

MISKOLCI EGYETEM
GÉPÉSZMÉRNÖKI ÉS INFORMATIKAI KAR



HATVANY JÓZSEF INFORMATIKAI TUDOMÁNYOK
DOKTORI ISKOLA

Vezető: PROF. DR. SZIGETI JENŐ

Az artikuláció geometriai és akusztikai jellemzőinek kapcsolata

PhD értekezés

Készítette: DR. TRENCSENYI RÉKA

Témavezető: PROF. DR. CZAP LÁSZLÓ

Miskolc
2024

*"Elő, elő az a nagy szép idő,
Amely felé reményeim szállanak..."*

Szüleimnek szeretettel

Tartalomjegyzék

1. Bevezető	1
2. A felhasznált UH- és MRI-felvételek	4
2.1. A felvételek vizuális jellemzése	4
2.2. A felvételek geometriai jellemzése	6
3. Automatikus nyelvkontúrkövetés	7
4. Szájpadkontúrok azonosítása	15
4.1. Az UH-szájpadkontúr megállapítása	16
4.2. Az MRI-szájpadkontúr megállapítása	24
5. UH- és MRI-nyelvkontúrok összehangolása optimalizációs eljárások segítségével	27
5.1. A transzformációk geometriai aspektusai és matematikai formulái	28
5.2. Az UH- és MRI-kontúrok globális távolságának értelmezése	31
5.3. A paraméterek optimalizációját segítő megfontolások	32
5.4. Az optimalizációs algoritmusok felépítése	35
5.5. A megvalósított optimalizációs algoritmusok	37
5.5.1. Optimalizáció 1	37
5.5.2. Optimalizáció 2	38
5.5.2.1. Optimalizáció 2.a	41
5.5.2.2. Optimalizáció 2.b	42
5.5.3. Optimalizáció 3	42
5.5.3.1. Optimalizáció 3.a	42
5.5.3.2. Optimalizáció 3.b	43
5.5.4. Optimalizáció 4	44
5.5.5. Az optimalizáció eredményeinek vizuális validálása	47
5.5.6. Az optimalizáció eredményeinek numerikus validálása	48
5.6. T1 tézis	51
6. UH- és MRI-nyelvkontúrokkal megvalósított gépi tanítás	51
6.1. A gépi tanítás eszközei	51

6.2. A nyelvkontúrok paraméterei	52
6.2.1. Tanítópontok	52
6.2.2. DCT-együtthatók	53
6.3. A gépi tanítás módszere	54
6.4. A neurális hálózatok konstrukciója	55
6.4.1. Tanítás 1	56
6.4.1.1. V_1 verzió	56
6.4.1.2. V_2 verzió	58
6.4.2. Az UH-nyelvkontúrok szelekciója	59
6.4.2.1. A hangátmenetekhez tartozó UH-nyelvkontúrok kiszűrése	59
6.4.2.2. A negatív iránytangensű UH-nyelvkontúrok kiszűrése	60
6.4.2.3. A szájpaddockon túlcsorduló UH-nyelvkontúrok kiszűrése	61
6.4.2.4. A konvex UH-nyelvkontúrok kiszűrése	62
6.4.2.5. A kirívó UH-nyelvkontúrok kiszűrése	63
6.4.3. Tanítás 2	64
6.4.4. Tanítás 3	66
6.5. T2 tézis	69
7. Artikulációs beszéd szintézis	69
7.1. Az akusztikus csőmodell	70
7.2. A lineáris predikció elve	71
7.3. Az MRI-kontúrok kiterjesztése	73
7.4. Dinamikus távolságmérés az UH- és MRI-felvételeken	74
7.4.1. Távolságmérés 1	75
7.4.2. Távolságmérés 2	78
7.4.3. A távolságmérő módszerek eredményeinek összehasonlítása	81
7.5. Beszédszintézis	82
7.5.1. A neurális hálózatok felépítése	82
7.5.2. A neurális hálózat bemeneti és kimeneti paramétereinek kapcsolata UH-MRI viszonylatban	85
7.5.3. A beszédszintézis módszerei	86
7.5.3.1. Az akusztikus csőmodellen alapuló szintézis	86
7.5.3.2. A lineáris predikció elvén alapuló szintézis	87
7.5.4. A beszédszintézis eredményei	88

7.6. T3 tézis	93
8. A tézisek korlátai, érvényessége és újdonságai	93
9. Összefoglaló	93
10. Summary	94
Köszönetnyilvánítás	95
Hivatkozások	96
1. függelék	104
2. függelék	105
3. függelék	106
4. függelék	110

Rövidítések:

UH – ultrahang

MRI – Magnetic Resonance Imaging (mágneses rezonanciás képalkotás)

NND – Nearest Neighbour Distance (legközelebbi szomszéd távolsága)

DCT – Discrete Cosine Transform (diszkrét koszinusztranszformáció)

SCG – Scaled Conjugate Gradient (skálázott konjugált gradiens)

LPC – Linear Predictive Coding (lineáris predikciós kódolás)

MAKIMA – Modified Akima (módosított Akima)

PCHIP – Piecewise Cubic Hermite Interpolating Polynomial (szakaszonként köbös Hermite-féle interpoláló polinom)

IPA – International Phonetic Alphabet (nemzetközi fonetikai abc)

Mennyiségek jelölései:

$(r; \phi)$ – síkbeli polárkoordináták

$(x; y)$ – síkbeli descartes-i koordináták

C – a geometriai transzformációk középpontja az UH-kereteken

C' – a geometriai transzformációk középpontja az MRI-kereteken

G – a nyugalomban lévő gégefedő csúcspontja az UH-kereteken

G' – a nyugalomban lévő gégefedő csúcspontja az MRI-kereteken

R_1 – radiális távolság az UH-kereteken

R_2 – radiális távolság az MRI-kereteken

R – radiális nagyítás

ϕ_1 – síkbeli polárszög az UH-kereteken

ϕ_2 – síkbeli polárszög az MRI-kereteken

Φ_0 – síkbeli szögelfordulás

Φ – a szögtartomány skálafaktora

l – a vokális traktus csőszakaszainak hossza

L – a vokális traktus teljes hossza

A – a vokális traktus csőszakaszainak keresztmetszete

N – görbék pontjainak száma vagy a vokális traktus csőszakaszainak száma

r – reflexiós tényező

$s(n)$ – diszkrét idejű jel

$\bar{s}(n)$ – diszkrét idejű becsült jel

$h(n)$ – diszkrét idejű hibajel

α_i, β_i – lineáris predikciós együtthatók

$F(z)$ – a lineáris predikciós együtthatók Laplace-transzformáltja a diszkrét frekvenciatartományban

$S(z)$ – a diszkrét idejű jel Laplace-transzformáltja a diszkrét frekvenciatartományban

$\bar{S}(z)$ – a diszkrét idejű becsült jel Laplace-transzformáltja a diszkrét frekvenciatartományban

$H(z)$ – a diszkrét idejű hibajel Laplace-transzformáltja a diszkrét frekvenciatartományban

$A(z), B(z)$ – átviteli függvények a diszkrét frekvenciatartományban

T_s – mintavételi periódusidő

f_s – mintavételi frekvencia

v – hangsebesség levegőben

$u(x, t)$ – a levegő térfogatárama

τ – időbeli késleltetés

F_1, F_2 – formánsfrekvenciák

1. Bevezető

A beszédkeltés az egyik legalapvetőbb és legtermészetesebb mozzanatként épül be minden napjainkba, ennél fogva nem meglepő, hogy a beszéd fiziológiai, akusztikai, lingvisztikai, prozódiai, percepció és kognitív pszichológiai vonatkozásainak tanulmányozása évszázadokra nyúlik vissza. Ezen hosszú időszak tartalmas és sokrétű kutatási irányzatainak ellenére azonban a témát még napjainkban is intenzív érdeklődés övezi és rendkívül sok kihívást tartogat még a szakemberek számára.

A beszéd kutatás egyik legfontosabb tématerülete a beszéd szintézis, ami elemi alkotóját képezheti az ember-gép kapcsolatnak. Ez esetben a gép kommunikációs szerepe abban nyilvánul meg, hogy kódoló adóvá válik, azaz beszédet produkál. A beszéd szintézis története rendkívül terjedelmes korszakot fed le, hiszen a kezdeti lépések a XVIII. század második felében gyökereznek, amikor is megvalósult az első beszédkeltő gép (Kempelen Farkas, 1791). Kempelen Farkas szavakat produkáló mechanikus szerkezete, majd a folyamatos beszédet előállító elektromechanikus beszélőgép (D. Homer, R. Ries, S. Watkins, 1939) után az igazi fellendülés a XX. század második felében indult el, miután az informatika virágzásának kezdetén lehetővé vált a számítógéppel vezérelt beszéd szintetizátorok megalkotása. A megnövekedett komputerkapacitásnak köszönhetően a gyors gépi vezérlés mellett lehetőség nyílt a beszéd finomabb részleteinek kidolgozására is. A klasszikus mechanikus eszközöket felváltották az elektronikai készülékek, melyekkel zöngés és zöngétlen gerjesztést, illetve jelszűrést lehetett létrehozni. A modernizáció nemcsak gyakorlati szinten idézett elő változásokat, hanem az elméleti háttér átformálódásában és fejlődésében is megmutatkozott, hiszen a számítógépes programok megszerkesztéséhez elengedhetetlen volt a beszédképzés paramétereinek időbeli modellezése.

Napjainkban a beszéd szintézis legelterjedtebb irányzata a szövegfelolvasók készítése, melyek konkrét témakörre szűkített vagy általános témájú írott szöveget szólaltatnak meg. Ebbe az alkalmazási kategóriába sorolhatók például a szépirodalmi felolvasók, az utastájékoztató rendszerek, a hírolvasók, a hangos időjárás-jelentés vagy a telefonos tudakozó szolgáltatás. A beszéd szintetizátorok megalkotásának célja a természetes emberi beszéd közben kialakuló akusztikai produktum élethű utánzása. Ebben a megközelítésben a beszéd hullámformája adja a kiindulópontot, amit kétfajta megoldásban alkalmaznak gépi beszéd előállítására. Az egyik csoportba az úgynevezett forráskódolású technikák tartoznak, melyek segítségével a beszédjelből kivonják a lényegi információkat és ezeket bemeneti adatsorozatként kezelik a szintézis során [1]. Természetesen ezeknél a módszereknél gondoskodni kell arról, hogy az adatokból kimeneti végeredményként ismét hullámforma álljon elő. A másik megoldás az emberi hangot közvetlenül használja fel a beszédépítéshez olyan módon, hogy a beszédjelből különböző hosszúságú hullámforma-részleteket vágnak ki és tárolnak el, majd az így kapott elemek megfelelő kiválasztásával és összefűzésével megkonstruálják a kívánt beszédhullámot [2]. Ezeken túlmenően, tágabb módszertani szempontok alapján megkülönböztetünk még szabályalapú, illetve statisztikai elven működő beszéd-előállítási eljárásokat. Az előbbi esetében megfigyelések és tapasztalatok szerint felállított szabályokkal koordinálják a szintézis egyes lépéseit [3], az utóbbi esetében pedig valószínűségeken alapuló belső rendszerállapotok révén jutnak el a beszédprodukcióhoz. A statisztikai elvű módszerek egyik tipikus válfaja a gépi tanulóalgoritmusok szerkesztése és alkalmazása, ami a jelenlegi tudományos kutatások egyik legaktívabban prosperáló irányzataként tartható számon [4]. A szabályalapú módszerekkel szemben a statisztikai elvű

megoldások kétségtelen előnye, hogy nem mosódnak el a beszéd legfinomabb részletei sem (pl. hangmagasság, hangleyjtés, hangszín), ugyanakkor hátrányként könyvelhető el, hogy az esetleges hibajavítás meglehetősen nehéz, és rejtve maradnak olyan mennyiségi kapcsolatok és törvényszerűségek, melyek kulcsfontosságú szerepet játszanak a beszédszintézis fizikai hátterének átfogó megértésében. A magyarországi viszonyokat tekintve, a hazai kutatók az 1970-es években kapcsolódtak be a témakör tanulmányozásába. Az első magyar szövegfelolvasó rendszer, a HungaroVox 1980-ban készült el, majd folyamatos fejlesztések nyomán sorban követték egymást a ScriptoVox [5], a Brailab [6], a PC TALKER [7], a PCROBOT, a MultiVox [8], a ProfiVox [9], illetve a FlexVoice [10] rendszerek, melyek szépen kirajzolják a fejlődés ívét a nagyon robotos hangzású beszédétől a teljesen emberi hangzású, jól érthető beszédig.

A szövegfelolvasó rendszerek a beszédszintézis klasszikus ágát képviselik, ahol hosszú évtizedek során rendkívül sok tapasztalat és gazdag tudásanyag halmozódott fel, amit a szakirodalom számos közleménye is igazol [11]-[17]. Emellett azonban olyan területek is kezdenek egyre élénkebben előtérbe kerülni, melyek kevésbé kidolgozottak, és rengeteg nyitott probléma vár még megoldásra. Ide sorolható például az artikulációs beszédszintézis, ami az akusztikai produktum utánzását emberi hangminták helyett a hangképzés és artikuláció gépi leképezése révén próbálja megvalósítani. Ennek egyik technológiai vonulata a robotok beszédének előállításához szükséges artikulációs elektromechanikus beszéd-keltőkre irányuló kísérletezés [18]-[20], ami lényegében Kempelen Farkas szellemiségének modern kori adaptációja. Szintén a jövő tendenciáinak kedvez a gégtől a száj-, illetve orrnyílásig terjedő artikulációs csatorna, más néven vokális traktus modellezésére épülő beszédszintézis, ami főként vizuális információkra támaszkodik. Számos tanulmány hitelesen alátámasztja, hogy az emberi beszéd fiziológiai folyamatairól nyert vizuális információk nagymértékben elősegítik a beszédképzés komplex mechanizmusának megértését, és ezen keresztül a beszédszintézis módszereinek hatékony fejlesztését [21]-[28]. A napjainkban rendelkezésünkre álló radiológiai és monitorozó eljárások – úgymint mágneses rezonanciás képalkotás (MRI), komputertomográfia (CT), ultrahang (UH), elektropalatográfia (EPG), elektromágneses artikulográfia (EMA) vagy elektroglottográfia (EGG) – nélkülözhetetlen szerepet játszanak az akusztikai-artikulációs konverzió problémájának kezelésében. A felsorolt képalkotó eljárások közül az MRI, a CT és az UH a legközismertebb, hiszen mindennapi életünk során is találkozhatunk ezekkel az eszközökkel a klinikai gyógyászatban. A fentebb említett képalkotó és monitorozó technikák segítségével generált morfológiai és geometriai adatok felhasználásával maradéktalanul feltérképezhetők az adott beszédjelhez tartozó artikulációs mozgások. Nem triviális feladat azonban az artikuláció akusztikummal való összekapcsolása, azaz a vokális traktus morfológiai és geometriai adataira alapozott beszédprodukció megvalósítása. Ilyen jellegű kutatási eredmények már felfedezhetők a szakirodalomban [29]-[36], de ez a terület sok szempontból még a mai napig is nyitott. Az eddig publikált tanulmányok főképpen a vokális traktus geometriai modelljének megalkotására fókuszálnak, aminek alapját az esetek többségében MRI-, UH- vagy EMA-felvételek képezik [37]-[39], vagy pedig a beszédjeltől származó lényegi információkkal manipulálnak [40]-[41], a tökéletes minőségű gépi beszéd tényleges megvalósításához vezető út azonban bőven tartogat még kihívásokat. Ezenkívül az artikuláció geometriai és akusztikai jellemzői között fennálló fizikai kapcsolatok természetét és hátterét illetően is számos nyitott kérdés vár még megválaszolásra. A problémafelvetés aktualitását mutatja, hogy az artikulációs-akusztikai kapcsolatrendszer feltárása, illetve

gyakorlati leképezése alapvető fontosságú lehet például a klinikai célú beszédterápiában, a nem anyanyelvi nyelvtanulási tréningek kialakításában vagy a néma beszéd megszólaltatásához szükséges szintetizátorok konstrukciójában és fejlesztésében, ami szolgálhatja többek között a gégeeltávolításon átesett emberek rehabilitációját is.

A beszéd kutatás területén a fentebb listázott képalkotó eljárások közül főleg az UH- és MRI-technikákat tarthatjuk számon a vokális traktus vizuális leképezésének legkedveltebb eszközeiként. Az UH-felvételek sokféle szempont alapján nyerhetnek alkalmazást. Például, a videokeretek használatával különböző artikulációs hibák detektálhatók [42], fonetikai szempontok alapján vizsgálhatók a gyermekkori beszédhangzavarok [43], beszédfelismerő módszerek fejleszthetők nyelvkontúrkövetés révén [44], megbecsülhetők a magánhangzók formánsfrekvenciái, támogatva a némabeszéd-interfészek létrehozását [45], továbbá az UH-alapú stratégiák a beszéd- és nyelvterápiában is méltán helyet kaphatnak [46]. Emellett az MRI-felvételek alkalmazását is sokrétű célok tehetik indokolttá. Például, beszédprodukciónak valósítható meg neurális vokóderek segítségével [47], vizsgálhatók a dadogó alanyok hangképző szerveinek mozgásában megfigyelhető fluktuációk [48], tanulmányozható a nyelvizmok mechanikai kooperációja [49], kontúrszegmentáció valósítható meg a szájüregi szövetek határvonalainál gépi tanulás bevetésével [50], valamint elemezhető a vokális traktus alakja érzelmi töltetű beszéd közben [51].

A kutatómunka során érdemes figyelmet szentelni a különböző képalkotó eljárások segítségével előállított vizuális információk összehasonlítására és összehangolására is, hiszen a különböző forrásokból származó adatok szimultán elemzése tovább mélyítheti a beszéd artikulációs és akusztikai perspektíváihoz kapcsolódó tudást. A szakirodalom felsorakoztat néhány publikációt, melyek képalkotó forrásokból kinyert információk összeegyeztetése révén jutnak el döntő fontosságú megállapításokhoz [52]-[54]. Ez az irányvonal azonban még nincs teljesen kiforrott állapotban. A monitorozó technikák egyidejű komparatív alkalmazása ugyanis egyáltalán nem triviális feladat, mivel a források megfelelő és hiteles összeegyeztetése nagyon komoly és szakszerű anatómiai, geometriai, mérnöki és informatikai ismereteket igényel. A törekvés azonban gyakorlatilag elengedhetetlen a vokális traktus működésének részletes feltárásához, illetve az artikulációs-akusztikai összefüggések mélyebb megértéséhez.

A fentebbiek tükrében doktori kutatómunkám egyik fő célkitűzése a beszéd közben készült kétdimenziós UH- és MRI-felvételek radiális és négyszöges geometriáinak összehangolása volt, aminek alapját a felvételekre automatikus algoritmusokkal illesztett nyelv- és szájpadkontúrok képezték. Ezt a feladatot analitikus szabályszerűségekre és mesterséges intelligenciára támaszkodó megközelítésekben szándékoztam megvalósítani. Az analitikus irányvonalat követve olyan geometriai transzformációkat dolgoztam ki, melyek kölcsönösen egyértelműen és kétirányú módon összekapcsolják és egymásba ágyazzák a két forrás anatómiai környezeteit úgy, hogy a matematikai műveletek paramétereinek optimalizációja révén elérhető legyen az UH- és MRI-felvételek nyelv- és szájpadkontúrjai közötti lehető legjobb egybevághóság. A mesterséges intelligencia érvényesítéséhez gépi tanulóalgoritmusok alkalmazását irányoztam elő olyan neurális hálózatok megszerkesztésével, melyek az UH-nyelvkontúrokból kivont paraméterekre hagyatkozva realizálják az MRI-nyelvkontúrokból származó adatok betanulását.

Kutatásaim másik nagyobb témaköre az artikulációs beszéd szintézis kivitelezése volt, amihez szintén a fentebb említett UH- és MRI-felvételeket használtam fel. Első lépésben célul tűztam ki a szintézis alapjául szolgáló képalkotó források releváns geometriai ada-

tainak dinamikus kinyerését, melyek birtokában önálló beszédhangok, illetve folyamatos beszéd előállítását vettem tervbe. Ennek során ismét a mesterséges intelligenciát kívántam segítségül hívni. Olyan gépi tanulóalgoritmusok felépítésére törekedtem, melyek a vizuális geometriai jellemzőkből kiindulva végrehajtják az UH- és MRI-felvételek akusztikai jeleiből eredeztetett artikulációs paraméterek betanulását.

2. A felhasznált UH- és MRI-felvételek

2.1. A felvételek vizuális jellemzése

Vizsgálataim alapvető eszközeit olyan audiovizuális források képezték, melyek ultrahangos (UH), illetve mágneses rezonanciás képalkotási (MRI) eljárással készültek. A beszéd közben rögzített dinamikus felvételek képi formában megjelenítik a beszélő hangképző szerveit, miközben hallható a beszélő által kibocsátott akusztikus jel. A képi keretek sorozatán megfigyelhető az aktív hangképző szervek (ajkak, nyelv, lágy szájpad, gégefedő) folytonos mozgása és a passzív hangképző szervek (fogmeder, kemény szájpad, gége) statikus pozíciói. A hang mint beszédjel időbeli eltolódás nélkül igazodik a felvételek képkockáihoz, így pontosan követhetők és egymáshoz rendelhetők a beszéd artikulációs és akusztikai mozzanatai. Végeredményben tehát létrejön az adott bemondáshoz tartozó, időben szinkronizált hang- és képcsomag. Az UH- és MRI-felvételek kétdimenziós mozgóképek formájában álltak rendelkezésemre. A felvételek síkját, a 2.1. ábrának megfelelően, az emberi testet bal és jobb oldali részekre osztó függőleges szimmetriasík, az ún. középszagittális sík definiálja, amely lehetővé teszi a hangképző szervek kétdimenziós vetületi mozgásának, relatív elhelyezkedésének, illetve anatómiai szerkezetének részletes tanulmányozását.



