnyeréséig vezethettek volna.

Köszönetnyilvánítás

A tanulmány elkészítését a Magyar Tudományos Akadémia Bolyai János Kutatási Ösztöndíja (Bolyai₉₁₈₂₃) (G. T. E.), az Innovációs és Technológiai Minisztérium ÚNKP-23-5 Új Nemzeti Kiválósági Programja a Nemzeti Kutatási, Fejlesztési és Innovációs Alap forrásából (G. T. E.), a TKA-DAAD 177375

számú ösztöndíja (D. A., J. K.), valamint az NKFIH FK128814 számú és az EKÖP-24 számú pályázata (J. K.) támogatta.

Hivatkozások

- Ahn, S., & Davidson, L. (2016). Tongue root positioning in english voiced obstruents: Effects of manner and vowel context. *Journal of the Acoustical Society of America*, 140, 3221–3221. URL: <https://doi.org/10.1121/1.4970161>. doi:10.1121/1.4970161.
- Arthur, F., & Csapó, T. (2022). Deep learning alapú agyi jel feldolgozás és beszéd-szintézis előkészítő munkálatai. In G. Berend, G. Gosztolya, & V. Vincze (Eds.), *XVIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2022)* (p. 185–198). Szeged: Szegedi Tudományegyetem Informatikai Intézet.
- Ashby, P. (2011). *Understanding phonetics*. London: Hodder Education.
- Beňuš, S., & Gafos, A. (2007). Articulatory characteristics of hungarian ‘transparent’ vowels. *Journal of Phonetics*, 35, 271–300. URL: <https://doi.org/10.1016/j.wocn.2006.11.002>. doi:10.1016/j.wocn.2006.11.002.
- Bolla, K. (1981a). A magyar hosszú mássalhangzók képzése. (kinoröntgenografikus vizsgálat számítógéppel. *Magyar Fonetikai Füzetek*, 7, 7–55.
- Bolla, K. (1981b). A magyar magánhangzók és rövid mássalhangzók képzési sajátosságainak dinamikus kinoröntgenográfiai elemzése. *Magyar Fonetikai Füzetek*, 8, 5–62.
- Csapó, T. (2022). A nyelvmozgás ultrahangos vizsgálata és az automatikus elemzés alkalmazási lehetőségei a beszédtechnológiában. In A. Deme, & A. Kuna (Eds.), *Tanulmányok a nyelvészet alkalmazásainak területéről* (p. 197–219). Budapest: ELTE Eötvös Kiadó.
- Csapó, T., Arthur, F., Nagy, P., & Boncz, A. (2023). Towards ultrasound tongue image prediction from eeg during speech production. In *Procee-*

- dings of the 24th International Speech Communication Association, INTER-SPEECH 2023* (p. 1164–1168). Dublin: International Speech Communication Association. URL: <https://doi.org/10.21437/Interspeech.2023-40>. doi:10.21437/Interspeech.2023-40 iSCA).
- Csapó, T., & Csopor, D. (2015). Ultrahangos nyelvkontúrkövetés automatikusan: a mély neuronhálókra alapuló autotracing eljárás vizsgálata. *Beszédkutatás*, (p. 177–187).
- Csapó, T., Deme, A., Grácsi, T., Markó, A., & Varjasi, G. (2017a). Szinkronizált beszéd- és nyelvultrahang-felvételek a sonospeech rendszerrel. In V. Vincze (Ed.), *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017)* (p. 339–346). Szeged: Szegedi Tudományegyetem Informatikai Intézet. URL: <http://rgai.inf.u-szeged.hu/project/mszny2017/files/kotet.pdf>.
- Csapó, T., Grósz, T., Tóth, L., & Markó, A. (2017b). Beszédszintézis ultrahangos artikulációs felvételekből mély neuronhálók segítségével. In V. Vincze (Ed.), *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)* (p. 181–192). Szeged: Szegedi Tudományegyetem Informatikai Intézet. URL: <https://rgai.inf.u-szeged.hu/sites/rgai.sed.hu/files/kotet.pdf>.
- Csapó, T., & Lulich, S. (2015). Error analysis of extracted tongue contours from 2d ultrasound images. In *Proceedings of Interspeech 2015* (p. 2157–2161). Drezda: ISCA. URL: <https://doi.org/10.21437/Interspeech.2015-486>. doi:10.21437/Interspeech.2015-486.
- Csapó, T., & Xu, K. (2020). Quantification of transducer misalignment in ultrasound tongue imaging. In *Proc. Interspeech 2020* (p. 3735–3739). Sanghaj: Interspeech. URL: <https://doi.org/10.21437/Interspeech.2020-1672>. doi:10.21437/Interspeech.2020-1672.
- Csapó, T., Xu, K., Deme, A., Grácsi, T., & Markó, A. (2021). Transducer misalignment in ultrasound tongue imaging. In M. Tiede, D. Whalen, & V. Gracco (Eds.), *Proceedings of the 12th International Seminar on Speech Production* (p. 166–169). New Haven (CT: Haskins Press).

