# Artikulációs fonetikai jellemzők verifikálása kvantitatív adatokkal

Trencsényi Réka[1], Czap László[2]

[1] *Debreceni Egyetem, Villamosmérnöki Tanszék*
[2] *Miskolci Egyetem, Automatizálási és Infokommunikációs Intézet*

**Abstract**

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

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

## 1. Bevezető

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

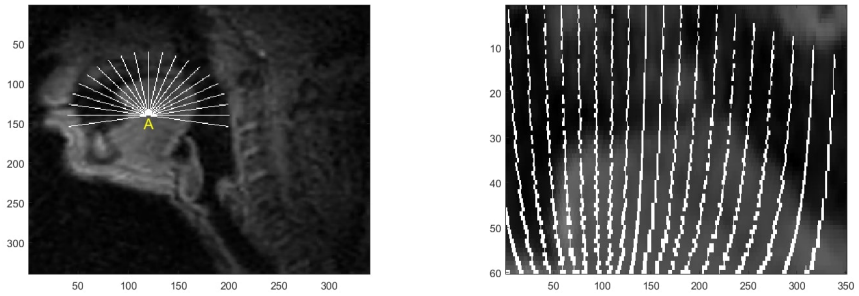
*Email addresses:* trencsenyi.reka@science.unideb.hu (Trencsényi Réka), czap@uni-miskolc.hu (Czap László)

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

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

## 2. Módszerek

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




(a) (b)

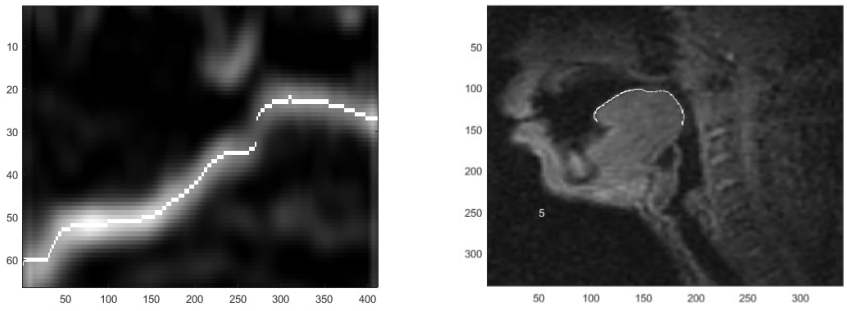
1. ábra. Az előfeldolgozás első lépése: (a) az MRI-kép radiális újra-mintavételezése, (b) az újra-mintavételezett kép descartes-i oszlopmátrixa. A könnyebb áttekinthetőség végett a radiális egyenesek 10°-onként vannak megrajzolva, de az újra-mintavételezés valójában 1°-onként történik.

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

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

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

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




(a) (b)

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

### *2.1. A nyelv pozíciójának vizsgálata*

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

*2.1.1. A magánhangzókhoz tartozó nyelvpozíciók meghatározásának kvantitatív módszere*

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

$$Cxy = [\overline{x}, \overline{y}] \tag{1}$$

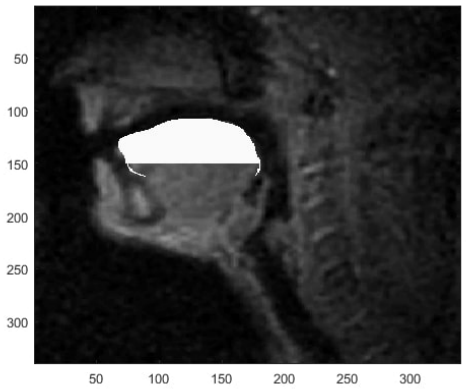
formula által adott súlypontjának $x$ és $y$ koordinátáit a 3. ábrán az $e$ hang esetében prezentált feltöltött nyelvtest fehér pontjainak horizontális és vertikális koordinátáival számított elsőrendű momentumként (Hu, 1962; Mukundan & Ramakrishnan, 1998) származtatjuk a