2.1. ábra. Az emberi test középszagittális síkja. [55]

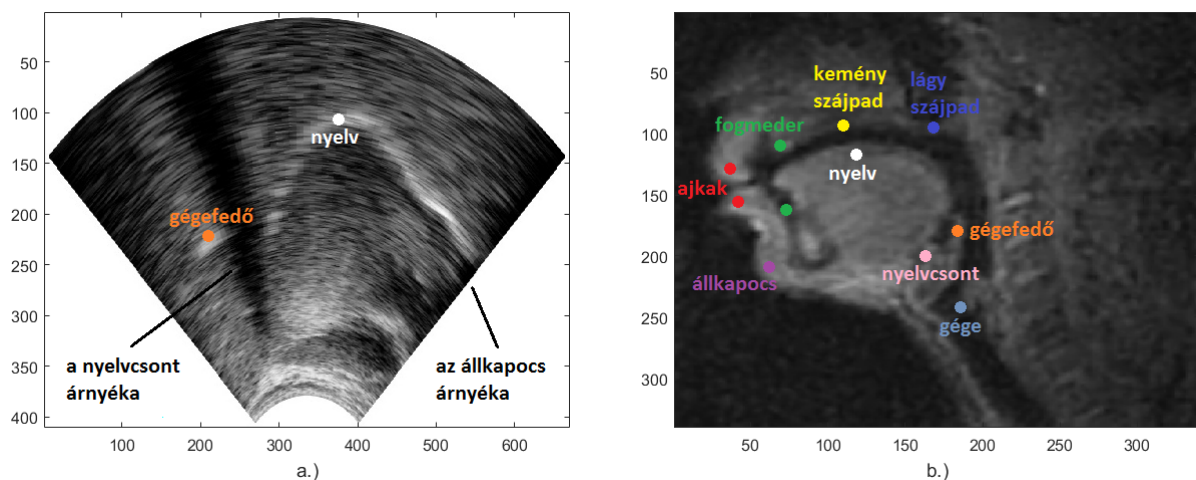
A két forrás azonban eltérő információtartalmat közvetít, mivel az UH- és MRI-keretek nem ugyanolyan módon ábrázolják a vokális traktus különböző elemeit. Az UH technikával többnyire csak a száj- és garatüreg egy része monitorozható kívülről, egy a beszélő álla alatt elhelyezett UH-fej alkalmazásával, melynek speciális elhelyezkedéséből adódóan az

UH-felvételeken csak a nyelv és a gégefedő mozgása jeleníthető meg. A többi hangképző szerv ellenben nem látható, hiszen az ajkak és a gége kívül esik az eszköz letapogatási zónáján, a kemény és lágy szájpad pedig nem detektálható közvetlenül az UH-hullámok sajátos szájüregi visszaverődései miatt. Mindemellett az UH-nyalábok nem képesek áthatolni a nyelvcsonton és az állcsonton, így a nyelv hátsó és elülső részeire árnyék vetül, aminek következtében a nyelv csak részlegesen mutatkozik meg. Ezzel szemben az MRI-felvételek az emberi fej egy jóval kiterjedtebb régióját képezik le, miután a teljes vokális traktus látható árnyékoló hatások nélkül. Ennek megfelelően a teljes nyelvszövet és annak mozgása, valamint a gégefedő, a lágy szájpad és az ajkak mozgása is megfigyelhető, illetve a kemény szájpad, a fogmeder és a gége helyzete is könnyedén azonosítható. Itt jegyzem meg, hogy a fogak sem az UH-, sem az MRI-kereteken nem láthatók.

2.1. táblázat. A felhasznált UH- és MRI-felvételek térbeli és időbeli felbontásához kapcsolódó paraméterek.

	Térbeli felbontás (mm/px)	Térbeli felbontás (px/mm)	Képméret (px × px)	Időbeli felbontás (keret/s)
UH	0,095	10,525	410 × 670	82
MRI	0,586	1,706	340 × 340	83

Mindkét képalkotó eljárás nagy előnye a jó térbeli és időbeli felbontás. A térbeli felbontást az határozza meg, hogy az adott kép 1 pixelje (px) hány mm-nek feleltethető meg a valódi térben, vagy – fordított irányú logikát követve – a valódi térben mért 1 mm-es távolság hány pixelt fed le a képen. A térbeli felbontáshoz kapcsolódik még a kép mérete, amit pixel mértékben is kifejezhetünk a kétdimenziós sík függőleges és vízszintes koordinátái mentén. Az időbeli felbontás a felvételen másodpercenként (s) felvillanó képkeretek számával jellemezhető. A 2.1. táblázat összefoglalja a felhasznált UH- és MRI-felvételek térbeli és időbeli felbontására vonatkozó adatokat.



2.2. ábra. A vokális traktus elemeinek elhelyezkedése egy statikus UH- (a.) és MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

A 2.2. ábra egy-egy statikus UH- és MRI-keretet mutat be, ahol a vokális traktus elemei különböző színekkel vannak felcímkézve, és az UH-képekhez viszonyítva szembevetve az MRI-képek vizuális megjelenésének részletgazdagabb struktúrája is. Látható, hogy az UH-felvételeken a nyelv hát és a gégefedő csúcspontja világos sávokként rajzolódik ki, a nyelvcsont és az állkapocs árnyékai pedig sötét zónák formájában észlelhetők. Az MRI-felvételeken az egyes hangképző szervek a levegő sötét tartományai és a testszövetek világosabb tartományai között létrejövő kontrasztok által különíthetők el egymástól. Ahogyan az a 2.2. ábrán is látható, a két kép egymásnak tükörképe, hiszen az UH-keret bal oldalán a nyelv hátsó része, jobb oldalán pedig a nyelv elülső része helyezkedik el, míg az MRI-kereten éppen fordított az elrendezés.

Az UH-felvételek elkészítése a technikai felszerelés szempontjából könnyebben véghezvihető, mivel egy hordozható, laptop-hoz csatlakoztatható UH-készülék tökéletesen megfelel a célnak, az MRI-felvételek megalkotása azonban kizárólag klinikai körülmények között kivitelezhető, amit sok esetben meglehetősen nehéz biztosítani. A 2.3. ábra bal oldali képen megfigyelhető a fejre illeszthető sisakhoz rögzített UH-fej, a jobb oldali kép pedig egy MRI-berendezést prezentál.



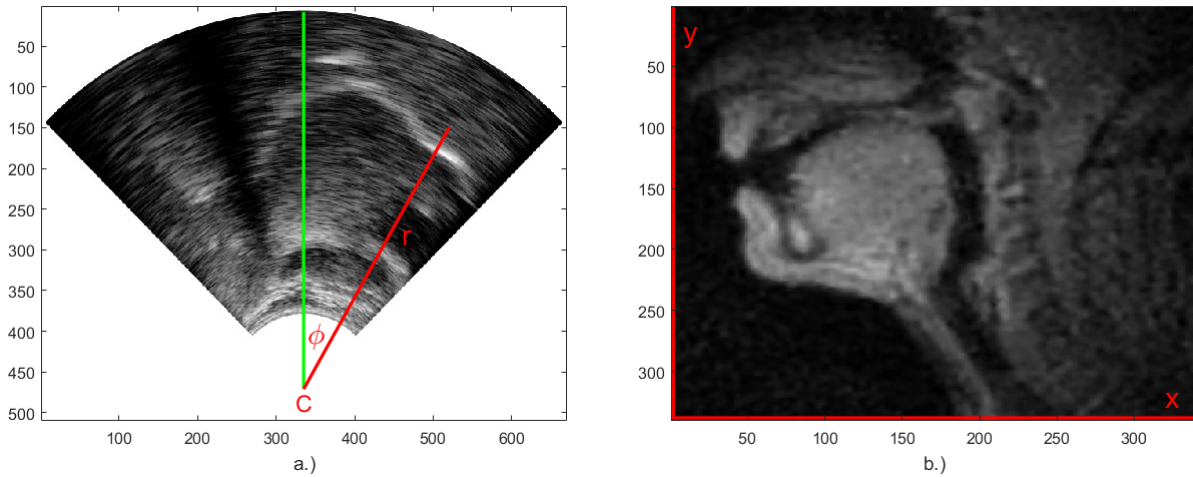
2.3. ábra. Dinamikus audiovizuális beszéd felvételek készítésére alkalmas UH-készülék (balra) és MRI-berendezés (jobbra). [56]

Az általam használt UH-csomag az MTA-ELTE Lendület Lingvális Artikuláció Kutatócsoportjának Micro rendszerével készült, az MRI-csomagot pedig a Dél-Kaliforniai Egyetem honlapján szabadon elérhető adatbázisból választottam ki [57]. Az UH-felvételeken különböző mondatok hangzanak el egy magyar női beszélő bemondásaiban, míg az MRI-felvételeken kitartott magánhangzókat, VCV struktúrákat, illetve szavakat artikulál egy amerikai angol férfi beszélő (V: magánhangzó, C: mássalhangzó).

2.2. A felvételek geometriai jellemzése

A 2.4.a ábra tanúsága szerint az UH-keretek egy körcikk által lefedett zónában vannak kifesztve, így az egyes képpontok pozíciójának jellemzésére kényelmesen alkalmazhatók a síkbeli polárkoordináták, melyek a kör C középpontjától mért r sugár (piros vonal), illetve a kép függőleges szimmetriatengelyéhez (zöld vonal) viszonyított ϕ szög segítségével definiálhatók. Ezáltal a képkocka síkjában tetszőlegesen felvett pixel helyzetét az $(r; \phi)$ koordinátapár egyértelműen meghatározza. A ϕ szög értéke minden UH-keret esetén -45°

és 45° között változhat, tehát a teljes régió 90° -os látószöggel pásztázható. A negatív szögértékek a függőleges szimmetriatengelyhez viszonyított bal oldali képtartomány, a pozitív szögértékek pedig a jobb oldali képtartomány pontjait lokalizálják. A 2.4.b ábrát alapul véve az MRI-keretek kezeléséhez szükséges legkomfortosabb vonatkoztatási rendszert egy síkbeli descartes-i koordináta-rendszer (piros tengelyek) adhatja, melynek origója a keret bal alsó sarkában helyezkedik el, és a képkocka kijelölt pontjának pozícióját az $(x; y)$ koordinátapár rögzíti. Az UH-felvételek tehát radiális geometriába vannak rendezve, az MRI-felvételek szerkezete pedig négyszöges geometriát követ.



2.4. ábra. Az UH-felvételek radiális geometriáját meghatározó $(r; \phi)$ polárkoordináták (a.), illetve az MRI-felvételek négyszöges geometriáját jellemző $(x; y)$ koordináták (b.). (Az ábra tengelyeinek mértékegysége pixel.)

3. Automatikus nyelvkontúrkövetés

A beszédhangok artikulációja során a nyelv alakja és pozíciója rendkívül fontos szerepet játszik, amit a nyelv felszíni formáinak tanulmányozásával lehet a legjobban leírni. Ebben nagy segítséget nyújthatnak az előző fejezetben említett UH- és MRI-felvételek, melyeken a nyelvszövet kétdimenziós vetületének határvonala nagyfokú pontossággal meghatározható a középszagittális síkban, ahol figyelemmel kísérhető a nyelv fel-le, illetve előre-hátra irányú mozgása. Az UH-keretek esetében a nyelvfelszín elmosódott, világos sávként azonosítható tartománya a nyelv és a felette lévő levegő határán kialakuló UH-hullámvisszaverődés eredményeként jön létre, így a nyelv felszíni vonalát a világos sáv alsó pereme jelöli ki (2.2.a ábra). Az MRI-kereteken a nyelvszövet világos tartománya és a sötét területként feltűnő levegő közötti kontraszt rajzolja ki a nyelv felületét (2.2.b ábra). Mindezeket figyelembe véve a nyelvkontúr követése mindkét esetben a nyelv hát vonalát meghatározó képpontok sorozatának megkeresését jelenti a releváns képtartományon belül.

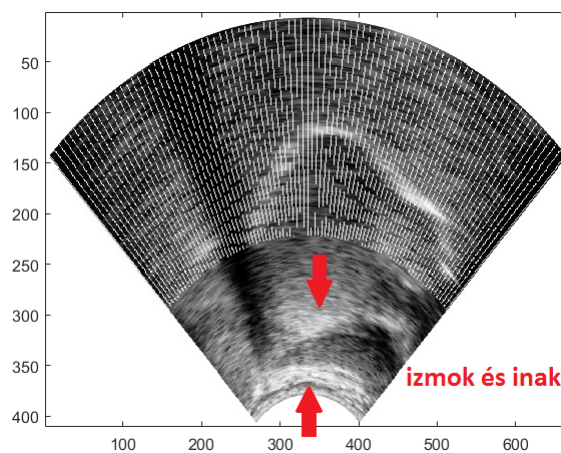
A nyelvkontúrkövetés elsődleges célja a különböző beszédhangokhoz tartozó nyelvállások és nyelv alakok statikus vagy dinamikus leírása, illetve a koartikuláció során létrejövő hangátmeneteket jellemző nyelvmozgások vizsgálata. A kvalitatív analízis mellett a nyelvkontúr a beszéd kvantitatív jellegű tanulmányozásának is jó kiindulópontja lehet, hiszen

a nyelvkontúrból származtatható számszerű értékek elősegíthetik az artikulációs modellek mélyebb megértését és fejlesztését.

A nyelvkontúr kijelölése történhet manuálisan vagy automatikus algoritmusok segítségével, bár az UH- és MRI-felvételeket alkotó képkockák számának ezres vagy akár százezres nagyságrendje indokolttá teszi az önműködő eljárások favorizálását a kézi erővel szemben [58]-[60]. A nyelvkontúr detektálásának hatékonyságát nagymértékben meghatározza a felvétel minősége, illetve a kontúrkövető algoritmus típusa, ezért gyakorlatilag elévülhetetlen ambíció a nyelvkontúrkövető programok állandó finomítása.

A vizsgálataim során egy olyan kontúrkövető algoritmust használtam fel és fejlesztettem tovább MATLAB környezetben, amely a dinamikus programozás technikáját alkalmazza [61]. A következőkben szeretném ismertetni az algoritmus főbb lépéseit egy-egy *o* hanghoz tartozó UH- és MRI-keret példáján keresztül. A nyelvkontúrkövető algoritmusok programkódjai az alábbi linken tekinthetők meg:

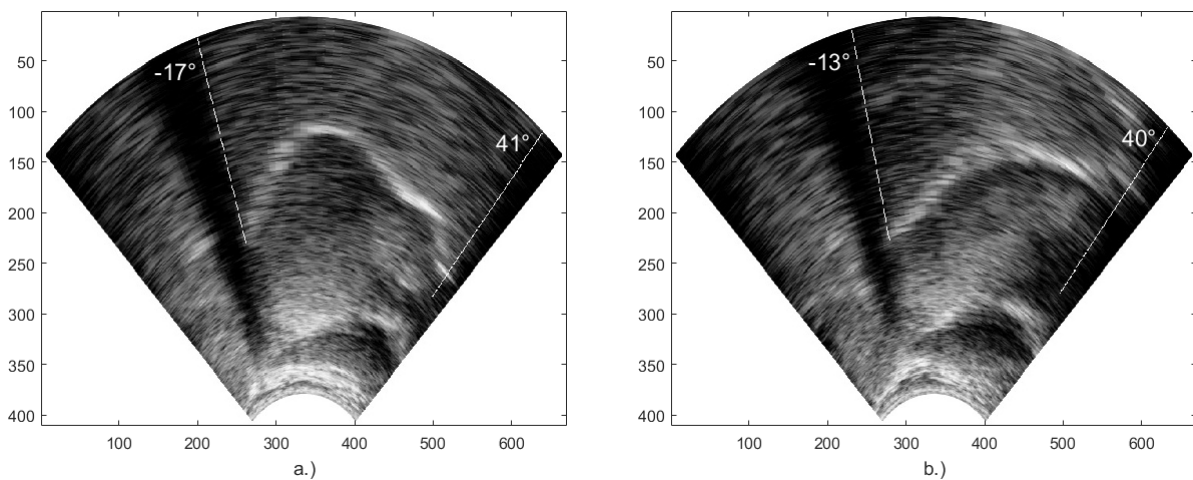
<https://drive.google.com/drive/folders/1QjT2GtClaadGzJdl6xMubWTD8AuRx?usp=sharing>



3.1. ábra. Az 1° -onként berajzolt sugárirányú metszetek egy *o* hanghoz tartozó UH-kereten. (Az ábra tengelyeinek mértékegysége pixel.)

Az algoritmus első szakaszában megkeressük a képek vízszintes dimenziója mentén azt a tartományt, ami a nyelvszövetre korlátozódik úgy, hogy a nyelvfelszín leghátsó és legelőlső pontjai a kívánt intervallum végpontjai közelében legyenek. Ezen vízszintes relevancia-tartomány kijelölése az UH- és MRI-felvételek esetében nem teljesen azonos mechanizmus szerint történik. Ennek oka az UH- és MRI-keretek vizuális megjelenésében megnyilvánuló különbségekre vezethető vissza. Ahogyan azt a 2.1. alfejezetben is részleteztem, az UH-felvételeken a nyelvcsont és az állkapocs árnyékot vet a nyelv hátsó és előlső részeire. A két sötét árnyéksáv helyzete azonban képkockáról képkockára változhat, hiszen beszéd közben egyrészt a nyelvcsont mozog, így hol kisebb, hol nagyobb mértékben takarja el a nyelv hátsó részét, másrészt a nyelv hegye is más-más pozícióban lehet, ezért az állkapocs is változó mértékben árnyékolja le a nyelv előlső részét. Ez azt eredményezi, hogy a nyelv felszíne mentén változó hosszúságú és változó elhelyezkedésű határvonal jelenhet meg az egyes kereteken, amit az algoritmusnak követnie kell. Ezért első lépésként a radiális geometriájú 2.4.a ábra *C* középpontjából kiindulva 1° -onként sugárirányú metszeteket képzünk a felvétel által definiált $[-45^\circ, 45^\circ]$ szögtartományban, majd metszetenként összegezzük az adott radiális vonal mentén található képpontok világosságértékeit, és megadjuk az így

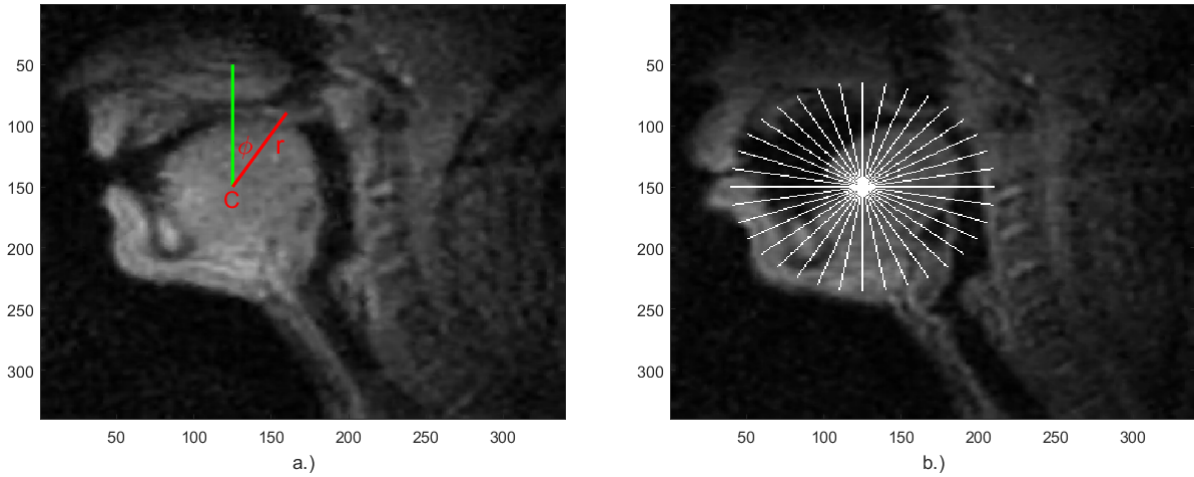
kiszámolt világosságértékek maximumát. Ennél a fázisnál – a fals és irreleváns eredmények elkerülése érdekében – gondosan be kell állítanunk a sugárszakasz hosszát, ügyelve arra, hogy a nyelv alatt található izmok és inak világos területeit kizárjuk a keresési tartományból. A 3.1. ábrán piros nyilak jelölik a nyelv alatti izmok és inak övezetét, és fehér vonalakkal van érzékeltetve a radiális metszetekkel letapogatott szektor. Ebből a szektorból olyan módon vágjuk ki a relevanciartományt, hogy meghatározzuk a nyelv hátsó és elülső részének pozícióját jellemző kezdő- és végszög értékét. Ez úgy történik, hogy az 1° -onként beosztott $[-45^\circ, 45^\circ]$ szögtartomány 91 metszetéből tetszőlegesen kiválasztunk egy metszetet (pl. az 50.-et), majd onnan indulva egyenként haladva lépegetünk előre és hátra a metszetek halmazában, és mindkét irányban megkeressük azt a metszetet, amely mentén a képpontok összegzett világossága már kisebb a maximális világosságérték harmadánál. Azt, hogy a maximális világosságérték hányad részét tekintjük megfelelőnek, tapasztalati úton állíthatjuk be. Általában a maximális világosságérték hatoda-harmada jó eredményre vezet. Az UH-felvételek esetében tehát a vízszintes relevanciartomány megkeresése, a síkbeli polárkoordináták rendszeréhez igazodva, a releváns szögtartomány kijelölését jelenti. A 3.2. ábra egy *o*, illetve egy *t* hanghoz tartozó UH-kereten szemlélteti az algoritmus által kalkulált kezdő- és végszögek értékeit.



3.2. ábra. Egy *o* (a.), illetve egy *t* (b.) hanghoz tartozó UH-kereten feltüntetett radiális metszetek a vízszintes relevanciartomány kezdő- és végszögének megjelölésével. (Az ábra tengelyeinek mértékegysége pixel.)

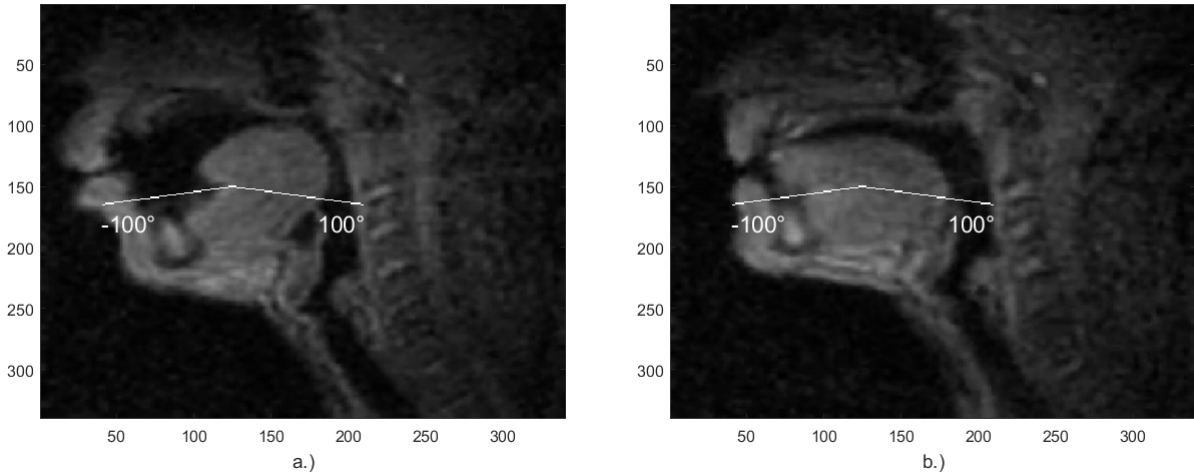
A relevanciartomány kezdő- és végszögét természetesen manuálisan is beállíthatjuk. Ekkor azonban a felvétel minden egyes kereten rögzített kezdő- és végszöggel fut az algoritmus függetlenül attól, hogy milyen mértékű területet foglal el a nyelvfelszín világos sávja. Ezen alternatívának akkor lehet jelentősége, ha az összes kereten azonos hosszúságú nyelvkontúrral szeretnénk vagy szükséges dolgozni.