- Csapó, T., Zainkó, C., Tóth, L., Gosztolya, G., & Markó, A. (2020). Ultrasound-based articulatory-to-acoustic mapping with waveglow speech synthesis. In H. Meng, B. Xu, & T. Zheng (Eds.), *Interspeech 2020* (p. 2727–2731). Sanghaj: International Speech Communication Association. URL: <https://doi.org/10.21437/Interspeech.2020-1031>. doi:10.21437/Interspeech.2020-1031 iSCA).
- Deme, A. (megj). *A magánhangzók változatossága a magyarban: Eredmények a koartikulációs ellenállásról és a koartikulációs agresszióról a magánhangzók közötti koartikulációban artikulációs és akusztikai adatok alapján*. Budapest: Akadémiai Kiadó.
- Deme, A., Bartók, M., Grácsi, T., Csapó, T., & Markó, A. (2019a). Articulatory organization of geminates in hungarian. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences 2019* (p. 1739–1743). Canberra: Australasian Speech Science and Technology Association Inc.
- Deme, A., Bartók, M., Grácsi, T., Csapó, T., & Markó, A. (2019b). Gemináták artikulációs szerveződése a magyarban. *Beszéd kutatás*, 27, 54–74.
- Deme, A., Greisbach, R., Markó, A., Meier, M., Bartók, M., Jankovics, J., & Weidl, Z. (2016). Tongue and jaw movements in high-pitched soprano singing: A case study. *Beszéd kutatás*, (p. 121–138).
- Deme, A., Greisbach, R., Meier, M., Bartók, M., Jankovics, J., Weidl, Z., & Markó, A. (2017). Tongue and jaw articulation of soprano singers at high pitch in hungarian and german. In J. Dang, & A. Li (Eds.), *11th International Seminar on Speech Production (ISSP 2017) Book of Abstracts* (p. 110–113). Tianjin: ISSP.
- Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J., & Brumberg, J. (2010). Silent speech interfaces. *Speech Communication*, 52, 270–287. URL: <https://doi.org/10.1016/j.specom.2009.08.002>. doi:10.1016/j.specom.2009.08.002.