$$\overline{x} = \frac{1}{n} \sum_{x} \sum_{y} x \cdot f(x, y), \tag{2}$$

$$\overline{y} = \frac{1}{n} \sum_{x} \sum_{y} y \cdot f(x, y) \tag{3}$$

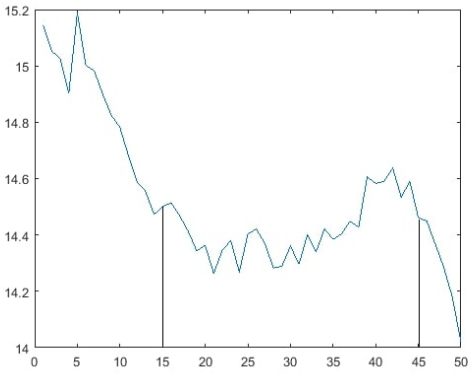
felírás szerint, ahol az $f$ függvény értéke $f(x, y) = 1$ a fehér területen belül, és $f(x, y) = 0$ azon kívül, $n$ pedig az adott tartományt lefedő fehér pontok számát megadó konstans. A nyelvtest feltöltése a nyelvfelszín legfelső pontjából indul, és a 3. ábrával összhangban a kép alsó része felé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

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




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

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




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

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

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

### *2.1.2. A mássalhangzókhoz tartozó artikulációs helyek meghatározásának kvantitatív módszere*

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

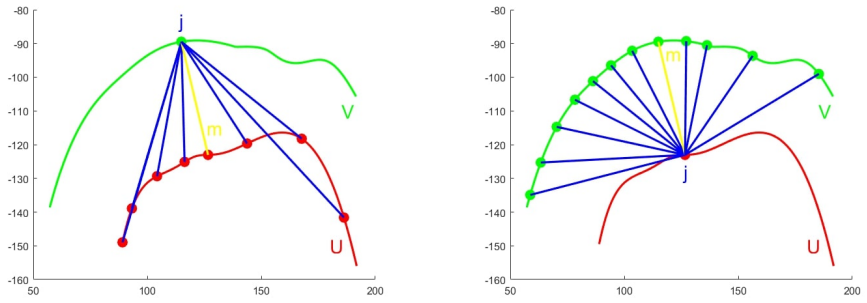
A nyelvkontúr automatikus követését lehetővé tevő algoritmust alapul véve, analóg módon meghatározható a fogmeder és a szájpad kontúrja is. A különbség mindössze abban nyilvánul meg, hogy az MRI-kereteken kiválasztott körközépponttól távolodva nem csökkenő, hanem növekvő világosságértéket kell detektálnunk.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

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

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

Nearest Neighbour Distance (NND) távolságmérték alkalmazásával, ami kiváltképp alkalmas eltérő számú pontból felépülő görbék távolságának meghatározására (Zharkova & Hewlett, 2009). Tegyük fel, hogy a nyelvkontúr az $U = [u1, u2, \ldots, un]$, a szájpadkontúr pedig a $V = [v1, v2, \ldots, vm]$ pontok által van kifeszítve. Ekkor a $V$ görbe adott pontjának az $U$ görbétől mért távolsága az $U$ görbe hozzá legközelebb eső pontjától mért távolsággal egyezik meg, amit matematikailag (3) első összegjárulékában szereplő minimum ír le, és grafikusan az 5.a ábra szerint értelmezhető. Hasonló gondolatmenetet követve, (3) második

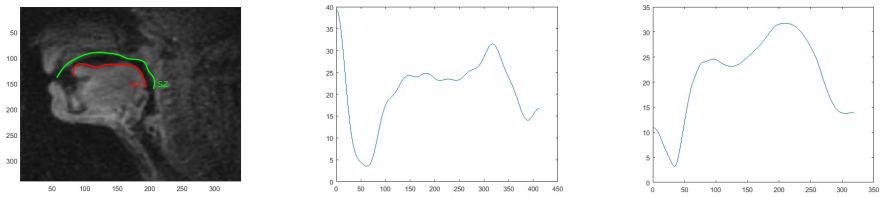
összegjárulékának minimuma a $V$ görbe adott pontjának és az $U$ görbe hozzá legközelebb eső pontjának távolságát adja meg, amit az 5.b ábra szemléltet. Az 5. ábra távolságainak $m$ minimuma sárgával van megjelölve. Gondoljuk meg, hogy a (3) kifejezés két összegjárulékában számított távolságok nem feltétlenül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