Az MRI-felvételek vízszintes relevanciartományának keresése a főbb lépések tekintetében hasonló az UH-keretekre érvényes, fentebb bemutatott eljáráshoz. Annak ellenére ugyanis, hogy az MRI-képek alapvetően négyzetes geometriával rendelkeznek (2.4.b ábra), ez esetben is célszerű síkbeli polárkoordinátákkal dolgozni, mivel a 3.3.a ábra szerint az egyes pixelek pozíciója egy tetszőlegesen felvett C körközeponttól mért r sugár (piros vonal), illetve a C középpontból induló függőleges tengelyhez (zöld vonal) viszonyított ϕ



3.3. ábra. A négyszöges geometriájú MRI-kereteken definiált $(r; \phi)$ polárkoordináták (a.), és a 10° -onként berajzolt sugárirányú metszetek egy o hanghoz tartozó MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

szög segítségével is megadható. Ekkor az $(r; \phi)$ koordinátapár az UH-keretek struktúrájához hasonló geometriai környezetet hoz létre az MRI-képeken, ahol azonban a C középpont nem az áll alatt, hanem a testszöveteken belül van lokalizálva, a ϕ szög értéke pedig az UH-keretek $[-45^\circ, 45^\circ]$ tartományához képest a jóval szélesebb $[-180^\circ, 180^\circ]$ intervallumban változhat. Az MRI-felvételek vízszintes relevanciatarományának meghatározásakor nem merülnek fel olyan jellegű problémák, mint amiket az UH-kereteken tapasztalunk, hiszen az MRI-képeket nem terhelik árnyékoló effektusok. Ennek következtében a nyelvhat kontúrvonala minden kereten teljes egészében látható, és nem tűnnek fel olyan árnyéksávok, amiket az algoritmusnak dinamikusan követnie kellene, mivel ezúttal a nyelvcsont és az állkapocs relatív mozgása nem hagy nyomot a nyelvfelszín határfelületén.

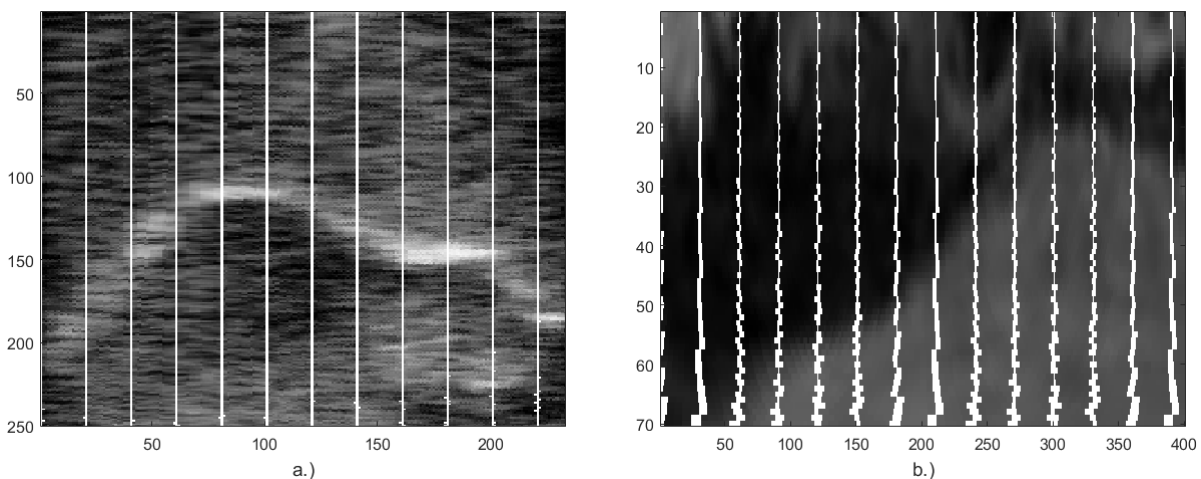


3.4. ábra. Egy o (a.), illetve egy t (b.) hanghoz tartozó MRI-kereten feltüntetett radiális metszetek a vízszintes relevanciataromány kezdő- és végzőgének megjelölésével. (Az ábra tengelyeinek mértékegysége pixel.)

Ezeket a körülményeket figyelembe véve a 3.3.a ábra C középpontjából indulva 10° -onként

sugárirányú metszeteket hozhatunk létre a polárkoordináta-rendszer origójának elhelyezkedéséből adódó $[-180^\circ, 180^\circ]$ szögtartományban, amit a 3.3.b ábra illusztrál. Ezt követően pedig vizuális megítélés alapján, az egyes képkockákon megjelenő különböző nyelvállások tanulmányozásával manuálisan beállíthatjuk a relevanciatartomány kezdő- és végszögét, ami minden keret esetén egységesen megfelelő lehet. A 3.4. ábra egy o , illetve egy t hanghoz tartozó MRI-kereten szemlélteti a manuálisan rögzített kezdő- és végszögek értékét.

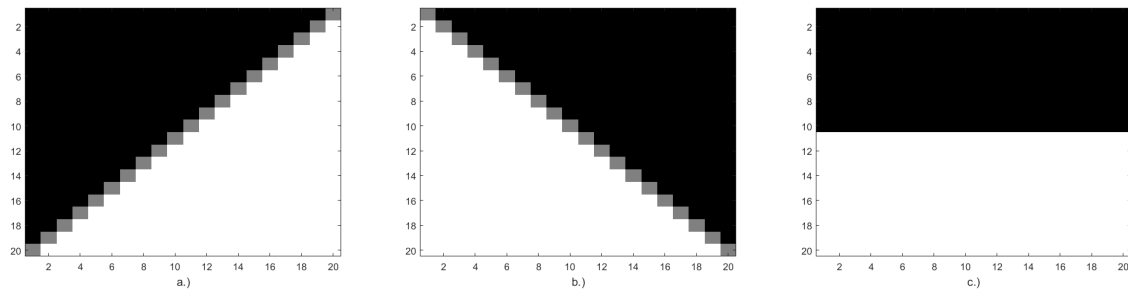
Az algoritmus második szakaszában az UH- és MRI-keretek vízszintes relevanciatartományának kezdő- és végszöge által határolt radiális szektorban ismét sugárirányú metszeteket alakítunk ki, melyeket függőleges oszlopokba rendezve egy mátrixstruktúrához jutunk. Ennek megfelelően a radiális metszetek mentén újra-mintavételezett képet egy descartes-i koordináta-rendszerben helyezünk el, vagyis a radiális geometriát négyyszöges geometriává alakítjuk át. A mátrixos szerkezet létrehozása nyomán áll elő a 3.5.a és 3.5.b ábrák, amik tehát a 3.2.a és a 3.4.a ábrákon felvett sugaras szektorok négyyszöges síkba történő kifeszítésével keletkeznek. A vizsgálataim azt mutatják, hogy az UH-felvételek esetében az $1/4^\circ$ -onként, az MRI-felvételek esetében pedig az $1/2^\circ$ -onként végrehajtott mintavételezés ideális, hiszen ekkor a mátrix szomszédos oszlopai között nem fordul elő két pixelnél nagyobb ugrás a kontúrban. Az áttekinthetőség kedvéért a metszeteket az UH-képen csak 5° -onként, az MRI-képen pedig csak 15° -onként ábrázoltam, amit a 3.5. ábra függőleges fehér vonalai demonstrálnak.



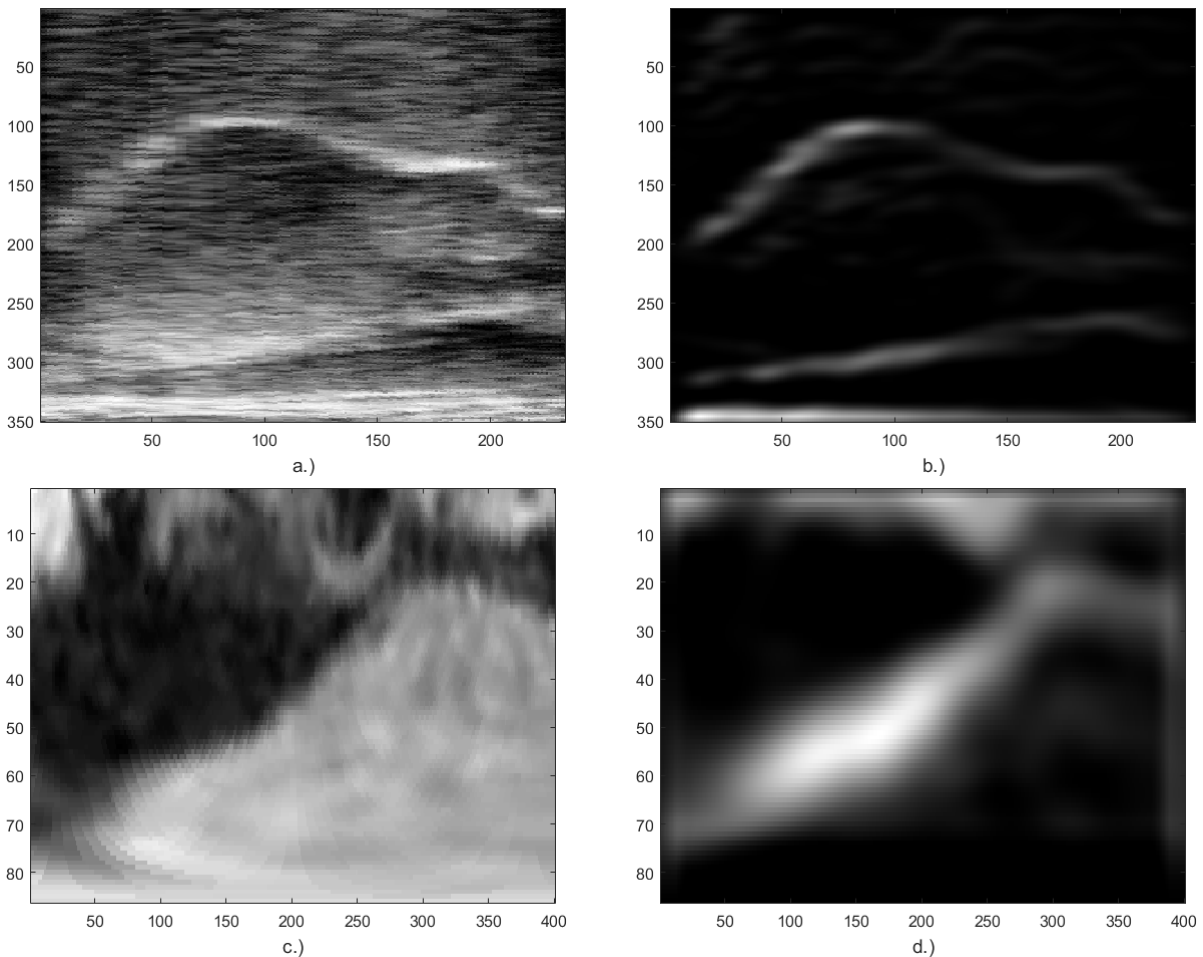
3.5. ábra. A radiális metszetek mátrixos szerkezetbe történő transzformációja egy o hanghoz tartozó UH- (a.) és MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

A fentebbieken túlmenően a felvételek előfeldolgozásának részét képezi a képképző eljárások tökéletlenségeiből adódó zajok és folytonossági hiányok csökkentése vagy megszüntetése is. A hibák redukálásának leghatékonyabb eszközei az élkiemelő és átlagoló operációk, amik matematikailag a konvolúció műveletével valósíthatók meg. Az élkiemelés a nyelvkontúr vonalának pontosabb és biztosabb megtalálását is támogatja, és az éleknek a különböző nyelvállások által meghatározott változatos orientációját figyelembe véve, három alapvető élkiemelő mátrixot értelmezünk, melyek konstrukcióját a 3.6. ábra vizualizálja. Az ábra tömbjein a fehér pixelek 1, a fekete pixelek -1, az átlók mentén sorakozó szürke pixelek pedig 0 értéknek felelnek meg. A fekete és fehér pixelekből álló mátrixblokkok pozíciói szimmetrikusan felcserélhetők annak függvényében, hogy az adott

típusú felvételen (UH vagy MRI) a nyelvkontúr a képen alulról felfelé haladva sötét-világos vagy világos-sötét képtartományok határvonala mentén húzódik.



3.6. ábra. Élkiemelő mátrixok grafikus megjelenítése három különböző konstrukcióban. (Az ábra tengelyeinek mértékegysége pixel.)

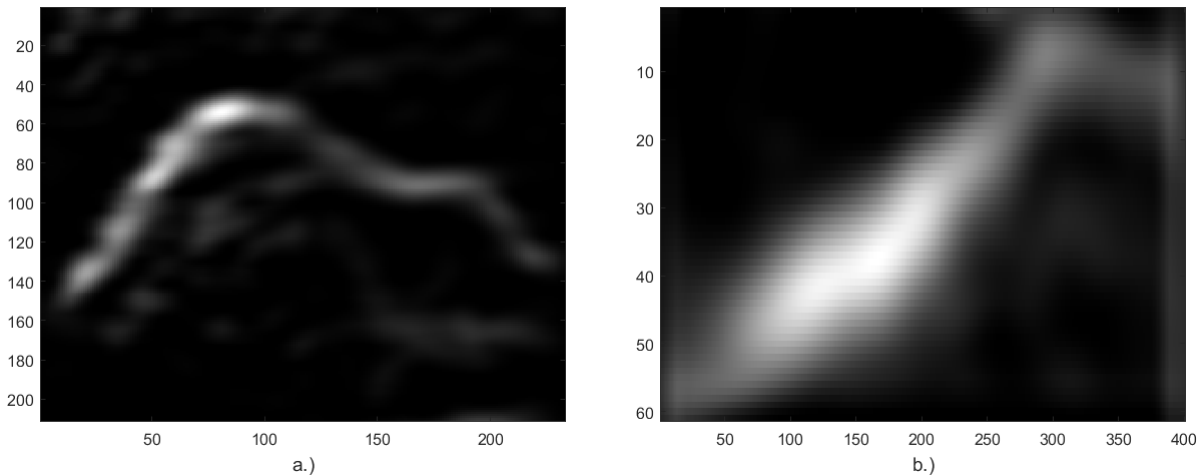


3.7. ábra. Élkiemelő és átlagoló operációk alkalmazása egy *o* hanghoz tartozó, négyzetes geometriájú UH- (a. és b.) és MRI-kereten (c. és d.). (Az ábra tengelyeinek mértékegysége pixel.)

A 3.5. ábra szerint transzformált négyzetes geometriájú keretekre együttesen alkalmazva a 3.6. ábra struktúráit követő élkiemelő mátrixokat, majd szinuszos és koszinuszos súlyfaktorok közbeiktatásával átlagoló operációkat végrehajtva eljutunk a 3.7.b és 3.7.d

ábrákhoz, melyek a 3.7.a és 3.7.c ábrák UH- és MRI-kereteinek élkkiemelt változatait prezentálják az *o* hanghoz rendelt nyelvállások esetén. Ezek után érthető, hogy azért van szükség a felvételek radiális irányú újra-mintavételezésére és négyszöges síkba való kifeszítésére, hogy elkerüljük a nyelvkontúr esetleges visszahajlásaiból adódó, a kép ugyanazon pixeloszlopában megjelenő, egynél több kontúrponthoz, amit az éldetektáló program nem tudna egyértelműen kezelni.

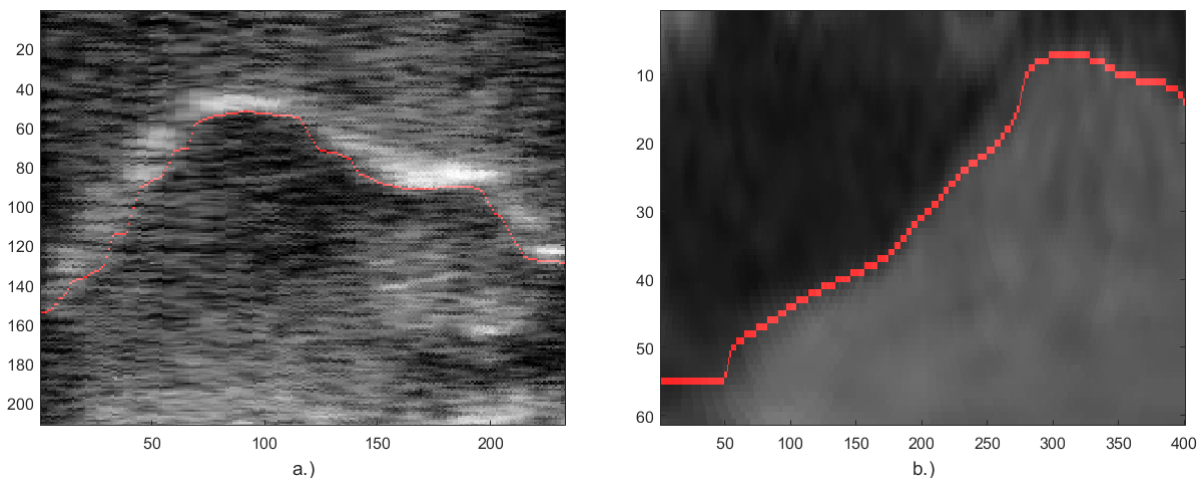
Az algoritmus harmadik szakaszában kijelöljük a képek függőleges dimenziója mentén azt a tartományt, ami magába foglalja a nyelvszövetet úgy, hogy a nyelvfelszín legfelső és legalsó pontjai a kívánt intervallum végpontjai közelében legyenek. Ezen függőleges relevanciatartományt olyan módon kell kimetszenünk a kép síkjából, hogy a vizsgált bementésekben előforduló legfelső és legalsó nyelvállás esetén se fusson ki a nyelvfelszín az intervallumból, illetve ne is kerüljenek be a látótérbe olyan képi elemek (pl. szájpád vagy nyelv alatti izmok és inak), melyek megzavarhatják a nyelvkontúr megbízható megállapítását. A függőleges relevanciatartományt keretről keretre haladva nem dinamikusán változtatjuk, hanem tapasztalati úton az adott bementés összes keretére egységesen állítjuk be. Jelen esetben ez azt jelenti, hogy a 3.7.b ábra mátrixszerkezetét követő UH-képek felső részéről 49, alsó részéről pedig 91 pixelsort eltávolítunk, és hasonlóan a 3.7.d ábra négyszöges elrendezése által adott MRI-képek felső részéről 14, alsó részéről pedig 11 pixelsort eliminálunk. A függőleges képlevágások eredményeként jönnek létre a 3.8. ábra UH- és MRI-keretei az *o* hang esetén. Ezáltal az algoritmus harmadik fázisának végén megkapjuk azokat az ablakolt képtartományokat, melyek a továbbiakban alapul szolgálhatnak a kontúrkereső programrészek futtatásához.



3.8. ábra. A függőleges relevanciatartomány kiszabása egy *o* hanghoz tartozó UH- (a.) és MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

Az algoritmus negyedik szakaszában a Viterbi-algoritmus [62] alapján kifejlesztett technika felhasználásával meghatározzuk az adott bementések keretein feltűnő nyelvkontúrok görbáját. A Viterbi-algoritmus egy dinamikus programozási technika, mely az adott probléma keretében értelmezett legvalószínűbb állapotok sorozatának feltárására szolgál. Az algoritmus megvalósításához szükséges kiindulópontot a 3.8. ábra szerint transzformált képek $m \times n$ méretű mátrixai adják, ahol az állapotokat az egyes pixelekhez rendelt világosságértékek testesítik meg, és a legvalószínűbb állapotsorozatot, azaz magát a nyelvkontúrt

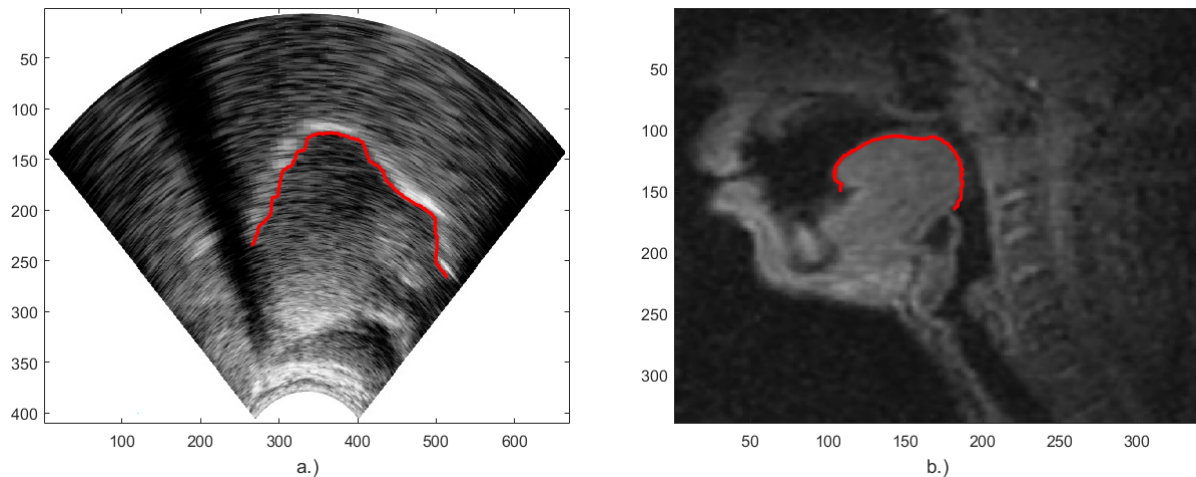
a maximális világosságú képpontok szekvenciája definiálja. Az analízis során a kép bal felső sarkából indulunk, és a jobb alsó sarokba érkezünk, miközben oszloponként haladva, az adott oszlop minden elemének helyén megvizsgáljuk, hogy az oda vezető lehetséges utak közül melyik rendelkezik a legnagyobb jósággal. Ezt technikailag úgy kivitelezük, hogy a kép j -edik oszlopának i -edik eleme körül értelmezünk egy néhány elemből álló, függőleges blokkot, és megkeressük a blokk legnagyobb világosságú képpontját, majd a megtalált maximumot hozzáadjuk a $j + 1$ -edik oszlop i -edik elemének értékéhez. Ezt az eljárást ciklikusan megismételjük a mátrix összes lehetséges j oszlop- és i sorindexére az $[1, n]$ és az $[1, m]$ intervallumokon belül. Az adott $(i; j)$ mátrixelem pozíciójába vezető legnagyobb jóságú út tehát megfelel a $j - 1$ -edik oszlop i -edik elemének környezetében megkeresett legnagyobb világosságú pixelből induló útnak. A folyamat eredményeként előáll egy olyan $m \times n$ méretű tömb, amely a fentebbi értelemben vett kumulált jóságokat tartalmazza az összes állapotra, vagyis az összes képpontra. Végül a kumulált jóságokat tartalmazó tömb utolsó oszlopának legvalószínűbb állapotából, azaz legnagyobb értékéből kiindulva, oszloponként egyesével visszafelé lépegetve, és a maximumokon végighaladva vissza tudjuk követni, hogy milyen úton kerültünk a legutolsó nyerő állapotba. Ezzel alakot ölt az ún. Viterbi-út, ami jelen esetben éppen a nyelvkontúr görbéje által kijelölt, legnagyobb világosságú pixelek sorozata mentén rajzolódik ki. A Viterbi-algoritmus kimeneteleként kapott nyelvkontúrokat a 3.9. ábra pirossal megjelölt pixelsorozata példázza az o hanghoz rendelt UH- és MRI-keretek esetében.