- Epstein, M., & Stone, M. (2005). The tongue stops here: Ultrasound imaging of the palate (1. *The Journal of the Acoustical Society of America*, 118, 2128–2131.
- Fasel, I., & Berry, J. (2010). Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. In *Proceedings of ICPR* (p. 1493–1496). Istanbul: ICPR. URL: <https://doi.org/10.1109/ICPR.2010.369>. doi:10.1109/ICPR.2010.369.
- Goffman, L. (2015). Effects of language on motor processes in development. In M. Redford (Ed.), *The handbook of speech production* (p. 555–577). Chichester: Wiley Blackwell. URL: <https://doi.org/10.1002/9781118584156.ch24>. doi:10.1002/9781118584156.ch24.
- Gosztolya, G., Grósz, T., Tóth, L., Markó, A., & Csapó, T. (2020). Applying dnn adaptation to reduce the session dependency of ultrasound tongue imaging-based silent speech interfaces. *Acta Polytechnica Hungarica*, 17, 109–124. URL: <https://doi.org/10.12700/APH.17.7.2020.7.6>. doi:10.12700/APH.17.7.2020.7.6.
- Grácz, T., Bóna, J., Csapó, T., Deme, A., Bartók, M., & Markó, A. (2018). Medicational effect on acoustic and articulatory vowel and voice parameters in young onset parkinson’s disease. case study. In *Előadás. 17th International Clinical Phonetics and Linguistics Association Conference (ICPLA)* (p. 23–26). St Julians, Malta.
- Grácz, T., Csapó, T., Bartók, M., Deme, A., & Markó, A. (2020a). The realization of voicing opposition in alveolar fricatives in hungarian: Preliminary study on articulation and acoustics. *Beszédtudomány / Speech Science*, 1, 22–56.
- Grácz, T., Csapó, T., Bartók, M., Deme, A., & Markó, A. (2021a). Articulatory and acoustic differentiation of /s/ and /ʃ/ in children’s speech: longitudinal case studies. In J. Bóna (Ed.), *Disfluencies in children’s speech*. Budapest:

- Akadémiai Kiadó. URL: https://mersz.hu/dokumentum/m873dfics__35/. doi:10.1556/9789634547099.4.
- Grácz, T., Csapó, T., Deme, A., Juhász, K., & Markó, A. (2020b). A réshangok zöngésségével összefüggő nyelvpozíciós jellemzők a megelőző magánhangzóban. *Nyelvtudományi Közlemények*, 116, 155–190.
- Grácz, T., Csapó, T., Deme, A., Juhász, K., & Markó, A. (2021b). Tongue root position in vc sequences with regard to the phonetic realization of obstruent voicing: A preliminary study on hungarian. In M. Tiede, D. Whalen, & V. Gracco (Eds.), *Proceedings of the 12th International Seminar on Speech Production* (p. 198–201). New Haven (CT: Haskins Press.
- Grácz, T., Juhász, K., Csapó, T., Deme, A., & Markó, A. (2023). Dynamic articulatory and acoustic features of hungarian sibilants as a function of phonological voicing. In R. Skarnitzl, & J. Volín (Eds.), *Proceedings of 20th International Congress of Phonetic Sciences (ICPhS)* (p. 1077–1081). Praha: Garant.
- Grósz, T., Tóth, L., Gosztolya, G., Csapó, T., & Markó, A. (2018). Kísérletek az alaphérfvencia beclslésére mély neuronhálós, ultrahang-alapú némabeszéd-interfészekben. In V. Vincze (Ed.), *XIV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2018)* (p. 196–2005). Szeged: Szegedi Tudományegyetem Informatikai Intézet.
- Kawahara, S. (2015). The phonetics of obstruent geminates, sokuon. In H. Kubozono (Ed.), *The Mouton Handbook of Japanese Language and Linguistics*. Berlin: Mouton Gruyter. URL: <https://doi.org/10.1515/9781614511984.43>. doi:10.1515/9781614511984.43.
- Kelsey, C., Minifie, F., & Hixon, T. (1969). Applications of ultrasound in speech research. *Journal of Speech and Hearing Research*, 12, 564–575. URL: <https://doi.org/10.1044/jshr.1203.564>. doi:10.1044/jshr.1203.564.