(a) (b)

5. ábra. Az NND távolságmérték grafikus reprezentációja

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



(a) (b) (c)

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

## 3. Eredmények

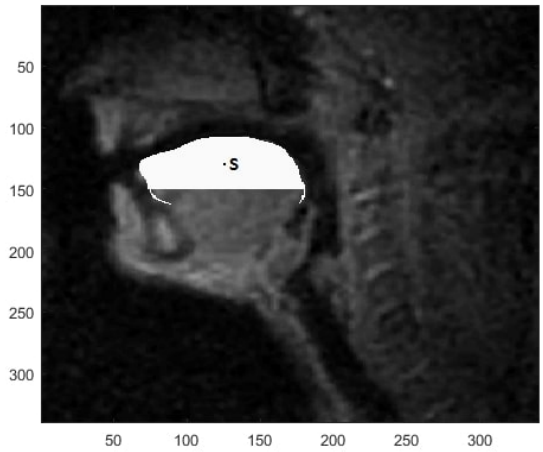
Méréseink során, a nyelv beszéd közbeni mozgását dinamikusan követő nyelvkontúrjaink segítségével olyan geometriai adatokra tettünk szert, melyek alapvető fontosságúak lehetnek az artikuláció kvantitatív jellemzésében. A magánhangzókra és mássalhangzókra kapott eredményeinket elkülönítve tárgyaljuk,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

### *3.1. Magánhangzók*

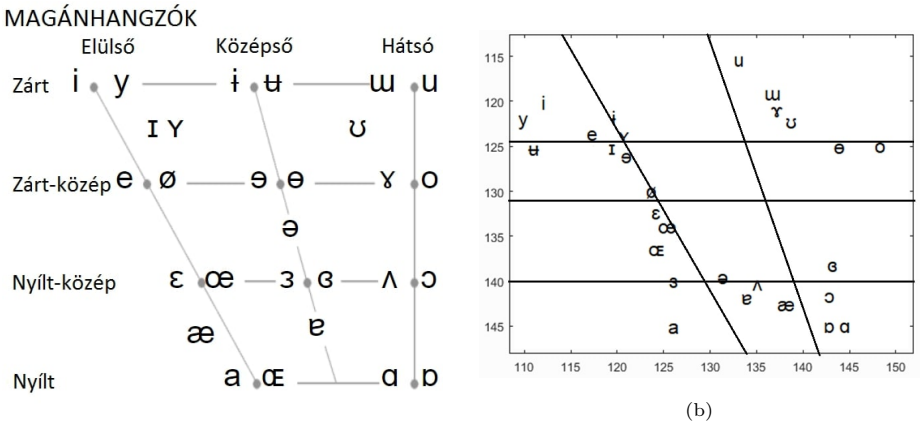
A magánhangzók esetében a vokális traktus longitudinális tengelye mentén képzett keresztmetszeti adatokat az állkapocs nyitottsága és a nyelv helyzete befolyásolja. A vokális traktus keskenyebb és szélesebb metszetei, valamint az ajkak alakja együttesen meghatározzák a gégéből kiinduló gerjesztőjel spektrális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

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

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




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




(a)

(b)