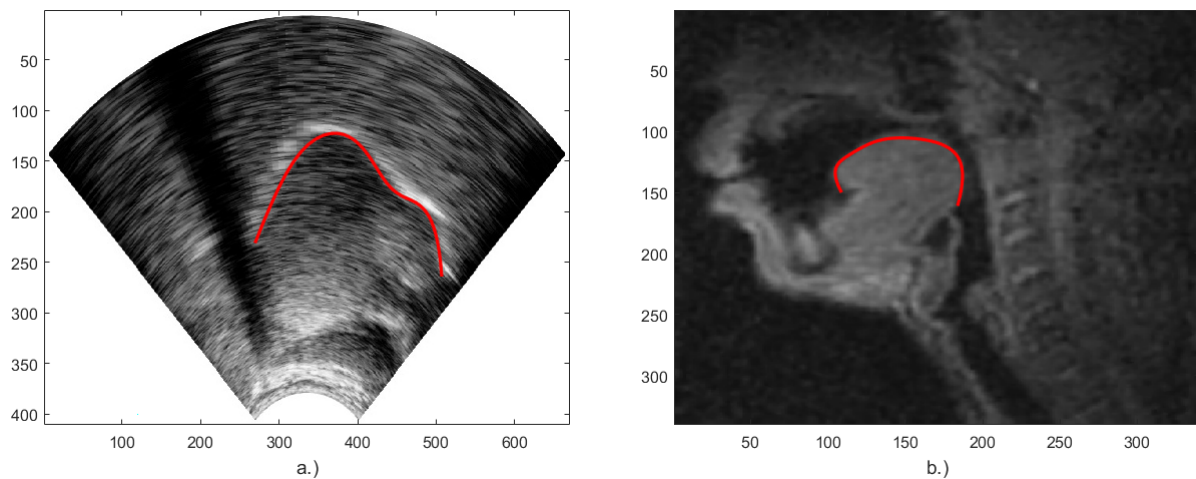
3.9. ábra. A Viterbi-algoritmus segítségével illesztett nyelvkontúrok egy o hanghoz tartozó UH- (a.) és MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

Az algoritmus ötödik szakaszában a 3.9. ábra szerinti négyszöges geometriában detektált nyelvkontúrokat visszaalakítjuk a radiális geometriához igazodó formába, amit az o hang esetében a 3.10. ábra UH- és MRI-keretei jelenítenek meg, ahol a kontúrok nyers görbéi egyelőre simítatlan módon illeszkednek a nyelv felszínére.

Az algoritmus hatodik és egyben utolsó szakaszában megtörténik a nyelvkontúr végleges alakjának realizálása, melynek során diszkrét koszinusztranszformáció segítségével elvégezzük a görbék finom simítását, hogy kiiktassuk az egyenetlenségeket. A végleges formájukat elnyert nyelvkontúrokat a 3.11. ábra UH- és MRI-keretein figyelhetjük meg az o hang esetében.



3.10. ábra. A radiális geometriában megjelenő nyelvkontúrok simítatlan görbéi egy *o* hanghoz tartozó UH- (a.) és MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)



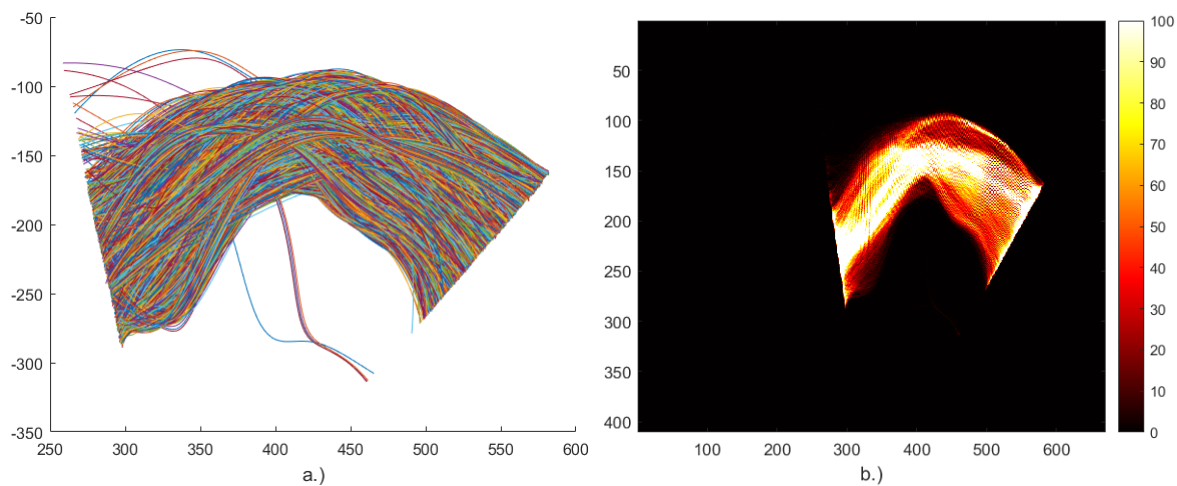
3.11. ábra. A radiális geometriában megjelenő nyelvkontúrok simított görbéi egy *o* hanghoz tartozó UH- (a.) és MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

4. Szájpadkontúrok azonosítása

Ahogy a 2.1. alfejezetben is részleteztem, számottevő különbségek észlelhetők az UH- és MRI-felvételek vizuális megjelenésében, ami többek között a szájpad anatómiai kontúrvonalának azonosíthatóságát is érinti. Láttuk, hogy az UH-képeken egyáltalán nem mutatkozik meg sem a kemény, sem a lágy szájpad, míg az MRI-képeken az ínyvitorlától a felső fogmederig kirajzolódik a szájpad teljes íve, és követhető a lágy szájpad mozgása, bár a kemény szájpad vonala olykor kevésbé határozottan, kissé elmosódottan bukkan fel a felvételeken. Ezen okoknál fogva különösen gondos megfontolásokat igényel a szájpadkontúr meghatározását célzó technika módszeres kidolgozása. Az UH- és MRI-keretek esetében merőben más elgondolások alapján fejlesztettem ki a szájpadkontúr lehető legpontosabb megkeresésére irányuló eljárást, amit a következő alfejezetekben szeparált módon taglalok. A szájpadkontúr-követő algoritmusok programkódjai az alábbi linken tekinthetők meg: <https://drive.google.com/drive/folders/1QjT2GtClaadGzJdl6xMubWTD0WD8AuRx?usp=sharing>

4.1. Az UH-szájpadkontúr megállapítása

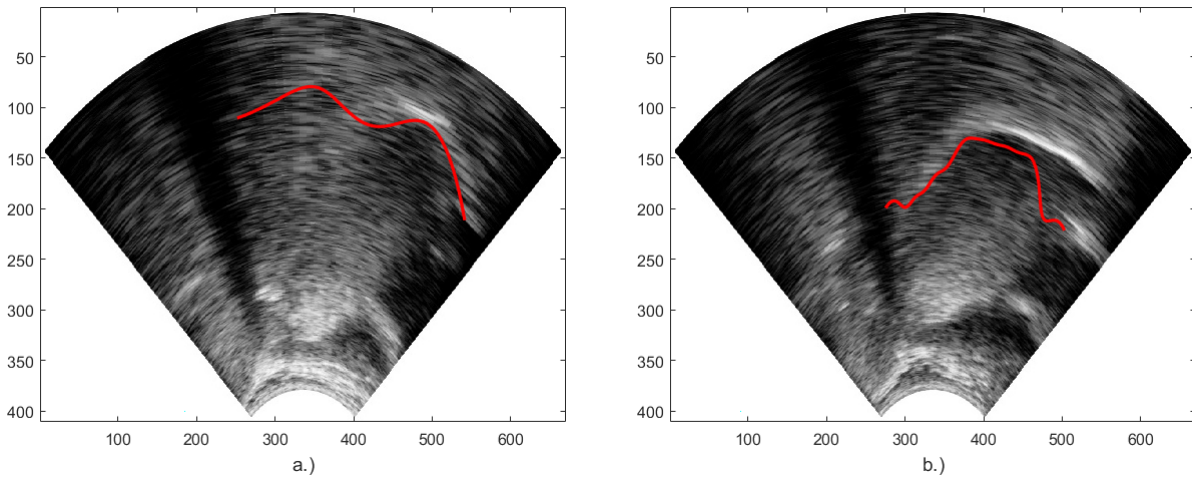
Az UH-szájpadkontúr elfogadható pontosságú becslésének az volt az alapvető koncepciója, hogy a felvételeken megtaláljam az artikuláció során a nyelvfelszín által érintett, legmagasabb helyzetben lévő szájüregi pontok halmazát, aminek nyomán valószínűsíthetjük a nyelv és a szájpada határvonalát. Ehhez természetesen olyan mássalhangzók vizsgálatára kell szorítkoznunk, melyek képzése során a nyelv biztosan érintkezik a szájpada palatális (kemény szájpada) és veláris (lágy szájpada) zónájával. Ez a feltétel a rendelkezésemre álló, különböző bemondásokat tartalmazó UH-csomag esetében automatikusan teljesül, hiszen a rögzített mondatokban szereplő mássalhangzók képzésekor (pl. *k*, *l*, *s*, *t*) a nyelv más-más helyeken kerül kontaktusba a szájpada ívével. Így a szájpada kontúrjának kirajzolását lényegében egy szélsőérték-keresési probléma megoldásaként valósítottam meg. Ehhez elsőként egy közös koordináta-rendszerben összegyűjtöttem a 40 UH-bemondás összes keretéhez tartozó 11 111 darab nyelvkontúr görbét, és ezzel párhuzamosan megalkottam az összes nyelvkontúr lenyomata által kialakuló sűrűségterképet is, melynek minden pixelje egy olyan valószínűségi mérték szerint kap értéket, hogy az adott képpont milyen gyakorisággal fordul elő lehetséges nyelvkontúrponként. Az összes nyelvkontúr halmazát, illetve a görbeseregnek megfelelő sűrűségterképet a 4.1. ábra demonstrálja. A 4.1.a ábrán látható, hogy a nyelvkontúrok vízszintes relevanciatartományának kezdő- és végszögét – a manuális opciót választva – minden kereten egységesen ugyanolyan értékekre állítottam be, hogy azonos hosszúságú görbékkel dolgozhassak (3. fejezet). A 4.1.b ábra jobb oldali színskálája alapján leolvasható, hogy a térkép adott pontja milyen gyakorisággal vesz részt a leképezésben, amennyiben a színskálán alulról felfelé haladva növekszik a képpontok érintésének gyakorisága, azaz a legsötétebb pixeleknél a legkisebb (nulla), a legvilágosabb pixeleknél pedig a legnagyobb az előfordulási valószínűség.



4.1. ábra. A 40 UH-bemondás összes keretéhez tartozó 11 111 darab nyelvkontúr halmaza (a.), illetve az összes nyelvkontúr által alkotott görbeseregnek megfelelő valószínűségi sűrűségterkép (b.). (Az ábra tengelyeinek mértékegysége pixel.)

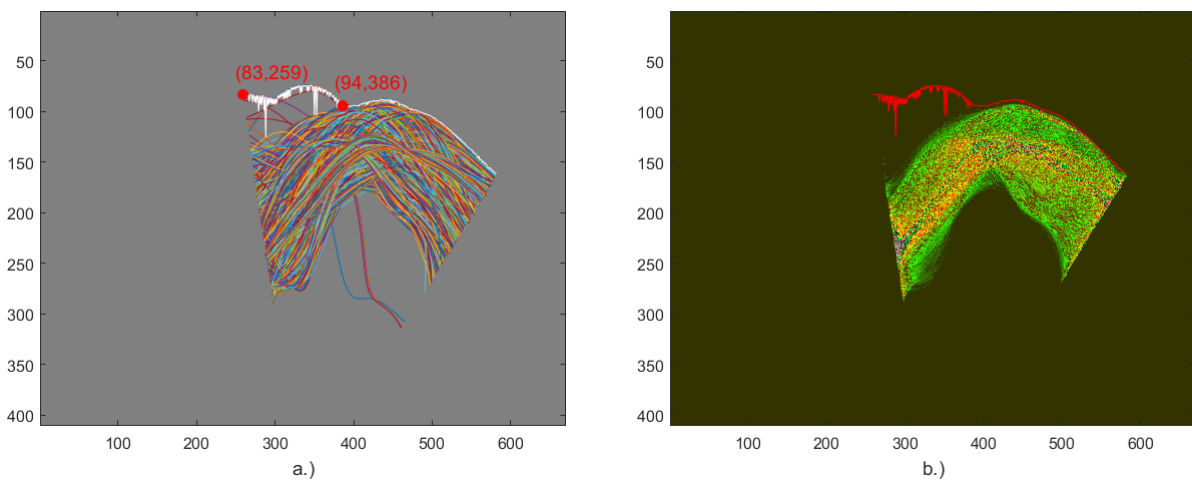
A szájpada kontúr felvételéhez azt a stratégiát vettem be, hogy a 4.1. ábrán illusztrált görbesereg és sűrűségterkép felhasználásával megkonstruáljam a nyelvkontúrok halmazának felső burkológörbét. Ez a feladat nem teljesen triviális, hiszen a kontúrkövető algoritmus futtatása során előfordulhatnak természetes hibák, melyek például egy adott

keret nem megfelelő képi minőségéből adódhatnak. Ezek általában egyáltalán nem vagy csak mérsékelten korrigálhatók a kontúrkereső program futásának eredményét meghatározó paraméterek módosításával. Emellett találkozhatunk olyan funkcionális hibákkal is, melyek a program paramétereinek helytelen beállításából fakadhatnak.



4.2. ábra. Természetes (a.) és funkcionális (b.) hibákból eredő fals UH-nyelvkontúrok. (Az ábra tengelyeinek mértékegysége pixel.)

Természetes és funkcionális kontúrillesztési hibákból eredő fals nyelvkontúrokat példáz a 4.2. ábra, ahol az a.) kereten megfigyelhető, hogy a kép túlságosan kontrasztszegény, így az algoritmus ideális beállítások mellett sem képes biztonsággal megtalálni a nyelv felszíni vonalát, a b.) képkockán azonban szabad szemmel is egyértelműen kivehető a nyelvfelszín világos sávja, ezért a paraméterek helyes konfigurációjával helyreigazítható a nyelvkontúr görbéje.



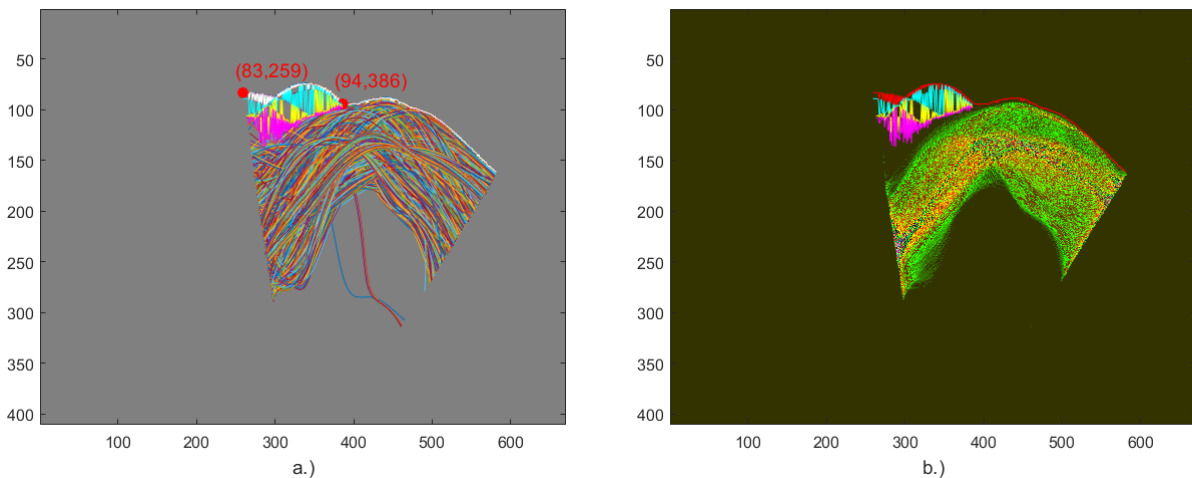
4.3. ábra. Az első közelítésben kapott burkológörbe az összes nyelvkontúr halmazán (a.) fehérrel, illetve a sűrűségterképen (b.) pirossal megrajzolva. (Az ábra tengelyeinek mértékegysége pixel.)

A fentebbiek értelmében munkám során természetesen arra törekedtem, hogy kontúrkereső algoritmusunk a lehető legkevesebb funkcionális hibával, ideális módon működjön,

a természetes hibákból származó nyelvkontúrok azonban egyelőre részét képezték a vizsgálandó görbeseregnek. Ezért a szájpadkontúr illesztésekor elengedhetetlen volt ezen fals nyelvkontúrok ignorálása. A szájpadkontúr elfogadható pontosságú közelítésére egy komplett algoritmust dolgoztam ki, melynek fázisait a következő szakaszokban ismertetem.

Az algoritmus első lépésében az összes nyelvkontúr által leképezett sűrűségterkép mátrixán oszloponként megkerestem az első egy nullától különböző mátrixelemet, ami az adott oszlopban éppen azt a legfelső pontot adja meg, ahol már megjelenik valamelyik nyelvkontúr pontja. Az ily módon megtalált pontok sorozatát a 4.3. ábra vizualizálja, ahol az összes nyelvkontúr halmazán fehérrel, a sűrűségterképen pedig pirossal van megrajzolva az első közelítésben kapott burkológörbe.

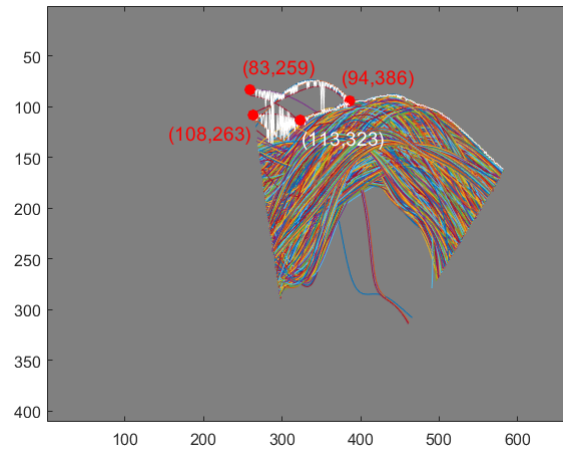
Az UH-felvételek elemzése során szerzett tapasztalataim szerint a nyelv hát közepső és elülső részein eléggé stabilan lokalizálható a korrekt kontúrvonal, a nyelv hátsó részén azonban gyakrabban fordulhatnak elő különböző képi anomáliák, amik fals nyelvkontúrokhoz vezetnek. Ez a helyzet áll elő a 4.3.a ábra görbeseregének esetében is, így a piros körökkel megjelölt pontok közötti szakaszon, azaz a képmátrix 259. és 386. oszlopai által határolt intervallumban nem reális az első közelítésben kapott görbe, és további finomításokra van szükség. Ezért az algoritmus második lépésében a sűrűségterkép 259. és 386. oszlopai közötti tartományban oszloponként megkerestem a második, harmadik és negyedik nullától különböző mátrixelemet is, szakaszosan lejjebb tolva ezzel a hátsó burkológörbe pontjait, amint azt a 4.4. ábra kék, sárga és rózsaszín görbéi is érzékeltetik.



4.4. ábra. A szájpadkontúr hátsó szakaszának másod- (kék), harmad- (sárga) és negyedrendű (rózsaszín) közelítése az összes nyelvkontúr halmazán (a.), illetve a sűrűségterképen (b.). (Az ábra tengelyeinek mértékegysége pixel.)

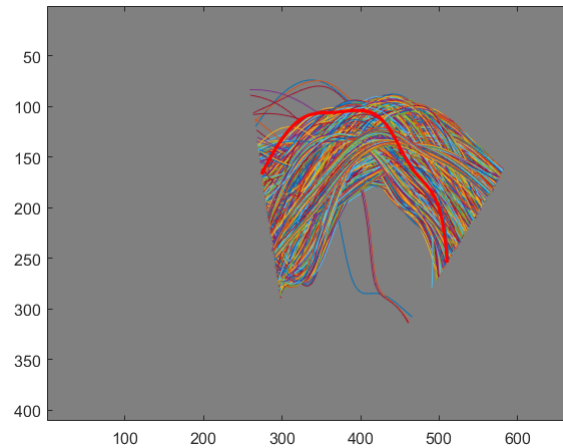
A 4.5. ábrán a (94;386) koordinátájú pontban egyesítettem a 4.4. ábra elsőrendű közelítésben nyert görbéjének jobb oldali részét (fehér) és a negyedrendű közelítésben keletkező hátsó görbeszakaszát (rózsaszín), aminek nyomán közelebb kerültünk a szájpadkontúr reális vonalához, hiszen a képmátrix 323. és 386. oszlopai közötti tartományban az alsó fehér görbe most már jobban illeszkedik a sűrű sokaság által megszabott peremhez.

Amint az a 4.5. ábrán is észrevehető, a szájpadkontúr letapogatását célzó törekvés egyre nehezebbé és bizonytalanabbá válik, ahogyan a görbe rögzítésével haladunk a nyelv hátsó részének irányába, mivel a görbesereg egymásra tevődő és egymást elfedő nagyszámú eleme gátolja a korrekt és fals nyelvkontúrok világos megkülönböztetését. Ezen



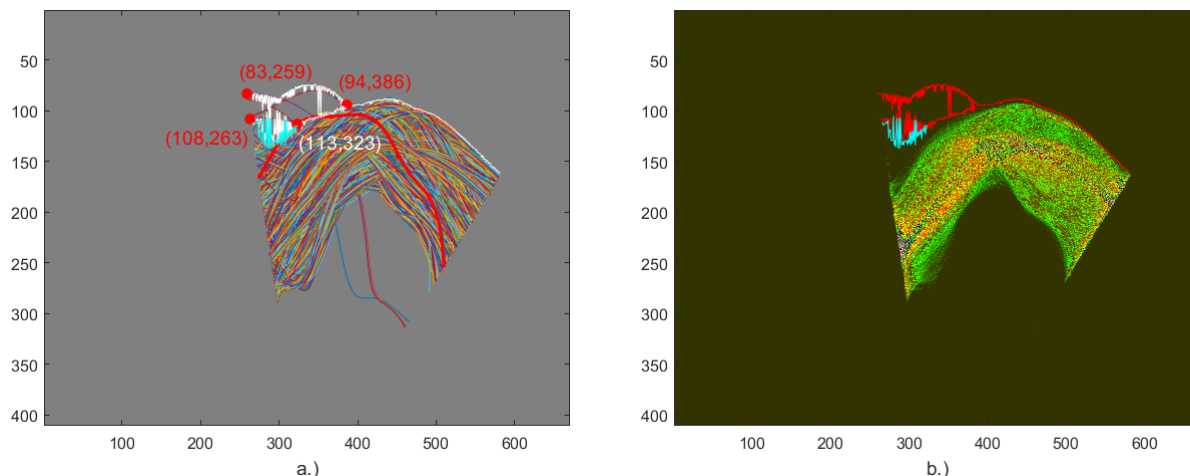
4.5. ábra. A szájjpadkontúr első- (felső görbe) és negyedrendű (alsó görbe) közelítése az összes nyelvkontúr halmazán. (Az ábra tengelyeinek mértékegysége pixel.)

a ponton tehát nem egyértelmű, hogy a 4.5. ábra képmátrixának 263. és 323. oszlopai közötti intervallumban melyik nyelvkontúr lehet mérvadó a görbe további finomításában. Ebből kifolyólag az algoritmus harmadik lépésében szükségessé vált a nyelvkontúr-halmaz felső sávjában lévő görbéinek vizsgálata, melynek során a görbesereg bal oldali határpontjai mentén felülről lefelé haladva, a sokaságból a nyelvkontúrokat egyenként kiragadva ellenőriztem az adott görbe helyességét. A procedúrát addig folytattam, amíg meg nem találtam az első olyan nyelvkontúrt, ami valóságos nyelvállást írhat le, és így etalonnak tekinthető a szájjpadkontúr hátsó részének megállapításában. A pirossal kiemelt etalongörbét a 4.6. ábra jeleníti meg.