- Kochetov, A. (2020). Research methods in articulatory phonetics i: Introduction and studying oral gestures. *Language and Linguistics Compass*, 14, 1–29. URL: <https://doi.org/10.1111/lnc3.12368>. doi:10.1111/lnc3.12368.
- Kochetov, A., & Kang, Y. (2017). Supralaryngeal implementation of length and laryngeal contrasts in japanese and korean. *Canadian Journal of Linguistics*, 62, 18–55. URL: <https://doi.org/10.1017/cnj.2016.39>. doi:10.1017/cnj.2016.39.
- Li, M., Kambhamettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics*, 19, 545–554. URL: <https://doi.org/10.1080/02699200500113616>. doi:10.1080/02699200500113616.
- Lotz, J. (1966). Egy magyar röntgen-hangosfilm és néhány fonológiai kérdés. *Magyar Nyelv*, 62, 257–266.
- Markó, A., Bartók, M., Csapó, T., Grácz, T., & Deme, A. (2019a). Az /i:/ artikulációs és akusztikai sajátosságai harmonikusan és antiharmonikusan toldalékolódó tövekben. *Nyelvtudományi Közlemények*, 115, 233–254.
- Markó, A., Csapó, T., Bartók, M., Grácz, T., & Deme, A. (2019b). Patterns of lingual cv coarticulation in hungarian children’s speech: The case of stops. In R. Rose, & R. Eklund (Eds.), *Proceedings of DiSS 2019. The 9th Workshop on Disfluency in Spontaneous Speech ELTE Eötvös Loránd University Budapest, Hungary 12–13 September, 2019* (p. 93–94). Budapest: ELTE Faculty of Humanities.
- Markó, A., Csapó, T., Deme, A., Bartók, M., & Grácz, T. (2020). A magánhangzók lingvális képzési jegyeinek tanítása az artikulációs tapasztalat, a kutatási eredmények és az elméleti kategóriák tükrében. In I. Zs. Ludányi, & A. Domonkosi (Eds.), *A nyelv perspektívája az oktatásban. Válogatás a PeLi-Kon2018 oktatásnyelvészeti konferencia előadásaiból* (p. 71–82). Eger: EKE Líceum Kiadó. URL: <https://doi.org/10.17048/Pelikon2018.2020.71>. doi:10.17048/Pelikon2018.2020.71.

- Markó, A., Csapó, T., Deme, A., Grácsi, T., & Bartók, M. (2019c). Gyermekek lingvális artikulációjának variabilitása magánhangzós nyelvkontúrok alapján. In J. Bóna, & V. Horváth (Eds.), *Az anyanyelvelsajátítás folyamata hároméves kor után* (p. 165–190). Budapest: ELTE Eötvös Kiadó.
- Markó, A., Csapó, T., Deme, A., Grácsi, T., & G, V. (2017). A gyermeki artikuláció vizsgálata – Új lehetőségek a hazai kutatásban. In J. Bóna (Ed.), *Új utak a gyermeknyelvi kutatásokban* (p. 65–95). Budapest: ELTE Eötvös Kiadó.
- Mády, K. (2008). Magyar magánhangzók vizsgálata elektromágneses artikulo-gráffal normál és gyors beszédben. *Beszédkutatás*, (p. 52–66).
- Narayanan, S., Alwan, A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, 98, 1325–1347. URL: <https://doi.org/10.1121/1.413469>. doi:10.1121/1.413469.
- Neuberger, T. (2015). Durational correlates of singleton-geminate contrast in hungarian voiceless stops. In *Proc. 18th ICPhS*. paper: ICPHS0422.
- Nittrouer, S., Studdert-Kennedy, M., & Neely, S. (1996). How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech and Hearing Research*, 39, 379–389. doi:10.1044/jshr.3902.379.
- Olaszy, G. (2006). Hangidőtartamok és időszerkezeti elemek a magyar beszédben. *Nyelvtudományi Értekezések*, 155.
- Payne, E. (2006). Non-durational indices in italian geminate consonants. *Journal of the International Phonetic Association*, 3, 83–95. URL: <https://doi.org/10.1017/S0025100306002398>. doi:10.1017/S0025100306002398.
- Percival, M., Csapó, T., Bartók, M., Deme, A., E., T., & Markó, A. (2020a). Tongue root and voicing in hungarian singleton and geminate