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

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

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

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

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

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

*3.2. Mássalhangzók*

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

2. táblázat. A mássalhangzók típusai és képzési helyei

MÁSSALHANGZÓK © 2005 IPA

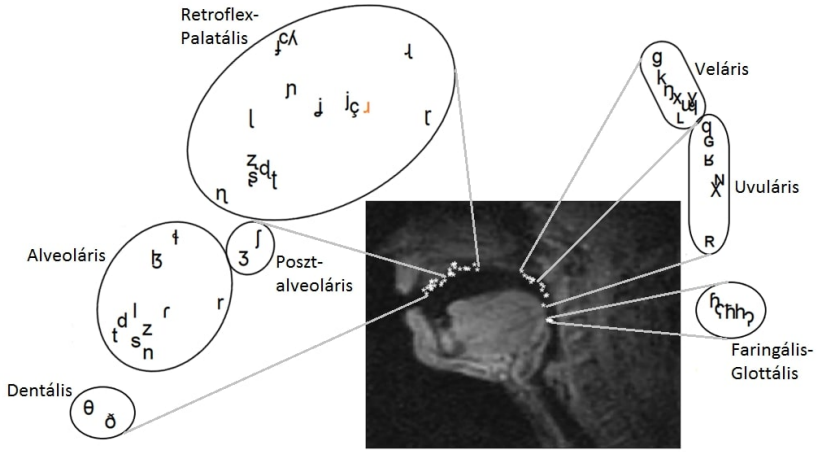
| | Bilabiális | Labiodentális | Dentális | Alveoláris | Posztalveoláris | Retroflex | Palatális | Veláris | Uvuláris | Faringális | Glottális |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Plozíva | p b | | | t d | | ʈ ɖ | c ɟ | k g | q ɢ | | ʔ |
| Nazális | m | ɱ | | n | | ɳ | ɲ | ŋ | ɴ | |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

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

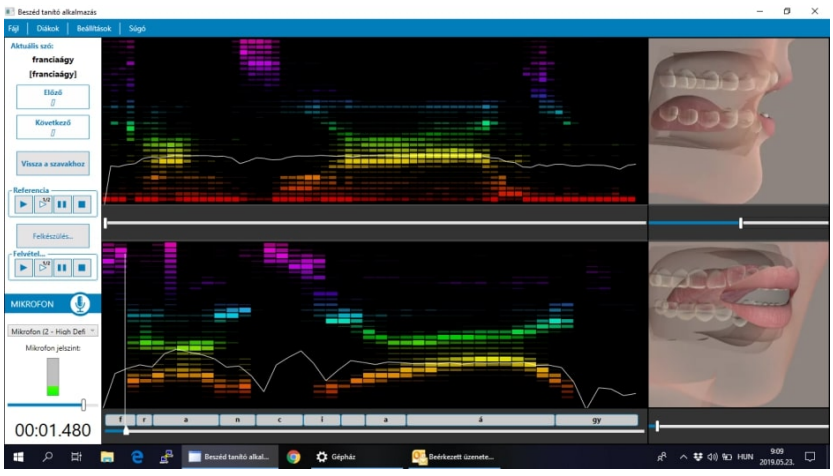
### *3.3. Az eredmények alkalmazhatósága a Beszédasszisztens rendszerben*

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

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




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




10. ábra. A Beszédasszisztens gyakorlás közbeni monitorképe

## 4. Összefoglaló

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

## Hivatkozások


Barnaud, M. L., Schwartz, J. L., Bessière, P., & Diard, J. (2019). Computer simulations of coupled idiosyncrasies in speech perception and speech production with COSMO, a perceptuo-motor Bayesian model of speech communication. *PLoS ONE*, *14*, 1.

Baum, S. R., Blumstein, S. E., Naeser, M. A., & Palumbo, C. L. (1990). Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language*, *39*, 33–56. URL: https://doi.org/10.1016/0093-934X(90)90003-Y. doi:10.1016/0093-934X(90)90003-Y.

Csapó, T. G., Al-Radhi, M. S., Németh, G., Gosztolya, G., Grósz, T., Tóth, L., & Markó, A. (2019). Ultrasound-Based Silent Speech Interface Built

on a Continuous Vocoder. *Proc. Interspeech*, (p. 894–898). doi:10.21437/Interspeech.2019-2046.

Czap, L. (2021). Impact of Preprocessing Features on the Performance of Ultrasound Tongue Contour Tracking, via Dynamic Programming. *Acta Polytechnica Hungarica*, *18*, 19. URL: http://acta.uni-obuda.hu/Czap_109.pdf.

Czap, L., Pintér, J. M., & Baksa-Varga, E. (2019). Features and results of a speech improvement experiment on hard of hearing children. *Speech Communication*, *106*, 7–20. URL: https://doi.org/10.1016/j.specom.2018.11.003. doi:10.1016/j.specom.2018.11.003.

Daassi-Gnaba, H., & Krahe, J. L. (2009). Universal combined system: speech recognition, emotion recognition and talking head for deaf and hard of hearing people. In *Conférence-AAATE* (p. 503–508).