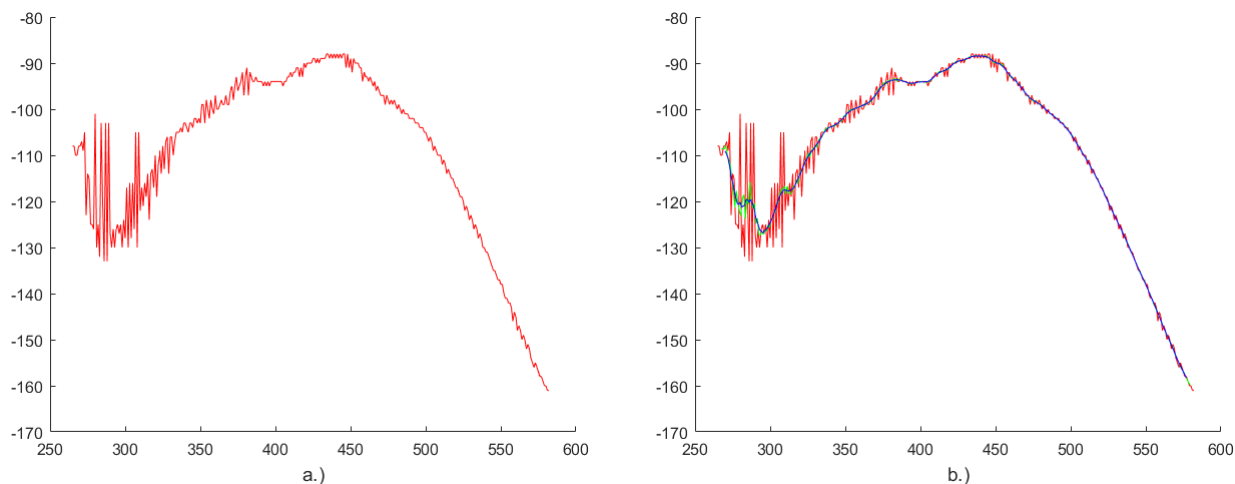
4.6. ábra. A pirossal kiemelt etalongörbe az összes nyelvkontúr halmazán. (Az ábra tengelyeinek mértékegysége pixel.)

Az etalongörbe fixálása lehetővé tette, hogy a szájjpadkontúr hátsó szakaszát célirányos transzformációknak vethessem alá. Ennek értelmében az algoritmus negyedik lépésében a sűrűségkép 263. és 323. oszlopai közötti tartományban oszloponként megkerestem az ötödik nullától különböző mátrixelemet is, még lejjebb csúsztatva ezzel a hátsó burkológörbe pontjait, amit a 4.7. ábra kék görbéje demonstrál.



4.7. ábra. A szájpadkontúr hátsó szakaszának ötödrendű (kék) közelítése az összes nyelvkontúr halmazán (a.), illetve a sűrűségképen (b.). (Az ábra tengelyeinek mértékegysége pixel.)

A 4.7.a ábra (113; 323) koordinátájú pontjában összekapcsoltam a negyedrendű közelítésből eredő görbe jobb oldali részét (fehér) és az ötödrendű közelítésben kialakuló hátsó görbeszakaszt (kék), majd az algoritmus ötödik lépésében kiszámítottam az így létrejövő görbe egyszeres és kétszeres gördülő átlagát. Ez jelen esetben azt jelenti, hogy végiggördítettem egy 5 szomszédos kontúrpontról álló intervallumot a görbe teljes hosszán úgy, hogy a görbe kezdőpontjából indulva minden ütemben eggyel arrébb toltam az 5 pontos intervallumot, és kiszámoltam az aktuális 5 pont függőleges koordinátáinak átlagát. Ezt az eljárást az egyszeresen átlagolt görbére megismételttem, aminek eredményeként előállt az eredeti görbe kétszeres gördülő átlaga.



4.8. ábra. A szájpadkontúr negyed- és ötödrendű közelítéséből származó görbék egyesítése (a.), illetve az eredeti görbe (piros), valamint annak egyszeres (zöld) és kétszeres (kék) gördülő átlagai (b.). (Az ábra tengelyeinek mértékegysége pixel.)

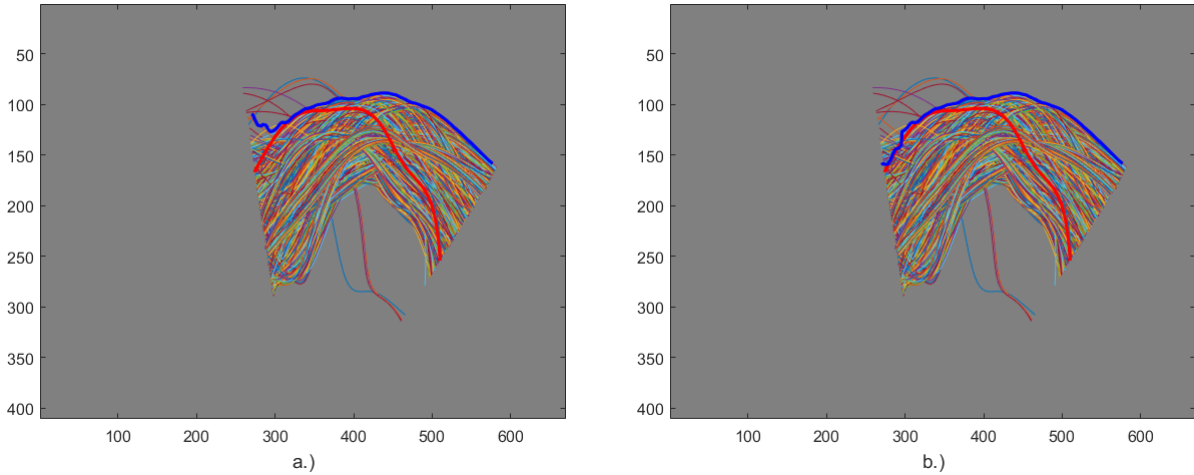
A 4.8.a ábrán látható a fentebb említett negyed- és ötödrendű közelítésből származó görbék fúziója, a 4.8.b ábrán pedig nyomon követhető az eredeti görbe (piros), illetve annak egyszeres (zöld) és kétszeres (kék) gördülő átlagai, és az is világosan kivehető, hogy

az átlagolás hatása egyfajta simításként értékelhető.

A 4.8.b ábrán feltüntetett simított görbe elhelyezkedését az összes nyelvkontúr halmaza a 4.9.a ábra mutatja be. Megfigyelhető, hogy a kék görbe felcsapódó far része egyelőre nem igazodik a piros etalongörbe hátsó szakaszához, ezért az algoritmus hatodik lépésében a szájpadkontúr leghátsó szakaszát speciális lineáris transzformációnak vettem alá. Ennek során az N pontból álló felcsapódó far rész pontjainak függőleges koordinátáit jelölő $Y = \{y_1, y_2, y_3, \dots, y_N\}$ halmaz elemeit felváltottam az

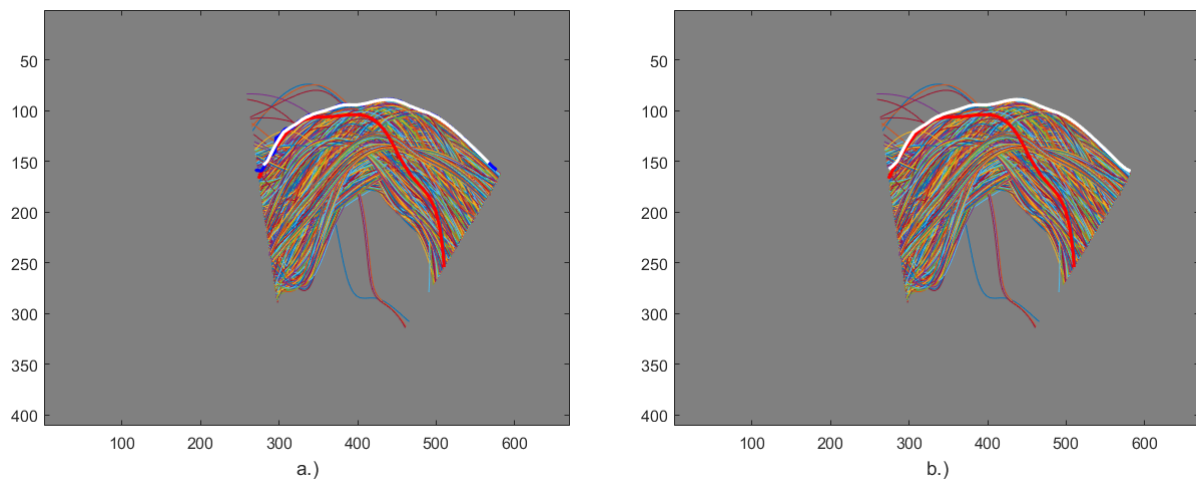
$$y'_n = y_n + 2N - 1,5n \quad (4.1)$$

konverzióval előállított $Y' = \{y'_1, y'_2, y'_3, \dots, y'_N\}$ elemekkel, ahol $n = 1, 2, 3, \dots, N$. A (4.1) összefüggés szerkezetének megalkotásakor nem standard módon rögzített analitikus formulákkal dolgoztam, hanem többkörös próbálgatások útján kristályosodott ki a transzformáció matematikai alakja azon célkitűzés mentén, hogy a görbe felcsapódó far része ideális mértékben kompenzálva legyen úgy, hogy a lehető legjobban illeszkedjen az etalongörbe hátsó szakaszához. A transzformáció eredményeképpen létrejövő kontúrgörbét a 4.9.b ábra illusztrálja.



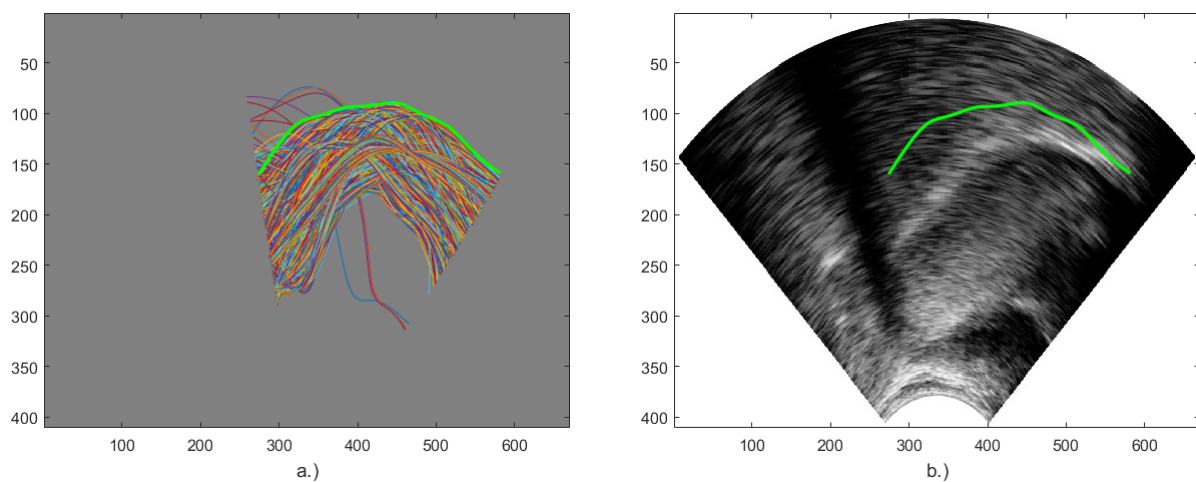
4.9. ábra. A szájpadkontúr kékkel megrajzolt kétszeres gördülő átlaga (a.), illetve a görbe felcsapódó far részének lineáris transzformációja nyomán létrejövő kontúr (b.). (Az ábra tengelyeinek mértékegysége pixel.)

A 4.9.b ábrán látható szájpadkontúr egyenetlenségeinek kiküszöbölése érdekében az algoritmus hetedik lépésében ismét gördülő átlagot alkalmaztam, de mivel a 4.8.a ábra görbéjéhez viszonyítva a kontúrt ezúttal sokkal mérsékeltebb ugrások törik meg, elegendő volt egy 20 szomszédos kontúrpontról álló intervallumot végigfuttatni a görbe teljes hosszán, amit – a korábbi átlagképzésekhez hasonlóan – a görbe kezdőpontjából indulva minden ütemben eggyel arrébb tolva, kiszámoltam az aktuális 20 pont függőleges koordinátáinak átlagát. Ezzel lényegében eljutottam a 4.8.a ábra görbéjének háromszoros gördülő átlagához, amit a 4.10.a ábra fehér görbéje vizualizál. Észrevehető, hogy a görbe a korábbi átlagolások eredményeihez képest a nagyobb léptékű átlagolás következtében erőteljesebben megrövidült, ezért elsőrendű extrapolációt alkalmazva kiterjesztettem a görbét az összes nyelvkontúr halmaza által meghatározott tartomány kezdő- és végpontjáig, hogy a lehető leghosszabb szakaszon rendelkezésre álljon a szájpadkontúr. Az extrapolált görbét a 4.10.b ábra szemlélteti.



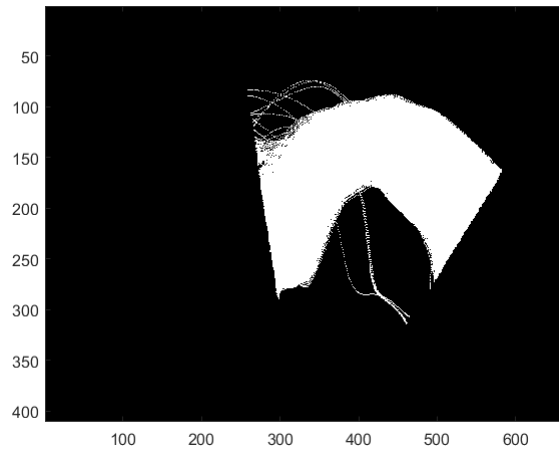
4.10. ábra. A szájpaddocktúr fehérrel megrajzolt háromszoros gördülő átlaga (a.), illetve annak lineáris extrapolációval meghosszabbított változata (b.). (Az ábra tengelyeinek mértékegysége pixel.)

Az algoritmus nyolcadik és egyben utolsó lépésében diszkrét koszinusztranszformáció segítségével utósimítást hajtottam végre a 4.10.b ábra görbéjén. Ezzel alakot öltött az UH-szájpaddocktúr végleges formája, amit a 4.11. ábra zöld görbéi elevenítenek meg az összes nyelvkontúr halmazán, illetve egy nyugalmi nyelválláshoz tartozó UH-kereten.



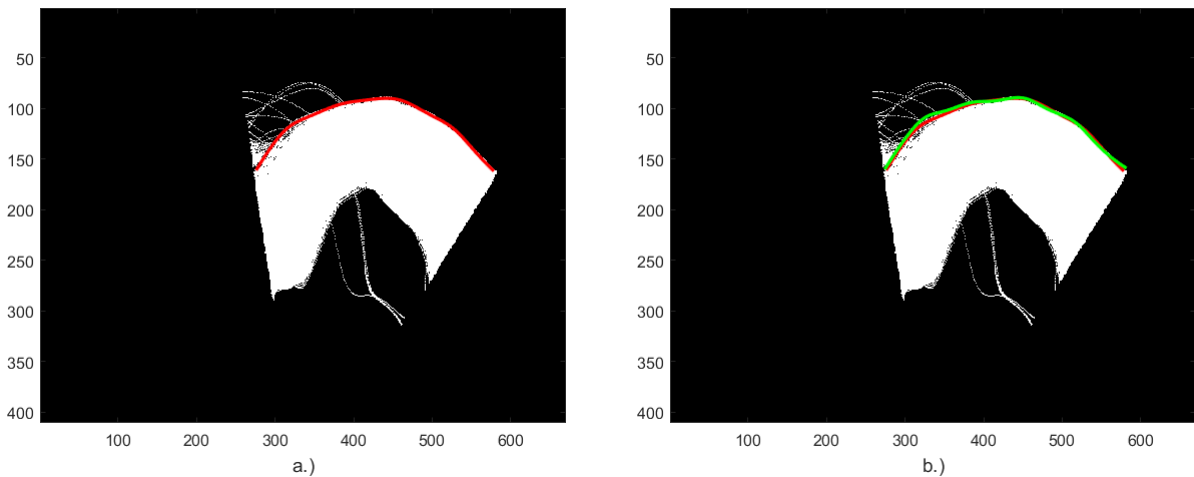
4.11. ábra. A diszkrét koszinusztranszformációval simított szájpaddocktúr végleges alakja az összes nyelvkontúr halmazán (a.), illetve egy nyugalmi nyelválláshoz tartozó UH-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

Eredményeim kontrollálása céljából a fentebb részletezett algoritmus lépéseitől független úton is megpróbáltam megkeresni a szájpaddocktúrvonalát. Ehhez az összes nyelvkontúr által adott görbesereget egy olyan bináris sűrűségterképen ábrázoltam, amely csak 0 és 1 értékeket vehet fel aszerint, hogy a képmátrix adott pontját érinti-e bármely nyelvkontúrnak bármely pontja. Ennek értelmében a nyelvkontúrok által lefedett pixeleket 1 értékek jellemzik, a fennmaradó képpontokat pedig 0 értékekkel látjuk el, aminek folytán, a 4.12. ábrával összhangban, a kép fekete háttéréből fehér tartományként emelkedik ki az összes nyelvkontúr halmaza.



4.12. ábra. Az összes nyelvkontúr által alkotott görbeseregnek megfelelő bináris sűrűségkép. (Az ábra tengelyeinek mértékegysége pixel.)

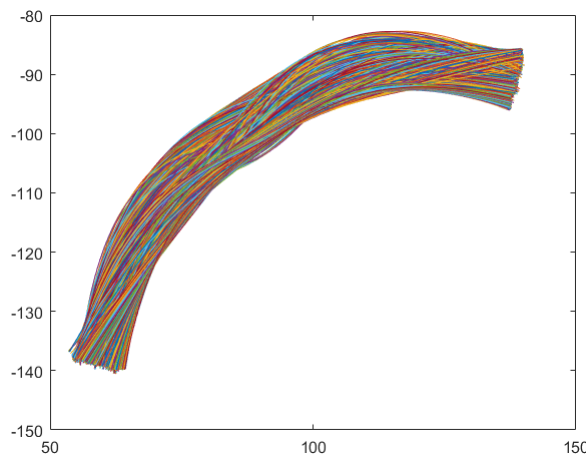
A 4.12. ábra bináris sűrűségképének fekete-fehér kontrasztja kiváló terepet biztosít a 3. fejezetben taglalt kontúrkereső algoritmus futtatására, hiszen az eljárás segítségével detektálható a fekete és fehér domének maximális világosságú felső határvonala, ami éppen a szájpadkontúrt közelíti. A kontúrkereső algoritmussal kirajzolt szájpadkontúrt a 4.13.a ábra piros görbéje rögzíti, ami a 4.13.b ábrán összehasonlítható a fentebbi szakaszokban kifejtett szélsőérték-probléma keretében megállapított zöld görbével. Látható, hogy a két különböző nézőpontból indított módszer nagyon hasonló eredményre vezetett, miután a két szájpadkontúr néhány pixelyi hibahatáron belül fedésbe hozható egymással. Mindezek után úgy ítélem meg, hogy a szájpadkontúrt az UH-felvételek által adott lehetőségekhez mérten sikerült elfogadható pontossággal kijelölni.



4.13. ábra. A szélsőérték-probléma keretében meghatározott (zöld), illetve a kontúrkereső algoritmussal detektált (piros) szájpadkontúr a bináris sűrűségképén. (Az ábra tengelyeinek mértékegysége pixel.)

4.2. Az MRI-szájpadkontúr megállapítása

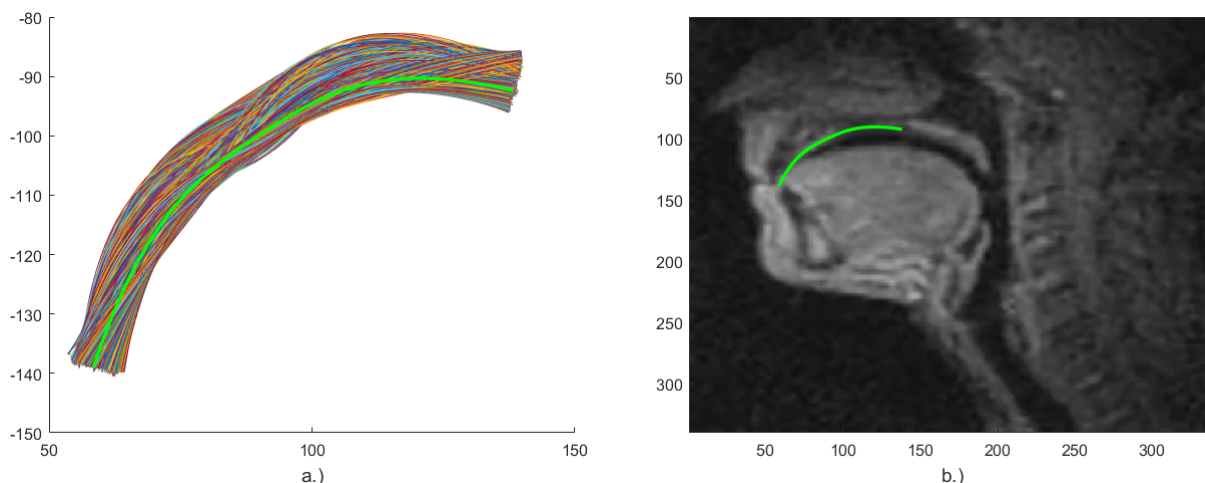
Az MRI-szájpadkontúr kijelölése jóval egyszerűbb feladatnak bizonyult, mint az UH-szájpadkontúr pozíciójának becslése, bár kétségtelen, hogy ez az út sem triviális, mivel az MRI-szájpadkontúrt is alapos körütekintéssel kell rögzíteni. Ez esetben a görbe konstrukcióját, a 3. fejezetben ismertetett kontúrkövető algoritmust adaptálva, két különálló szakaszban végeztem el, ugyanis a kemény és a lágy szájpad megrajzolásához eltérő módszereket alkalmaztam. Ezt az eljárást több körülmény is indokolta. Egyrészt beszéd közben a kemény szájpad helyzete rögzített, míg a lágy szájpad képes mozogni, így az előbbi statikus, az utóbbi pedig dinamikus görbekeresési technikát igényel. Ezenkívül az MRI-keretek egy részén a kemény szájpad kissé bizonytalanul, diffúz módon jelenik meg, a lágy szájpad kontúrja viszont sokkal élesebben ábrázolódik. Harmadrészt pedig a teljes szájpadkontúr egyetlen, osztatlan görbeként történő megszerkesztése nem vezetne egzakt megoldásokra, mivel a felső fogmedertől az ínyvitorláig terjedő szakasz túlságosan széles szögtartományt fed le, ezért a rendelkezésemre álló kontúrkereső algoritmus paramétereit nem igazán lehetne úgy beállítani, hogy a kontúr minden hajlata hitelesen, torzítások nélkül követhető legyen.



4.14. ábra. A 28 MRI-bemondás összes keretéhez tartozó 3 395 darab keményszájpad-kontúr halmaza. (Az ábra tengelyeinek mértékegysége pixel.)