- obstruents. In *12th International Seminar on Speech Production*. URL: https://issp2020.yale.edu/S08/percival_08_15_156_abstract.pdf.
- Percival, M., Csapó, T., Markó, A., Deme, A., Grácz, T., & Bartók, M. (2020b). Gemination as fortition? articulatory data from hungarian. In *Előadás. LabPhon17 virtual conference* (p. 6–8). Vancouver. URL: https://labphon.org/sites/default/files/previous_conferences/LP17/abstracts/LabPhon_17_paper_308.pdf.
- Percival, M., G., C. T., Bartók, M., Deme, A., Grácz, T., & Markó, A. (2020c). Ultrasound imaging of hungarian geminates. előadás. ultrafest ix. *Ultrasound Imaging for Speech and Language. Virtual Conference*, (p. 21–24). URL: https://ultrafest2020.indiana.edu/abstracts/UltraFest_IX_Percival_Csapo_Bartok_Deme_Graczi_Marko_Hungarian.pdf.
- Percival, M., Kochetov, A., & Kang, Y. (2018). An ultrasound study of gemination in coronal stops in eastern oromo. *Proceedings of Interspeech*, (p. 1531–1535). URL: <https://doi.org/10.21437/Interspeech.2018-2512>. doi:10.21437/Interspeech.2018-2512.
- Porrás, D., Sepulveda-Sepulveda, A., & Csapó, T. (2019). Dnn-based acoustic-to-articulatory inversion using ultrasound tongue imaging. In *Proceedings of the International Joint Conference on Neural Networks* (p. 1–8). New Jersey: IEEE. URL: <https://doi.org/10.1109/IJCNN.2019.8851769>. doi:10.1109/IJCNN.2019.8851769.
- Prenger, R., Valle, R., & Catanzaro, B. (2019). Waveglow: A flow-based generative network for speech synthesis. In *Proc ICASSP* (p. 3617–3621). Brighton: ICASSP. URL: <https://doi.org/10.1109/ICASSP.2019.8683143>. doi:10.1109/ICASSP.2019.8683143.
- Pycha, A. (2009). Lengthened affricates as a test case for the phonetics–phonology interface. *Journal of the International Phonetic Association*, 39, 1–31.

- Pycha, A. (2010). A test case for the phonetics–phonology interface: Gemination restrictions in hungarian. *Phonology*, 27, 119–152.
- Recasens, D., & Rodríguez, C. (2016). A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound. *Journal of Phonetics*, 59, 58–75. URL: <https://doi.org/10.1016/j.wocn.2016.09.002>. doi:10.1016/j.wocn.2016.09.002.
- Ridouane, R. (2007). Gemination in tashlhiyt berber: an acoustic and articulatory study. *Journal of the International Phonetic Association*, 37, 119–142.
- Scobbie, J., Lawson, E., Cowen, S., Cleland, J., & Wrench, A. (2011). A common co-ordinate system for mid-sagittal articulatory measurement. In *Working Paper WP-20*. Edinburgh: Queen Margaret University. doi:<https://core.ac.uk/reader/161924590>.
- Seikel, J., King, D., & Drumright, D. (2010). *Anatomy & physiology for speech, language and hearing*. (4th ed.). Delmar: Cengage Learning.
- Siptár, P., & Grácz, T. (2014). Degemination in hungarian: Phonology or phonetics? *Acta Linguistica Hungarica*, 61, 443–471. URL: <https://doi.org/10.1556/aling.61.2014.4.4>. doi:10.1556/aling.61.2014.4.4.
- Siptár, P., & Törkenczy, M. (2000/2007). *The Phonology of Hungarian*. Oxford: Oxford University Press.
- Stevens, K. (1998). *Acoustic Phonetics*. Massachusetts: MIT Press. URL: <https://doi.org/10.7551/mitpress/1072.001.0001>. doi:10.7551/mitpress/1072.001.0001.
- Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19, 455–501. URL: <https://doi.org/10.1080/02699200500113558>. doi:10.1080/02699200500113558.
- Stone, M. (2010). Laboratory techniques for investigating speech articulation. In W. Hardcastle, J. Laver, & F. Gibbon (Eds.), *The Handbook*