Douros, I. K., Kulkarni, A., Dourou, C., Xie, Y., Felblinger, J., Isaieva, K., Vuissoz, P., & Laprie, Y. (2020). Using Silence MR Image to Synthesise Dynamic MRI Vocal Tract Data of CV. *Proc. Interspeech*, (p. 3730–3734). doi:10.21437/Interspeech.2020-1173.

Erdogan, N., & Wei, M. (2019). Articulatory Phonetics: English Consonants. In N. Erdogan, & M. Wei (Eds.), *Applied Linguistics for Teachers of Culturally and Linguistically Diverse Learners* (p. 263–284). IGI Global. URL: http://doi:10.4018/978-1-5225-8467-4.ch011.

Fagel, S., & Clemens, C. (2004). An articulation model for audiovisual speech synthesis—Determination, adjustment, evaluation. *Speech Communication*, *44*, 141–154. URL: https://doi.org/10.1016/j.specom.2004.10.006. doi:10.1016/j.specom.2004.10.006.

Hu, M. K. (1962). Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, *8*, 179–187. URL: https://doi.org/10.1109/TIT.1962.1057692.

Ivanova, S. A., & Hasko, V. (2019). Articulatory Phonetics: English Vowels. In N. Erdogan, & M. Wei (Eds.), *Applied Linguistics for Teachers of Culturally and Linguistically Diverse Learners* (p. 285–301). IGI Global. URL: http://doi:10.4018/978-1-5225-8467-4.ch012.

Mattheyses, W., & Verhelst, W. (2015). Audiovisual speech synthesis: An overview of the state-of-the-art. *Speech Communication*, *66*, 182–217. URL: https://doi.org/10.1016/j.specom.2014.11.001.

Mukundan, R., & Ramakrishnan, K. R. (1998). *Moment functions in image analysis*. Singapore: World Scientific Press.

Peng, X., Chen, H., Wang, L., Tian, F., & Wang, H. (2020). Talking Head-based L2 Pronunciation Training: Impact on Achievement Emotions, Cognitive Load, and Their Relationships with Learning Performance. *International Journal of Human–Computer Interaction*, *36*, 1487–1502. URL: https://doi.org/10.1080/10447318.2020.1752476.

Recasens, D. (1991). On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics*, *19*, 267–280. URL: https://doi.org/10.1016/s0095-4470(19)30344-4.

Segaran, K., Ali, A. Z. M., & Hoe, T. W. (2014). Usability and user satisfaction of 3D talking-head mobile assisted language learning (MALL) app for non-native speakers. *Procedia-Social and Behavioral Sciences*, *131*, 4–10. URL: https://doi.org/10.1016/j.sbspro.2014.04.069.

Serrurier, A., Badin, P., Barney, A., Boë, L. J., & Savariaux, C. (2012). The tongue in speech and feeding: Comparative articulatory modelling. *Journal of Phonetics*, *40*, 745–763. URL: https://doi.org/10.1016/j.wocn.2012.08.001.

Wang, J., Green, J. R., Samal, A., & Yunusova, Y. (2013). Articulatory distinctiveness of vowels and consonants: a data-driven approach. *Journal*

*of Speech, Language and Hearing Research*, *56*, 1539–1551. URL: https://doi.org/10.1044/1092-4388(2013/12-0030).

Wang, X., Hueber, T., & Badin, P. (2014). On the use of an articulatory talking head for second language pronunciation training: the case of Chinese learners of French. In *10th international seminar on speech production (issp 2014)* (p. 449–452).

website1 (n.d.). URL: https://www.internationalphoneticalphabet.org/ipa-charts/vowels Accessed 22.04.2020.

website2 (n.d.). URL: https://www.internationalphoneticalphabet.org/ipa-charts/consonants/ Accessed 22.04.2020.

website3 (n.d.). URL: sail.usc.edu/span/rtmri_ipa/je_2015.html Accessed 15.11.2009.

Zhao, J., Lirong, W., Chao, Z., Lijuan, S., & Jia, Y. (2010). Pronunciation of rehabilitation methods based on 3d-talking head. In *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering* (p. 17–20).

Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/. *Journal of Phonetics*, *37*, 248–256. URL: https://doi.org/10.1016/j.wocn.2008.10.005.