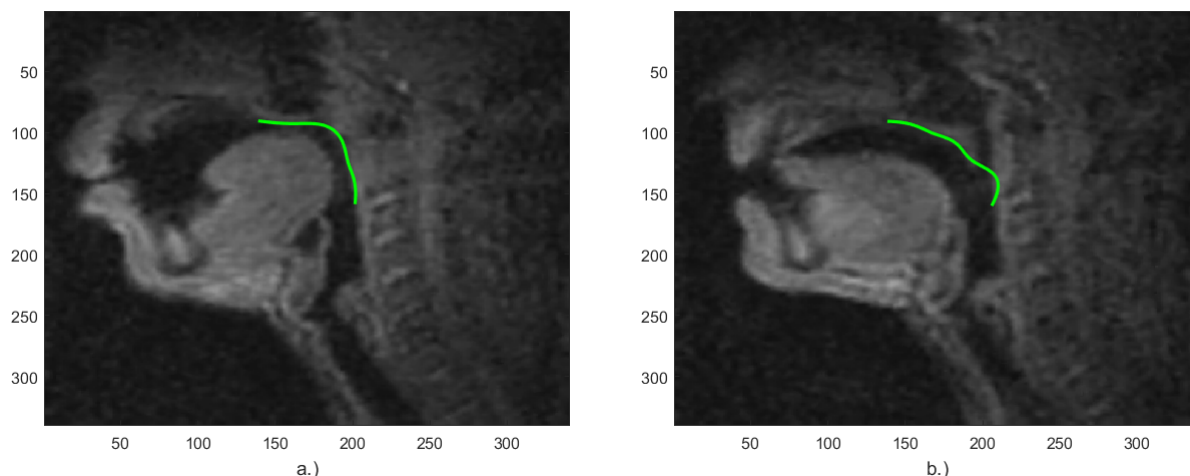
A kemény szájpad körvonalának felvételéhez egy közös koordináta-rendszerben összenyaláboltam a vizsgált 28 MRI-bemondás összes keretén azonosított 3 395 darab szájpadkontúrt, amit a 4.14. ábra prezentál. A görbesereg bal oldali végénél található a felső fogmeder, a köteg jobb oldali vége pedig a kemény és lágy szájpad közötti átmenet közelében lokalizálható. Megfigyelhető, hogy a szájpadkontúrok halmaza egy viszonylag széles sávot definiál, ami minden bizonnyal a kemény szájpad fentebb jelzett keretenkénti határozatlanságából adódik. Ennek okán, hogy megszüntessem a sokaság által generált életlenséget, kiszámítottam a halmaz elemeinek átlagát, amit a 4.15. ábra zöld görbéi érzékeltetnek az összes szájpadkontúr halmazán, illetve egy nyugalmi nyelválláshoz tartozó MRI-kereten. Az átlaggörbe megfelelően illeszkedik az MRI-keret anatómiai kontúrjához, ami azt mutatja, hogy sikerült a kemény szájpad statikus körvonalát kielégítő pontossággal realizálni.

A lágy szájpad görbét szintén a kontúrkövető algoritmus felhasználásával alkottam meg. Ebben az esetben azonban nem végeztem átlagolást a kapott görbék halmazán,



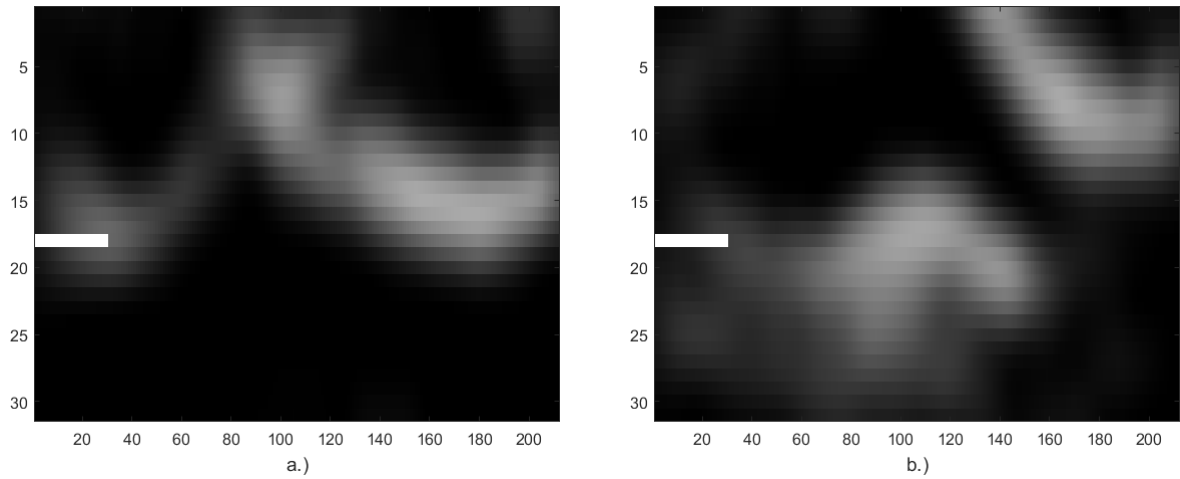
4.15. ábra. A kemény szájpad kontúrja az összes szájpadkontúr halmazán (a.), illetve egy nyugalmi nyelválláshoz tartozó MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

hanem tekintetbe véve a lágy szájpad mozgékonyosságát, dinamikus letapogatást érvényesítettem. A 4.16. ábra egy-egy lágyszájpad-kontúrt mutat be az *o* és *n* hangok példáján keresztül.

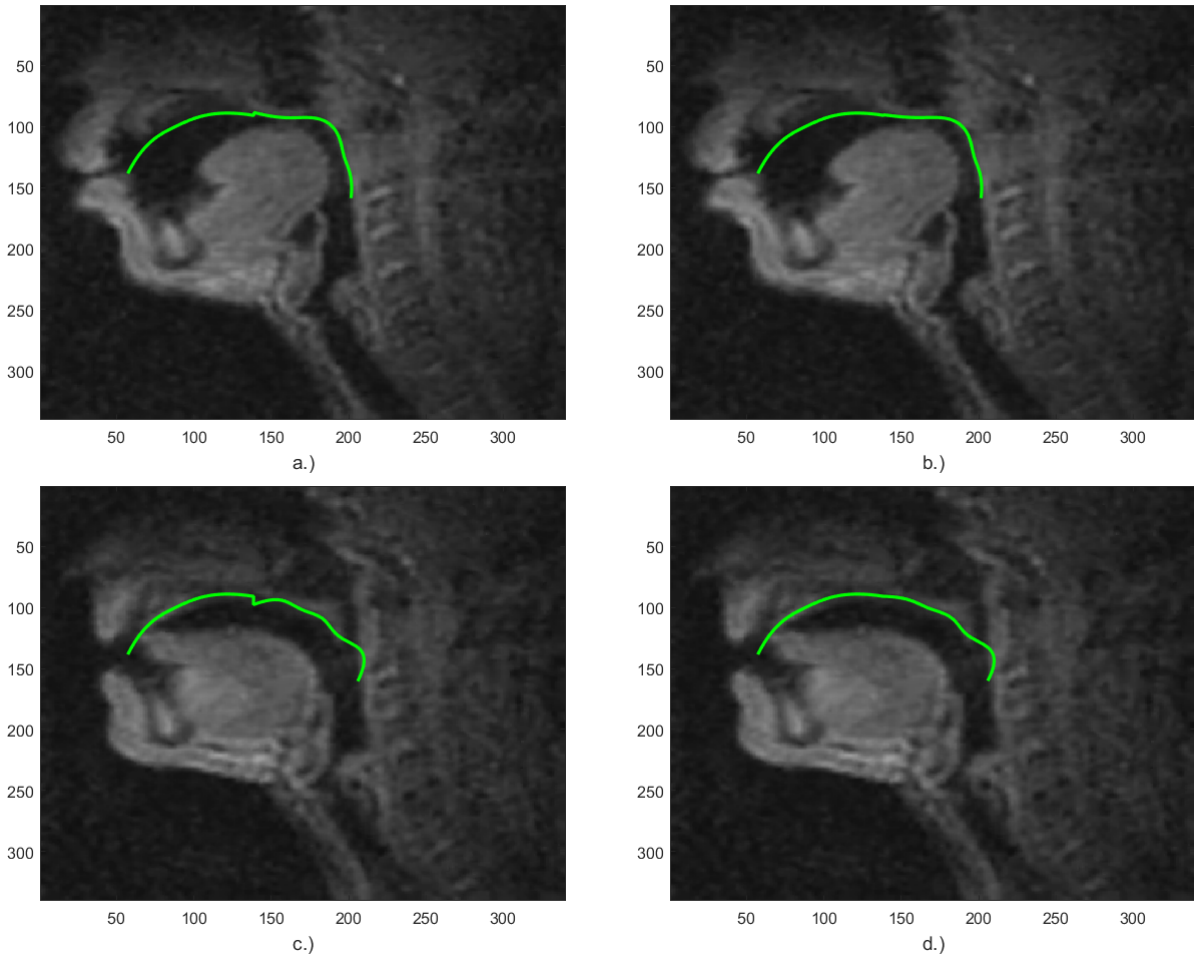


4.16. ábra. Automatikusan illesztett lágyszájpad-kontúrok egy *o* (a.), illetve egy *n* (b.) hanghoz tartozó MRI-kereten. (Az ábra tengelyeinek mértékegysége pixel.)

A kemény és lágy szájpad görbéinek illesztésekor pontosan ugyanazon lépések szerint valósítottam meg a kontúrkövető algoritmust, mint ahogyan azt a 3. fejezet szakaszai diktálják. A különbség pusztán a program paramétereinek eltérő beállításában rejlik. Ez például abban nyilvánul meg, hogy más helyre kerül a sugaras mintavételezés középpontja (3.3. ábra), módosulnak a vízszintes relevanciateromány határszögeinek értékei és a függőleges relevanciateromány, az élkimelő mátrixok elemeinek pozíciói szimmetrikusan felcserélődnek (3.6. ábra), illetve változhat az élkimelő és átlagoló mátrixok mérete, a radiális mintavételezés során beállított sugárirányú hatótávolság, valamint a görbék simításához szükséges diszkrét koszinusztranszformáció együtthatóinak száma.



4.17. ábra. A művi úton beszűrt maximális világosságú tartomány (vízszintes fehér csík) egy o (a.), illetve egy n (b.) hanghoz tartozó MRI-kereten. (Az ábra tengelyeinek mértékegysége pixel.)



4.18. ábra. A teljes szájpadkontúr diszkontinuitást tartalmazó (a. és c.) és folytonos (b. és d.) görbéje egy o (a. és b.), illetve egy n (c. és d.) hanghoz tartozó MRI-kereten. (Az ábra tengelyeinek mértékegysége pixel.)

A teljes szájjpadkontúr folytonosságának biztosítása érdekében gondoskodnom kellett a keményszájpad-kontúr végpontjának és a lágyszájpad-kontúr kezdőpontjának szakadásmentes összeillesztéséről is. Ezt úgy oldottam meg, hogy a lágy szájjpad görbájének keresésekor, a 3.8. ábra szerkezetének megfeleltethető négyszöges geometriájú mátrixban a kép bal oldali részén tapasztalati úton beállítottam egy olyan tartományt, melynek világossága jelentősen meghaladja a kép legnagyobb világosságú pixeljének értékét. Ezzel a beavatkozással garantálható, hogy a Viterbi-algoritmus által generált kontúr kezdőpontja minden MRI-kereten a mesterségesen beiktatott maximális világosságú zónában rögzül, amit a 4.17. ábra vízszintes fehér csíkjai lokalizálnak az o és n hangokhoz rendelt MRI-keretek esetében. A fentebb említett eljárással sikerült kiküszöbölni a kemény és lágy szájjpad körvonalainak találkozásánál fellépő diszkontinuitást. A módszer hatékonyságát a 4.18. ábra igazolja az o és n hangok példái által. A 4.18.a és 4.18.c kereteken az egyenetlen görbék, a 4.18.b és 4.18.c kereteken pedig már a folytonosan csatlakoztatott kontúrszakaszok figyelhetők meg.

5. UH- és MRI-nyelvkontúrok összehangolása optimalizációs eljárások segítségével

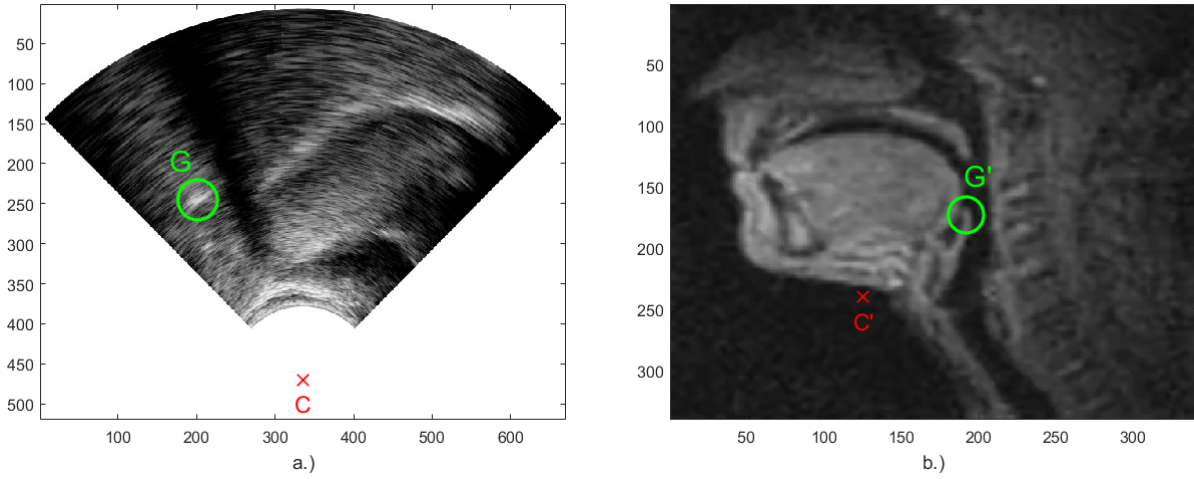
Kutatómunkám első nagyobb volumenű szakaszában az UH- és MRI-felvételek szimultán elemzésére és összehangolására fókuszáltam, melynek célja az UH-keretek radiális geometriájának, illetve az MRI-keretek négyszöges geometriájának kölcsönösen egyértelmű megfeleltetése volt. Ez a feladat több szempontból is komoly kihívásokat tartogathat, hiszen egyrészt – a 2.1. alfejezetben tárgyalt ismereteknek megfelelően – az MRI-felvételekkel szemben az UH-felvételekből jóval szűkebb információhalmaz nyerhető ki, így például a nyelv felszíne is csak részlegesen mutatkozik meg, tehát a nyelv vonatkozásában egy szűkebb tartományt kell összeegyeztetni az MRI által leképezett teljes régióval. Másodsorban az UH-képeken megjelenő hangképző szervek orientációja, és ennél fogva a látótérbe eső zóna nagyban függ attól, hogy az UH-fej milyen pozícióban van rögzítve a beszélő állala alatt. Harmadrészt pedig az UH- és MRI-felvételek természetesen beszélőspecifikusak, mivel az adatközlők neme, anyanyelve és egyéni szervi adottságai végtelenül változatossá tehetik a képek anatómiai részleteit.

A két forrás összehangolásának egy lehetséges módja az artikuláció szempontjából releváns és azonos típusú szájjüregi kontúrvonalak kölcsönösen egyértelmű konverziója. Ezt az elgondolást speciális geometriai transzformációk kivitelezésével valósítottam meg, melynek kézenfekvő eszközei a 3. és 4. fejezetek szerint meghatározott nyelv- és szájjpadkontúrok voltak. A transzformációk vizuális sémáit és jellemző paramétereit magába foglaló matematikai keret kijelölését követően pedig olyan optimalizációs technikákat alkottam meg, melyek eltérő perspektívák alapján megkeresik a transzformációk által definiált paraméterhalmaz legkedvezőbb értékeit. Az optimalizációs algoritmusok programkódjai az alábbi linken tekinthetők meg:

<https://drive.google.com/drive/folders/1pGo8u3BjFjGJpIXBJQYsXWIdZGpAW218?usp=sharing>

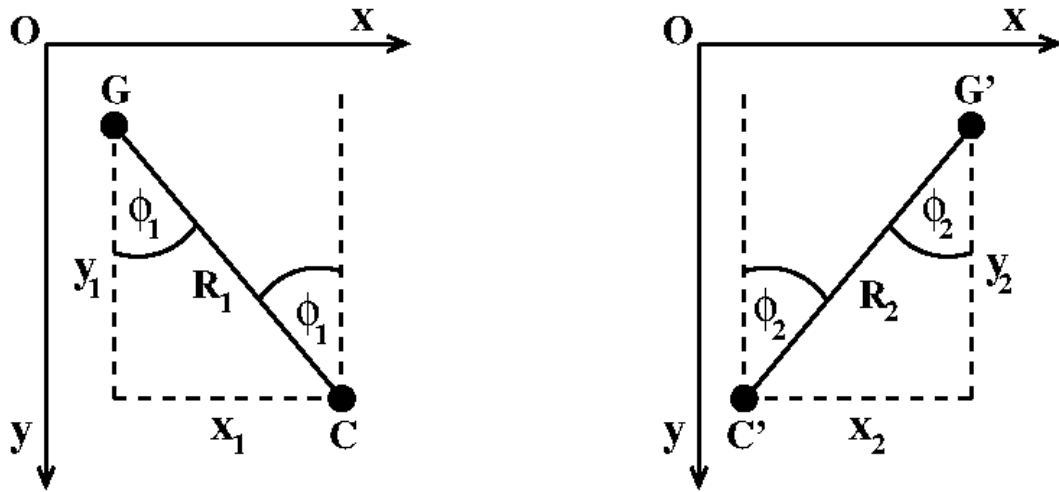
5.1. A transzformációk geometriai aspektusai és matematikai formulái

A nyelv- és szájpaddocktúrok transzformációjához elsőként egy olyan viszonyítási pontot kerestem, ami az UH- és MRI-kereteken is meggyőző biztonsággal azonosítható. Ezt a pontot a nyugalomban lévő gégefedő csúcspontjánál definiáltam, melynek helyzetét az 5.1. ábra képein megrajzolt zöld körvonal lokalizálja, a transzformáció középpontja pedig piros keresztjellel van ellátva. Ebből kiindulva a $\{C, G\}$, illetve a $\{C', G'\}$ pontpárok egymás transzformáltjainak tekinthetők.



5.1. ábra. A gégefedő zöld körvonallal határolt csúcspontja és a transzformáció piros keresztjellel megjelölt középpontja egy-egy nyugalmi UH- (a.) és MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

A transzformáció középpontjának és a gégefedő csúcspontjának bevezetése után létrehoztam az 5.2. ábra által vázolt geometriai struktúrát, ami megteremtette annak a lehetőségét, hogy értelmezhessem a transzformáció alapvető paramétereit.



5.2. ábra. Az UH- és MRI-keretekhez rendelt pitagorasz geometria (bal és jobb) elemeinek grafikus megjelenítése. (Az ábra tengelyeinek mértékegysége pixel.)

Az ábra bal és jobb oldali blokkjai az 5.1.a és 5.1.b ábrák orientációját követik sematikus módon, és a bal oldali rajz $\{C, G\}$, illetve a jobb oldali rajz $\{C', G'\}$ pontjai megfelelnek az 5.1. ábra ugyanazon jelölésekkel felcímkézett pontjainak. Az R_1 és R_2 radiális távolságokat a transzformáció középpontja és a gégefedő csúcspontja között mérhetjük a négyszöges geometria szerint felvett derékszögű háromszögek átfogói mentén, ahol az R_1 és R_2 szakaszok vízszintes és függőleges irányú $(x_1; y_1)$ és $(x_2; y_2)$ vetületei a háromszögek befogóit adják. A ϕ_1 és ϕ_2 polárszögeket a transzformáció középpontján átmenő függőleges tengely, illetve az R_1 és R_2 sugarak által bezárt szögek hozzák létre. Miután az UH- és MRI-felvételek esetében a polárszög előjeles mennyiségként volt bevezetve a leírásba, és figyelembe véve az 5.2. ábra elemeinek relatív pozícióit, ϕ_1 negatív, ϕ_2 pedig pozitív értékű. Az O pont az UH- és MRI-keretek bal felső sarkában rögzített origót szimbolizálja. A C, G, G' pontok x és y irányú koordinátái rögzített értékekkel rendelkeznek, hiszen C éppen megegyezik az UH-felvételek által definiált körcikk középpontjával, G és G' pozíciói pedig vizuálisan leolvashatók az 5.1. ábra nyugalmi UH- és MRI-kereteinek felhasználásával. Ennek megfelelően a C, G, G' pontok koordinátái a

$$\begin{aligned} C_x &= 335, & C_y &= 471, \\ G_x &= 207, & G_y &= 245, \\ G'_x &= 192, & G'_y &= 168 \end{aligned} \quad (5.1)$$

értékeket kapják, melyek pixel egységben mérendők az 5.2. ábra O origójából. Az (5.1) által adott numerikus adatokat alapul véve kiszámítható az UH-kereten megrajzolt derékszögű háromszög összes paramétere az

$$\begin{aligned} x_1 &= C_x - G_x = 335 - 207 = 128 \\ y_1 &= C_y - G_y = 471 - 245 = 226 \\ R_1 &= \sqrt{x_1^2 + y_1^2} = 259,73 \\ \phi_1 &= -\arctan\left(\frac{x_1}{y_1}\right) = -29,53^\circ \end{aligned} \quad (5.2)$$

formulák alkalmazásával. A C' pont pozíciója nem ismert, de (5.2) szerkezetéhez hasonló módon kifejezhetők az MRI-kereten megrajzolt derékszögű háromszög paraméterei is az

$$\begin{aligned} x_2 &= G'_x - C'_x = 192 - C'_x \\ y_2 &= C'_y - G'_y = C'_y - 168 \\ R_2 &= \sqrt{x_2^2 + y_2^2} \\ \phi_2 &= \arctan\left(\frac{x_2}{y_2}\right) \end{aligned} \quad (5.3)$$

egyenlőségek szerint, melyek tartalmazzák a C' pont x és y irányú koordinátáit. Az (5.2)-(5.3) matematikai összefüggések rögzítését követően az MRI- és UH-keret radiális távolságainak hányadosaként definiáltam az

$$R = \frac{R_2}{R_1} \quad (5.4)$$

nagyítást, illetve előjeles polárszögeinek összegeként értelmeztem a

$$\Phi_0 = \phi_1 + \phi_2 \quad (5.5)$$

szögelfordulást. Ezzel explicit és implicit módon is összekapcsoltam az UH- és MRI-keret paramétereit, hiszen (5.4) és (5.5) közvetlen összefüggései (5.2) és (5.3) révén indirekt módon tartalmazzák a $\{C, G\}$ és $\{C', G'\}$ pontok koordinátáit is.