- of *Phonetic Sciences* (p. 9–38). Oxford: Willey-Blackwell. URL: <https://doi.org/10.1002/9781444317251.ch1>. doi:10.1002/9781444317251.ch1.
- Szende, T. (1974). A magyar hangrendszer néhány összefüggése röntgenográfiai vizsgálatok tükrében. *Magyar Nyelv*, 70, 68–77.
- Tang, L., Bressmann, T., & Hamarneh, G. (2012). Tongue contour tracking in dynamic ultrasound via higher-order mrfs and efficient fusion moves. *Medical Image Analysis*, 16, 1503–1520. URL: <https://doi.org/10.1016/j.media.2012.07.001>. doi:10.1016/j.media.2012.07.001.
- Terband, H., Maassen, B., Lieshout, P., & Nijlan, D. (2011). Stability and composition of functional synergies for speech movements in children with developmental speech disorders. *Journal of Communication Disorders*, 44, 59–74. URL: <https://doi.org/10.1016/j.jcomdis.2010.07.003>. doi:10.1016/j.jcomdis.2010.07.003.
- Trencsényi, R. (2020a). Mri- és uh-felvételek geometriai elemzése a beszéd-szintézisben. *Acta Medicinae et Sociologica*, 11, 55–65.
- Trencsényi, R. (2020b). A nyelvkontúrkövető algoritmusok és a gépi tanulás összekapcsolhatóságának vizsgálata. In G. Berend, G. Gosztolya, & V. Vincze (Eds.), *XVI. Magyar Számítógépes Nyelvészeti Konferencia: MSZNY 2020* (p. 23–24 233–244). Szeged: Szegedi Tudományegyetem Informatikai Intézet.
- Trencsényi, R., & Czap, L. (2021). Possible methods for combining tongue contours of dynamic mri and ultrasound records. *Acta Polytechnica Hungarica*, 18, 143–160.
- Vorperian, H., & Kent, R. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, 50, 1510–1545. URL: [https://doi.org/10.1044/1092-4388\(2007/104\)](https://doi.org/10.1044/1092-4388(2007/104)). doi:10.1044/1092-4388(2007/104).
- Vorperian, H., Wang, S., Chung, M., M., S. E., B., D. R., D., K. R., Ziegert, A., & Gentry, L. (2009). Anatomic development of the oral and pharyngeal

- portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*, 125, 1666–1678. URL: <https://doi.org/10.1121/1.3075589>. doi:10.1121/1.3075589.
- Whalen, D., Kang, J., Iwasaki, R., Shejaeya, G., Kim, B., Roon, K., & Tiede, M. (2019). Accuracy assessments of hand and automatic measurements of ultrasound images of the tongue. In S. Calhoun, P. M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (p. 542–546). Canberra: Australasian Speech Science and Technology Association Inc.
- Zharkova, N. (2016). Ultrasound and acoustic analysis of sibilant fricatives in preadolescents and adults. *The Journal of the Acoustical Society of America*, 139, 2342–2351. URL: <https://doi.org/10.1121/1.4947046>. doi:10.1121/1.4947046.
- Zharkova, N., Gibbon, F., & Hardcastle, W. (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical Linguistics and Phonetics*, 29, 249–265. URL: <https://doi.org/10.3109/02699206.2015.1007528>. doi:10.3109/02699206.2015.1007528.
- Zharkova, N., Hardcastle, W., & Gibbon, F. (2018). The dynamics of voiceless sibilant fricative production in children between 7 and 13 years old: An ultrasound and acoustic study. *The Journal of the Acoustical Society of America*, 144, 1454–1466. URL: <https://doi.org/10.1121/1.5053585>. doi:10.1121/1.5053585.
- Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using english /t/ and /a/. *Journal of Phonetics*, 37, 248–256. URL: <https://doi.org/10.1016/j.wocn.2008.10.005>. doi:10.1016/j.wocn.2008.10.005.
- Zharkova, N., Hewlett, N., & Hardcastle, W. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study.

Motor Control, 15, 118–140. URL: <https://doi.org/10.1123/mcj.15.1.118>.
doi:10.1123/mcj.15.1.118.

Zharkova, N., Hewlett, N., & Hardcastle, W. (2012). An ultrasound study of lingual coarticulation in /sv/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, 42, 193–208. doi:10.1017/S0025100312000060.