A fentebb bemutatott geometriai keretrendszerben megalkottam egy olyan transzformációs mechanizmust, amely alapvetően három operációt foglal magába. A három művelet a radiális tartomány skálázása, a szögtartomány skálázása, valamint a szögtartomány forgatása által deklarált mozzanatokból tevődik össze, ami matematikailag az

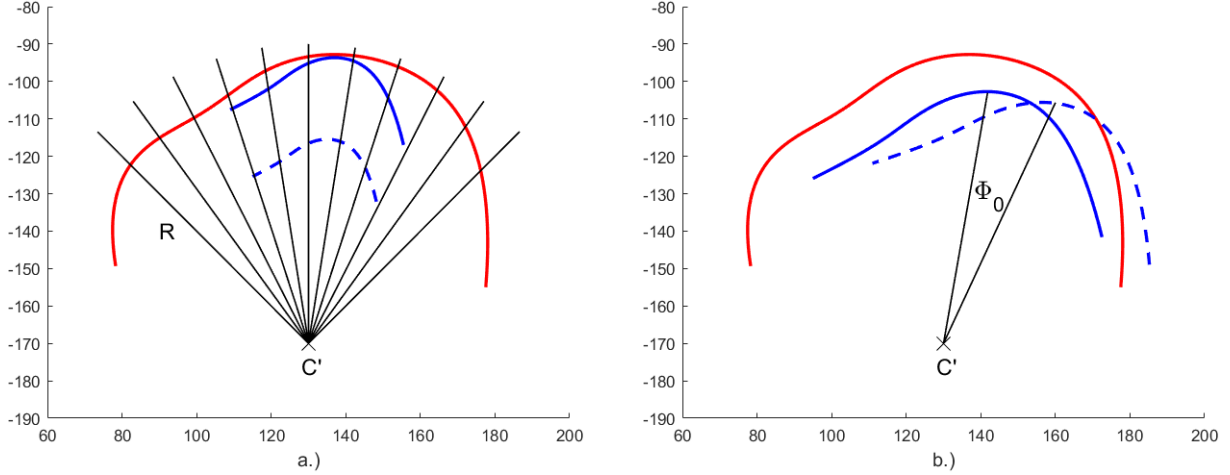
$$\begin{aligned} r' &= R r, \\ \phi' &= \Phi \phi, \\ \phi' &= \phi + \Phi_0, \end{aligned} \quad (5.6)$$

relációcsoport segítségével realizálható. Az (5.6) által kijelölt operációkból származtathatók az

$$\begin{aligned} r &= \frac{r'}{R}, \\ \phi &= \frac{\phi'}{\Phi}, \\ \phi &= \phi' - \Phi_0 \end{aligned} \quad (5.7)$$

inverz műveletek, melyek az (5.6) egyenlőségek egyszerű átrendezésével adódnak. Az (5.6)-(5.7) összefüggések az UH-felvételek nyelv- és szájpadkontúrjainak pontjait jellemző $(r; \phi)$, illetve az MRI-felvételek nyelv- és szájpadkontúrjainak pontjaihoz rendelt $(r'; \phi')$ polárkoordináták által meghatározott sugár- és szögtartományt transzformálják. Ebből az következik, hogy az (5.6) operációk segítségével az UH-kontúrok, míg az (5.7) műveletek felhasználásával az MRI-kontúrok transzformálása végezhető el, tehát kétirányú konverzió valósul meg abban az értelemben, hogy az első eset $\text{UH} \rightarrow \text{MRI}$, a második eset pedig $\text{MRI} \rightarrow \text{UH}$ irányú leképezést biztosít. Másképpen megfogalmazva az egymásnak megfeleltetett UH- és MRI-görbéket kölcsönösen egyértelműen egymásra vetítjük. Az (5.6)-(5.7) relációk R és Φ skálafaktorai a sugár- és szögtartomány normálását teszik lehetővé, a Φ_0 translációs tag pedig a szögtartomány elforgatásáért felel. Az $\{R, \Phi, \Phi_0\}$ paraméterhalmaz elemeinek az $\text{UH} \rightarrow \text{MRI}$ és $\text{MRI} \rightarrow \text{UH}$ irányok mentén végrehajtott transzformációk esetében szükségszerűen ugyanazokkal az értékekkel kell rendelkezniük, hiszen az UH- és MRI-környezet relatív skálaaránya és relatív orientációja a konverzió irányától függetlenül állandó. Vizsgálataim során a Φ faktor értékét egységnyiinek választottam $\Phi = 1$ szerint, ami azt fejezi ki, hogy a szögtartomány szélessége változatlan marad az átalakítások során, azaz a transzformációt szögtartó leképezésként értelmeztem. Ennélfogva csak a sugártartomány skálafaktora és a szögtartomány translációs tagja módosulhat az 5.3. ábrával összhangban, ahol a folytonos kék és piros vonalak a kiinduló görbepárként felvett UH- és MRI-nyelvkontúrt érzékeltetik. Az 5.3.a ábra a radiális metszeteket illusztrálja, melyek mentén a kék görbe az R skálafaktor szerint méretváltozást szenved, az 5.3.b ábra pedig a kék görbe Φ_0 szöggel történő síkbeli elforgatását demonstrálja a transzformáció

C' középpontja körül. A szaggatott kék görbék az operációk eredményeképpen létrejövő nyelvkontúrokat szemléltetik. Összefoglalva tehát a fentebb megkonstruált geometriai transzformációk az adott görbe nagyításának és síkbeli elforgatásának műveleteiben öltenek alakot.



5.3. ábra. A kontúrtranszformációk grafikus vizualizálása: a kék görbe R skálafaktorral történő kicsinyítése (a.), illetve Φ_0 szöggel történő síkbeli elforgatása (b.) a piros görbéhez viszonyítva. (Az ábra tengelyeinek mértékegysége pixel.)

Az 5.3. ábrát tanulmányozva felfigyelhetünk arra, hogy egy párba rendezett UH- és MRI-kontúr esetében a két görbe közötti globális távolság nagymértékben függ a C' pont helyzetétől, valamint az R nagyítás és a Φ_0 szögelfordulás értékétől. Erre a felismerésre alapozva az általam kidolgozott optimalizációs algoritmus lényege abban áll, hogy megkeressük a $\{C', R, \Phi_0\}$ paraméterek ideális értékeit oly módon, hogy az UH- és MRI-kontúrok között mérhető globális távolság minimális legyen. A feladat megvalósításakor az (5.2)-(5.7) összefüggések együttesére támaszkodva olyan algoritmusokat fejlesztettem ki, melyek a $\{C', R, \Phi_0\}$ adathalmaz elemeinek optimalizációja révén minimalizálják az UH- és MRI-kontúrok globális távolságát.

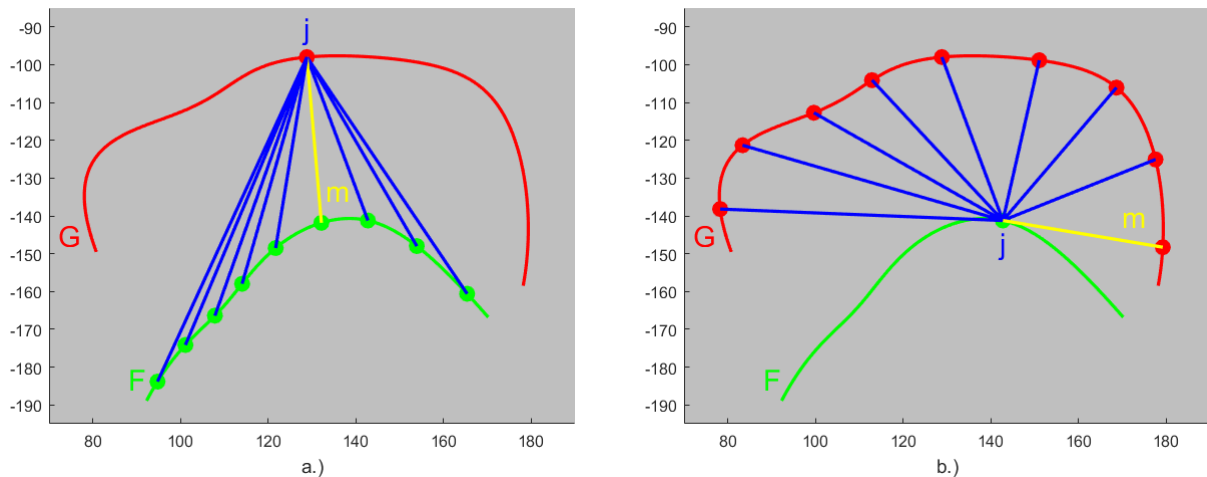
5.2. Az UH- és MRI-kontúrok globális távolságának értelmezése

Az UH- és MRI-kontúrok között mérendő távolság definiálására az úgynevezett NND mértéket (Nearest Neighbour Distance) használtam, ami kiváltképp alkalmas eltérő számú pontból felépülő görbék távolságának meghatározására a

$$D_{F,G} = \frac{1}{N_1 + N_2} \left(\sum_{i=1}^{N_1} \min_j |f_i - g_j| + \sum_{j=1}^{N_2} \min_i |g_j - f_i| \right) \quad (5.8)$$

utasítás szerint [63]. (5.8) értelmében a tetszőleges N_1 és N_2 elemet tartalmazó F és G görbék az $F = \{f_1, f_2, \dots, f_{N_1}\}$ és $G = \{g_1, g_2, \dots, g_{N_2}\}$ koordináták által vannak kifesztve a kétdimenziós síkban. Ekkor (5.8) első járuléka G adott pontja és F összes lehetséges pontja között mérhető távolságok minimumainak az összegét adja meg G összes lehetséges pontjára vonatkoztatva. Hasonlóképpen (5.8) második járuléka F adott pontja és G összes lehetséges pontja között mérhető távolságok minimumainak az összegét hordozza F

összes lehetséges pontjának figyelembevételével. A teljes összeg a két görbét alkotó pontok számának összegével van normálva. Az NND mérték kétirányú számításának egyes részletei grafikusán az 5.4. ábra példája szerint értelmezhetők, ahol az a.) és b.) rajzokon a G és F görbék egy-egy tetszőlegesen felvett j pontjának esetében az aktuálisan értelmezhető távolságok m minimuma sárgával van megjelölve. A j pont F és G görbéken történő végigcsúsztatásával minden j ponthoz hozzárendelhetjük az m minimumot, majd ezen értékek kumulálásával jutunk el (5.8) két összegjárulékához. A számítást természetesen az adott görbepár összes lehetséges pontpárjára el kell végezni, az 5.4. ábrán azonban a könnyebb áttekinthetőség kedvéért csak néhány pontot tüntettem fel az a.) rajz F , illetve a b.) rajz G görbéje mentén. Mindezekkel egyetértésben G adott pontjának F -től mért távolsága F -nek G kijelölt pontjához legközelebb eső pontjától mért távolsággal egyezik meg, és ugyanígy F adott pontjának G -től mért távolságát G -nek F kijelölt pontjához legközelebb eső pontjától mért távolság definiálja. Gondoljuk meg, hogy az (5.8) kifejezés két összegjárulékában számított távolságok nem feltétlenül egyeznek meg egy rögzített görbepár esetében.



5.4. ábra. A két tetszőleges görbe viszonylatában értelmezett NND távolságmérték kétirányú számításának szemléltetése. (Az ábra tengelyeinek mértékegysége pixel.)

Ebben a megközelítésben az F és G görbék az UH- és MRI-kontúrok szerepét játsszák. Egy kiválasztott görbepár $\{f_i, g_j\}$ pontpárjának $x - y$ síkban mért távolsága az $f_i = (x_i; y_i)$, illetve a $g_j = (x_j; y_j)$ descartes-i koordináták felhasználásával egyszerűen kifejezhető a

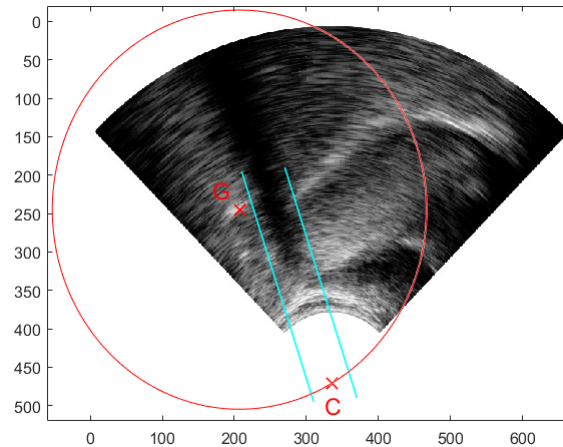
$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (5.9)$$

alakban felírt Pitagorasz-tétel segítségével.

5.3. A paraméterek optimalizációját segítő megfontolások

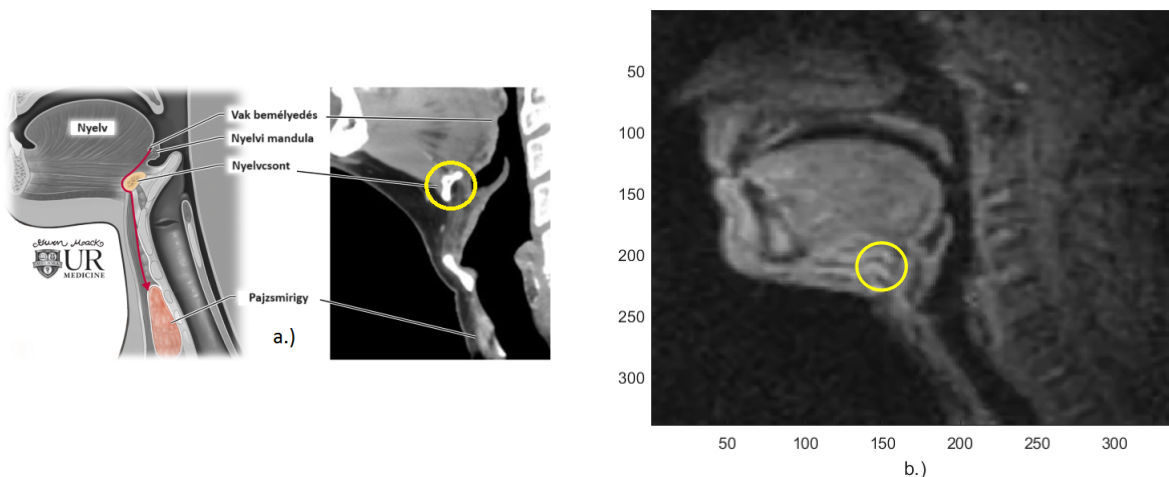
Mielőtt hozzáláttam a $\{C', R, \Phi_0\}$ adathalmaz optimalizálásához, alaposan tanulmányoztam az UH- és MRI-felvételek felépítését anatómiai és geometriai szempontok figyelembevételével. Mivel az (5.2) egyenlőségek tanúsága szerint az UH-kereteken minden releváns paraméter rögzített értékkel rendelkezik, elsősorban arra törekedtem, hogy megvizsgáljam

az UH-képeken azonosítható fixpontok relatív helyzetének sajátosságait. Hogy az 5.1. alfejezet megállapításaival konzisztens legyen a gondolatmenetem, egy olyan UH-keretet elemeztem, amelyen a nyelv nyugalmi állapotban van, azaz nem történik hangképzés.



5.5. ábra. A fixpontok (a gégefedő G csúcspontja és a transzformáció C középpontja) relatív helyzetének szemléltetése egy nyugalmi UH-kereten a nyelvcsont árnyéksávjához viszonyítva. (Az ábra tengelyeinek mértékegysége pixel.)

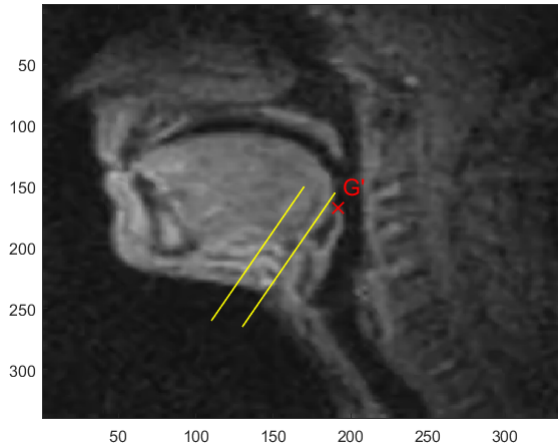
Ezt alapul véve, az 5.5. ábrát követve, a nyugalmi UH-kereten piros keresztjellel bejelöltem a gégefedő G csúcspontját és a transzformáció C középpontját, majd a G pontból kiindulva berajzoltam az (5.2) által meghatározott R_1 sugarú kört. Ezenkívül két párhuzamos kék vonal segítségével behatároltam a nyelvcsont árnyékának radiális vetületét, és figyeltem arra a körülményre, hogy a transzformáció középpontja a nyelvcsont árnyék-sávjában helyezkedik el, a gégefedő csúcspontja viszont ezen a zónán kívül esik.



5.6. ábra. A nyelvcsont elhelyezkedése egy sematikus képen és egy CT-felvételen (a.), illetve egy nyugalmi MRI-kereten (b.). (A (b.) ábra tengelyeinek mértékegysége pixel.) [64]

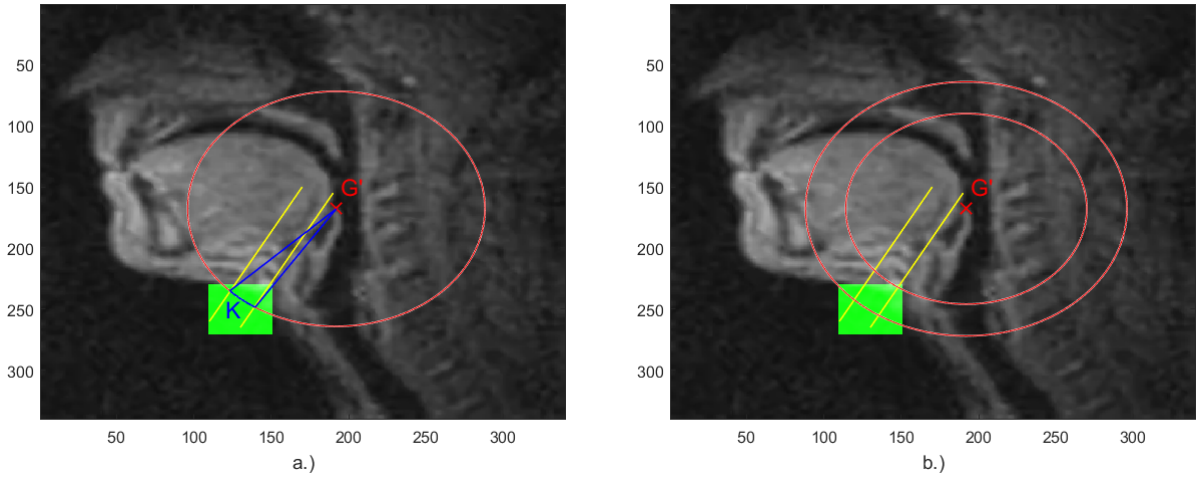
A nyugalmi UH-keret fixpontjainak relatív helyzetére kinyert információkat felhasználva, hasonló szempontokat tekintetbe véve megvizsgáltam egy nyugalmi MRI-keret ugyanazon pontjainak pozicionálását is. Elsőként az anatómiai szakirodalom tanulmányozásával

a nyelvcsontot próbáltam meg azonosítani, ami az 5.6.a ábra sematikus képe, illetve CT-felvétele alapján az 5.6.b ábra MRI-keretén a sárga körvonallal határolt tartományban lokalizálható. Az 5.6.b ábrából kiindulva megrajzoltam a nyelvcsont árnyéksávját is oly módon, hogy a gégefedő csúcspontja az 5.5. ábra UH-keretéhez hasonlóan az árnyéksáv határvonalán kívül, de annak közelében legyen. Ennek nyomán a gégefedő piros keresztjellel megjelölt G' csúcspontjának és a nyelvcsont két párhuzamos sárga vonallal érzékeltetett árnyéksávjának relatív helyzetét az 5.7. ábra tünteti fel.



5.7. ábra. A gégefedő G' csúcspontjának és a nyelvcsont árnyéksávjának relatív helyzete egy nyugalmi MRI-kereten. (Az ábra tengelyeinek mértékegysége pixel.)

A transzformáció C' középpontját az 5.2. ábrával egyetértésben, és az 5.5. ábra geometriai elrendezéséhez igazodva egy R_2 sugarú kör mentén kell elhelyezni úgy, hogy a C' pont a nyelvcsont árnyéksávjában legyen. Ennek megfelelően a C' pont potenciális pozíciói az 5.8.a ábra szerint a nyelvcsont árnyéksávja, az R_2 sugarú kör, valamint a zöld téglalap által meghatározott terület közös metszetében realizálhatók, amit a kézzel megjelölt K körívek definiálnak. A zöld téglalap területét a $Z_x = [110, 150]$ és a $Z_y = [230, 270]$ koordináták szerint rögzített intervallumok egyértelműsítik az x és y irányok mentén. Ebben a fázisban észrevehetjük azt, hogy nemcsak a C' pont lehetséges pozíciójában mutatkozik meg egyfajta szabadsági fok, hanem az R_2 sugár nagysága is szabadon mozoghat a fentebb említett közös metszet által adott jó tartományon belül. Ennek felismerésével lehetőség nyílik arra, hogy az R_2 sugár értékére alsó határt szabjunk ki, és ezáltal, az (5.4) összefüggést alkalmazva, információt nyerjünk az R nagyítás minimumára is. Ezen a szinten az R_2 sugár és így az R nagyítás felső határára vonatkozóan nem tudunk észszerű becslést adni, mivel vizuálisan csak azt tudjuk megállapítani, hogy melyik az a legkisebb sugarú kör, amivel már el tudjuk érni a jó tartományt, azonban sem anatómiai, sem geometriai megfontolások alapján nem áll módunkban semmilyen felső korlát kijelölése. Az R_2 sugár minimumához tartozó kört, illetve ezen minimális sugárnál nagyobb, de tetszőlegesen megválasztott értékű sugárral felvett kört az 5.8.b ábra demonstrálja. Az ábra két körnek sugara pixel egységben kifejezve 77,92, illetve 103,89 értékű, melyekhez a 0,3 és 0,4 nagyításértékek tartoznak. Ezek a következtetések mindenképpen irányadónak tekinthetők a $\{C', R, \Phi_0\}$ paraméterek optimalizációja során, hiszen kizárólag olyan eredmények fogadhatók el, amik arra vezetnek, hogy az R nagyítás értéke 0,3-nél nagyobb, a C' pont pedig a beszélő álla alatti jó tartományon belül marad.



5.8. ábra. Egy G' középpontú, R_2 sugarú kör (piros), a nyelvcsont árnyéksávja (sárga), a C' pont potenciális tartománya (zöld) és az előző három geometriai halmaz közös K metszete a körív mentén (kék) egy nyugalmi MRI-kereten (a.) változtatható R_2 sugarú körök illusztrálásával (b.). (Az ábra tengelyeinek mértékegysége pixel.)

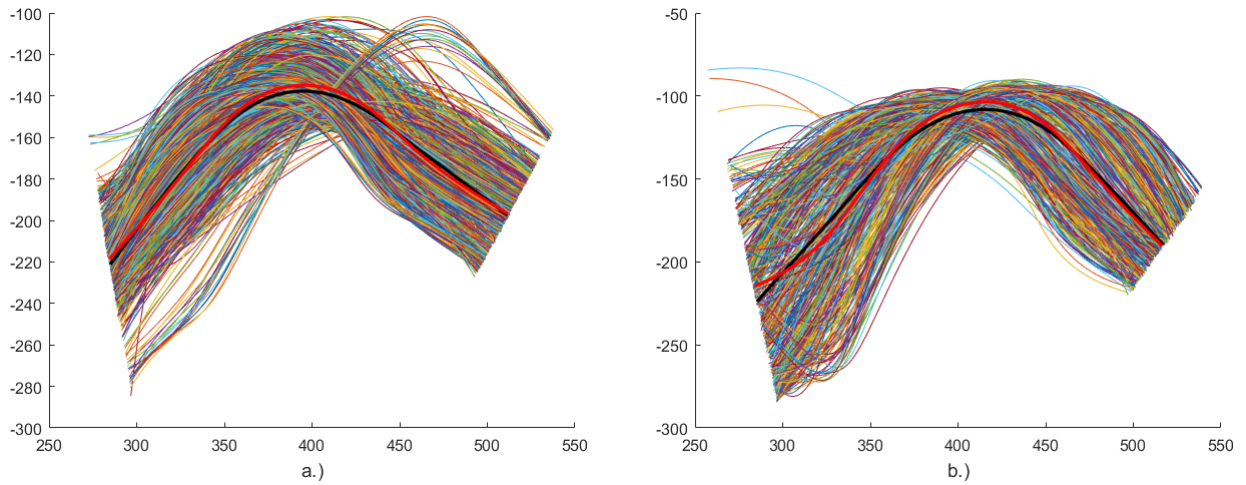
A fentebb kibontott anatómiai és geometriai töltetű elgondolások figyelembevételével számos különböző utat dolgoztam ki a $\{C', R, \Phi_0\}$ paraméterek lehetséges optimalizálására, amiket a következő alfejezetekben ismertetek.

5.4. Az optimalizációs algoritmusok felépítése

A $\{C', R, \Phi_0\}$ paraméterek optimalizációját minden esetben egy olyan kontúrhalmazzra szorítkozva valósítottam meg, mely az $\{a, e, o, k, s, t\}$ hangokhoz tartozó hat-hat UH- és MRI-nyelvkontúrt, valamint az UH- és MRI-szájpadkontúrokat tartalmazza. A rendelkezésemre álló hangkészletből tehát három magánhangzót és három mássalhangzót emeltem ki. A hangok kiválasztásának fő szempontja az volt, hogy a nyelv legszélső pozícióit leíró nyelvkontúrok szerepeljenek, elősegítve ezzel a lehető legkiterjedtebb szájüregi tartomány bevonását az optimalizációs procedúrába. Ezt a törekvést tükrözik az a és o hangok, melyek hátsó nyelvállással képződnek, az e hangot pedig elülső nyelvállással artikuláljuk. Ezenkívül a k hangot, veláris természetének megfelelően, a szájüreg hátsó részén formázzuk, mivel a nyelv legmagasabb helyzetben lévő pontja a lágy szájpadot érinti. Ezzel szemben a t hang, alveoláris típusának köszönhetően, a másik végetet képviseli, miután a nyelvhegy a felső fogmederrel érintkezik a szájüreg elülső traktusában. Végül a poszt-alveoláris osztályba tartozó s hang is a szélső nyelvállások közelében van, hiszen a felső fogmeder környezetében, az alveoláris régió mögött képezzük.

Az optimalizációban részt vevő UH- és MRI-nyelvkontúrokat eltérő technikával választottam ki az adott hanghoz rendelt képkockákon megjelenő kontúrok összességéből. Ennek oka az, hogy míg az UH-nyelvkontúrok alakja a kontextusfüggőség miatt keretről keretre haladva rendkívül nagy változatosságot mutat még egyetlen hang esetében is, addig az MRI-nyelvkontúrok meglehetősen stabilak, mert csak nagyon elenyésző mértékben módosul az alakjuk az adott hanghoz tartozó képkockák sorozatán. Ezeket a körülményeket figyelembe véve az UH-nyelvkontúrokat az adott hanghoz rendelt görbesereg átlagához legközelebb eső kontúrként azonosítottam. A 40 különböző bemondást tartalmazó

UH-csomag 1 084 a -kontúrt, 386 e -kontúrt, 853 o -kontúrt, 489 k -kontúrt, 205 s -kontúrt, valamint 1 062 t -kontúrt foglal magába, melyek felhasználásával a hangonkénti átlaggörbe meghatározása után az (5.8) NND mértékre támaszkodva megkerestem az átlaggörbétől legkisebb távolságban lévő jelöltet. Az 5.9. ábra az a és k hangok példáján keresztül prezentálja az adott hanghoz tartozó összes nyelvkontúr halmazán feketével megrajzolt átlaggörbét, illetve az átlaghoz legközelebb eső, pirossal kiemelt nyelvkontúrt. Az MRI-felvételek esetében nem volt szükség a fentebbi átlagolás elvégzésére, hiszen a görbék kiválasztási mechanizmusa jóval egyszerűbb az MRI-nyelvkontúrok hangonkénti állandósága folytán. Ebből adódóan az adott hanghoz rendelt nyelvkontúrt a releváns keretek csoportjának középső elemén rögzített görbeként definiáltam.



5.9. ábra. Az a (a.) és k (b.) hanghoz tartozó összes nyelvkontúron felvett átlaggörbe (fekete), illetve az átlaghoz legközelebb eső nyelvkontúr (piros). (Az ábra tengelyeinek mértékegysége pixel.)

A kijelölt hanghalmazt kiegészítő UH-szájpadkontúr tekintetében a 4.1. alfejezetben megszerkesztett görbét alkalmaztam, az MRI-szájpadkontúrt pedig egy nyugalmi MRI-keretből származtattam.

Az optimalizációnak alávetett görbék rögzítését követően egymásba ágyazott *for* ciklusok keretében futtattam az optimalizálandó paramétereket úgy, hogy a legbelső *for* ciklusban érvényesítettem az UH- és MRI-kontúrok között az 5.1. alfejezetben értelmezett geometriai transzformációkat. Ahogyan azt a jelzett alfejezetben is konstatáltam, az $UH \rightarrow MRI$ és $MRI \rightarrow UH$ irányokban végrehajtott transzformációknak ugyanazon $\{R, \Phi, \Phi_0\}$ paraméterértékekkel kell megvalósulniuk, melynek folytán az (5.6) és (5.7) operációcsoport egymásnak egyértelmű inverzei. Ennélfogva az optimalizációs algoritmust elegendő az egyik irány szerint teljesíteni, hiszen az optimalizáció eredményeképpen kapott paraméterek a másik irány mentén történő transzformációkat is világosan deklarálják. Ennek jegyében az $UH \rightarrow MRI$ utat választottam, így az optimalizációs algoritmusba közvetlenül csak az (5.6) relációk épülnek be. A transzformációt egy olyan programszakasz készíti elő, aminek segítségével az adott UH-kontúr C középpontból mért előjeles $(x; y)$ koordinátáinak előállítását követően kiszámítjuk a görbe pontjait jellemző sugár- és polárszögértékeket, azaz minden pontban meghatározzuk az $(r; \phi)$ polárkoordinátákat. Ezután pedig az (5.6) összefüggések felhasználásával a görbe minden pontjában elvégezzük a sugár skálázását, illetve a szög translációját, tehát – az 5.3. ábrára visszautalva – globálisan megtörténik a

görbe R skálafaktorral való nagyítása és Φ_0 szöggel való síkbeli elforgatása. Ezt követően a transzformált UH-kontúrokat rávetítjük a megfelelő MRI-kontúrra, majd kiszámítjuk az összetartozó kontúrpárok mindegyikére az (5.8) NND mértékre alapozott távolságokat, és összegezzük az értékeket. Végül az algoritmus legbelső *for* ciklusában egy *if* feltétel keretében egy a kumulált távolságra kezdetben szabadon fixált felső korlát fokozatos leszorításával megkeressük az összegzett távolságok minimumát úgy, hogy amennyiben a kumulált távolság kisebb az előző lépésben kapott kumulált távolság által definiált felső határnál, akkor az iteráció ismétlődik, és addig tart, amíg meg nem találjuk azt a legkisebb összegzett távolságot, amire még teljesül a feltétel. Miután a *for* ciklusok változói éppen az optimalizálandó paraméterek, az összegzett távolság minimumát beállító ciklusváltozók megadják az (5.6) transzformációkhoz kapcsolódó $\{C', R, \Phi_0\}$ paraméterek optimalizált értékeit.

5.5. A megvalósított optimalizációs algoritmusok

5.5.1. Optimalizáció 1

Elsőként egy olyan optimalizációs programstruktúrát alakítottam ki, melyben a $\{C', R, \Phi_0\}$ paraméterek mindegyike szabadon futhat a kijelölt intervallumokon belül. Ezzel a megoldással az volt a célom, hogy az (5.2)-(5.5) transzformációs összefüggésektől független úton tesztelhessem az optimalizáció sikerességét, mivel az (5.2)-(5.5) relációkat egyáltalán nem vettem figyelembe az algoritmusban. Az optimalizáció eredményeit az 5.1. táblázat foglalja össze, ahol az első oszlopban a $C' = (C_x; C_y)$ koordinátapár látható, a második és harmadik oszlopok az R nagyítást és a Φ_0 szögelfordulást tartalmazzák, az utolsó oszlop pedig az NND mérték minimumát tünteti fel.

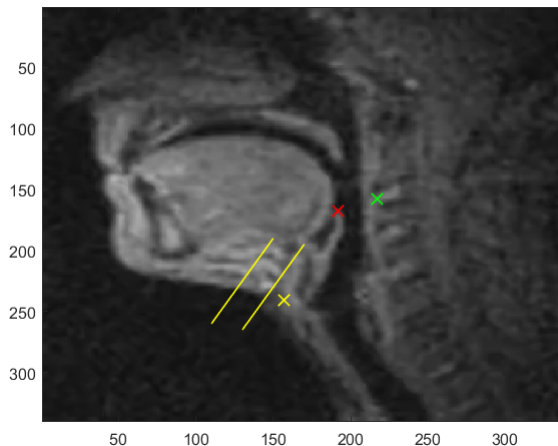
5.1. táblázat. **Az Optimalizáció 1 algoritmus futtatásával kapott eredmények.**

C' (px)	R	Φ_0 (fok)	MIN (px)
(157; 241)	0,39	6,42	28,78

Az optimalizálást követően több irányból is megpróbáltam ellenőrizni az 5.1. táblázatban foglalt numerikus értékek hitelességét. Egyfelől azt vizsgáltam meg, hogy a kapott C' pont a korábban is felhasznált nyugalmi MRI-kereten valóban a nyelvcsont árnyéksávjában helyezkedik-e el. Másrészt pedig azt is kontrolláltam, hogy az 5.1. táblázat paramétereit alapul véve visszaadódik-e az MRI-környezetben nyugalomban lévő gégefedő csúcspontjának helyes pozíciója. A gégefedőcsúcspont $G' = (G'_x; G'_y)$ koordinátáinak kiszámításához az (5.3)-(5.5) formulákat tekintettem etalonnak, és ezekből az egyenletekből, (5.2) rögzített értékeinek megtartásával, egyszerű átrendezéssel kifejeztem a

$$\begin{aligned}
G'_x &= C'_x + R R_1 \frac{\tan(\Phi_0 - \phi_1)}{\sqrt{1 + \tan^2(\Phi_0 - \phi_1)}}, \\
G'_y &= C'_y - R R_1 \frac{1}{\sqrt{1 + \tan^2(\Phi_0 - \phi_1)}}
\end{aligned} \tag{5.10}$$

koordinátákat. Elemzéseim során azt tapasztaltam, hogy sem a C' , sem a G' pont optimalizált pozíciója nem validálható, hiszen C' a nyelvcsont árnyéksávján kívül, a gége pajzsporcában lokalizálódik, és G' -nek is irreális a nyaki csigolyákban való elhelyezkedése. Ezeket a megállapításokat az 5.10. ábra támasztja alá, ahol C' sárga, G' zöld keresztjellel van megjelölve, a piros keresztjel a gégefedő csúcspontjának valódi pozícióját jelzi, a két párhuzamos sárga vonal pedig a nyelvcsont árnyéksávját határolja be.



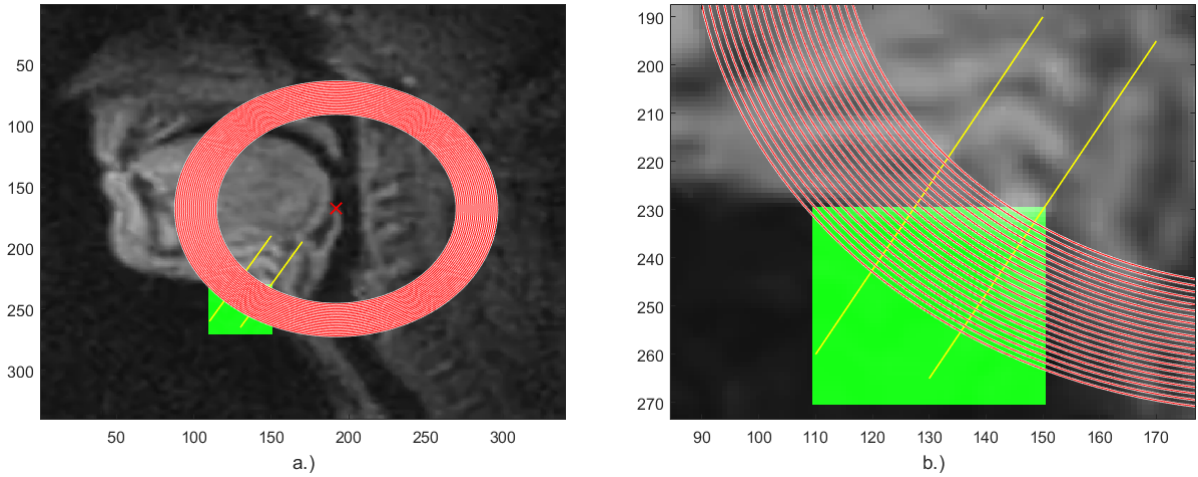
5.10. ábra. Az optimalizáció eredményeképpen kapott C' (sárga) és G' (zöld) pontok helyzete a nyelvcsont árnyéksávjához (sárga), illetve a gégefedő csúcspontjának valódi pozíciójához (piros) viszonyítva. (Az ábra tengelyeinek mértékegysége pixel.)

A fentebbi analízis tehát azt mutatja, hogy a $\{C', R, \Phi_0\}$ paraméterek szabad futtatásával nem biztosíthatók az 5.3. alfejezetben megkövetelt kritériumok, így más megközelítések alapján kellett továbbhaladnom.

5.5.2. Optimalizáció 2

Az előző alfejezet kimenetele alapján arra a következtetésre jutottam, hogy a $\{C', R, \Phi_0\}$ paraméterek optimalizációja során célszerű lehet bizonyos feltételek kijelölése, amik megkönnyíthetik a paraméterek értékeinek helyes meghatározását úgy, hogy reális eredmények születhessenek. Így ezen indítatásból elsőként visszakanyarodtam az 5.3. alfejezet megfontolásaihoz, és az 5.8.b ábrán kialakított geometriai környezetben előre meghatározott értékek szerint rögzítettem az R nagyítást az $I_R = [0,3; 0,4]$ intervallumban 0,005-es lépésközzel, ami összesen 21 különböző értéket definiál a 0,300, 0,305, 0,310, ..., 0,395, 0,400 számsorozatnak megfelelően. Az I_R intervallum elemeihez tartozó R_2 sugarú köröket az 5.11.a ábra illusztrálja egy nyugalmi MRI-kereten, az 5.11.b blokk pedig az áll alatti tartomány kinagyított részletét vizualizálja.

Az R nagyítás potenciális értékeinek beállítása után a C' pont helyzetével kapcsolatosan megfogalmazható kritériumokat tanulmányoztam kvantitatív módon. Az 5.3 alfejezettel összhangban a C' pont lehetséges pozícióit az 5.11. ábra körjeinek azon szakaszai mentén valószínűsíthetjük, melyek a zöld téglalapról a két párhuzamos sárga vonal által kimetszett területen húzódnak. Ez a megszorítás geometriailag (5.3) harmadik egyenlősége által kirótt Pitagorasz-tétel segítségével vihető keresztül, amely a releváns R_2 sugar



5.11. ábra. Az $I_R = [0,3;0,4]$ intervallumban értelmezett 21 nagyításértékhez tartozó kör egy nyugalmi MRI-kereten egészében (a.) és egy kinagyított részletben (b.) ábrázolva. (Az ábra tengelyeinek mértékegysége pixel.)

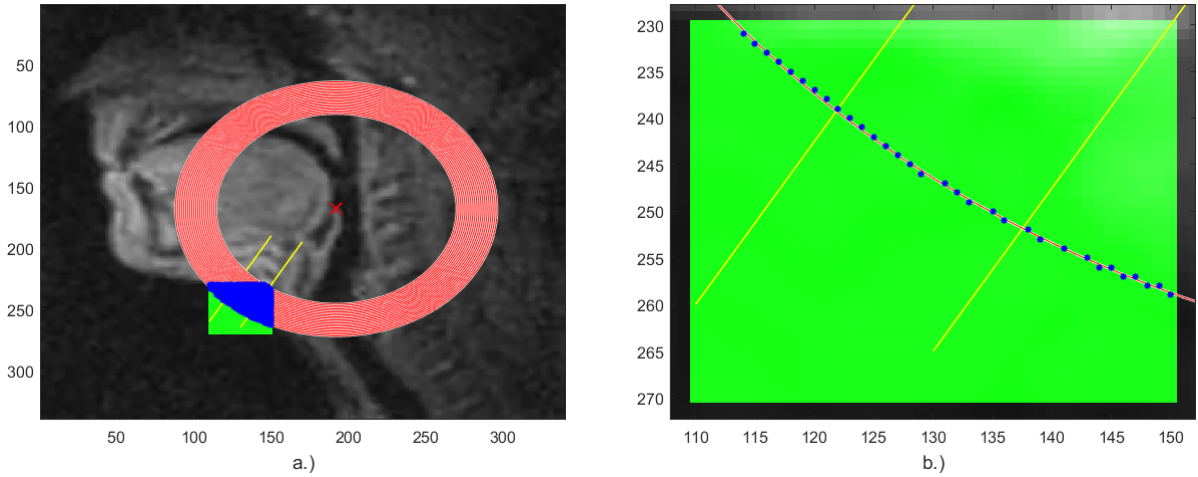
mellett (5.3) első két összefüggésének helyettesítésével tartalmazza C' koordinátáit is

$$R_2^2 = (192 - C'_x)^2 + (C'_y - 168)^2 \quad (5.11)$$

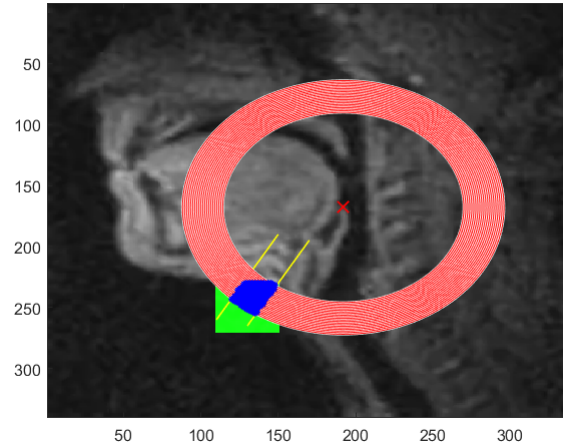
alakban. Az (5.11) feltétel azonban a MATLAB-algoritmusban nem teljesíthető egzakt egyenlőség formájában, mivel a szóban forgó mennyiségek értékeit reprezentáló tizedes törteknek csak a legkivételesebb esetekben egyezhet meg az összes digitje. Elegendő egy csekély eltérés a sokadik tizedesjegyben ahhoz, hogy a komputeres számításokban szigorúan már ne legyen igaz az (5.11) egyenlőség (pl. $2,328641 \neq 2,328642$). Ennélfogva C' illeszkedésében kénytelen voltam megengedni néhány tized pixelnyi pontatlanságot az

$$\begin{aligned} (R_2 - r)^2 &< (192 - C'_x)^2 + (C'_y - 168)^2, \\ (R_2 + r)^2 &> (192 - C'_x)^2 + (C'_y - 168)^2 \end{aligned} \quad (5.12)$$

relációk egyidejű kiszabásával, ahol az r paraméter egy $2r$ szélességű tűrési sávot határoz meg az R_2 sugarú körvonal mentén szimmetrikusan. Ez azt jelenti, hogy C' -nek az $R_2 - r$, illetve az $R_2 + r$ sugarú körök által formált vékony körgyűrű belsejében kell elhelyezkednie a beszélő álla alatt lévő jó tartományon belül. Az r paraméter értékének növelésével értelemszerűen növekszik C' lehetséges pozícióinak száma is egy adott R_2 sugarú körív mentén, hiszen egy szélesebb körgyűrűben eleve több pont található. Ezt figyelembe véve r értékének fixálásakor az volt a mérvadó szempont, hogy a körív mentén keresett pontoknak az adott körívtől mért távolsága ne haladja meg a pixelméretet, azaz ne essenek túlságosan távol a körívtől, illetve a pontok száma a pixelmérethez viszonyítva ne legyen se túl kevés, se túl sok, tehát a pontok közelítőleg egyenletesen sorakozzanak az adott szakaszon. Így vizuális megítélés alapján az $r = 0,3$ értéket jelöltem ki, melynek alkalmazásával az 5.12.a ábrán látható kék ponthalmaz adódik C' lehetséges pozícióira a 21 körív mentén a zöld tartományon belül. Az 5.12.b ábra példaként kiemelve az $R = 0,385$ nagyításértékhez rendelt körív mentén megtalált pontok sorozatát mutatja be felnagyított módon. Ezt követően az 5.12.a ábra pontsokaságából kizárólag azokat a pontokat



5.12. ábra. Az $r = 0,3$ tűrési paraméter beállításával kapott kék pontok halmaza C' potenciális pozícióira a 21 körív mentén a zöld tartományon belül (a.), illetve az $R = 0,385$ nagyításértékhez tartozó körív mentén megtalált pontok sorozatának kinagyítása (b.). (Az ábra tengelyeinek mértékegysége pixel.)



5.13. ábra. A nyelvcsont árnyéksávjába eső C' pontok kiválogatásának eredménye. (Az ábra tengelyeinek mértékegysége pixel.)

hagytam meg, amik a nyelvcsont árnyéksávjába esnek, a többi elemet pedig elimináltam. A válogatás eredményét az 5.13. ábra jeleníti meg.

A C' pont potenciális pozícióinak megállapítása megteremtette annak a lehetőségét is, hogy kiszámítsam az egyes C' pontokhoz tartozó Φ_0 szögelfordulásokat is (5.2)-(5.3) utolsó összefüggéseinek az (5.5) kifejezésbe történő helyettesítésével. Számításaim eredményeit az 5.2. táblázat sűríti példaként az $R = 0,385$ nagyításértékre vonatkoztatva. A táblázat első és második oszlopai az 5.12.b ábra ponthalmazából az 5.13. ábra szerint kiválogatott C' pontok x és y koordinátáit tartalmazzák, a harmadik oszlop pedig a megfelelő szögelfordulásokat listázza.

A fentebb bemutatott koncepciórendszerhez kapcsolódóan kétféle optimalizációs variánszt valósítottam meg, melyeket a következő két paragrafusban ismertetek.

5.2. táblázat. A C' pontok x és y koordinátái, illetve a hozzájuk tartozó szögelfordulások az $R = 0,385$ nagyításérték esetében.

R = 0,385		
C'_x (px)	C'_y (px)	Φ_0 (fok)
123	240	14,26
124	241	13,44
125	242	12,63
126	243	11,82
127	244	11,01
128	245	10,20
129	246	9,40
131	247	8,15
132	248	7,35
133	249	6,54
135	250	5,28
136	251	4,48

5.5.2.1. Optimalizáció 2.a ♦ Kidolgoztam egy olyan algoritmust, melyben az R nagyítás az $I_R = [0,3; 0,4]$ intervallum előzőekben kijelölt 21 különböző értékén fut végig, a Φ_0 szögelfordulás az egyes nagyításértékekhez rendelt C' pontok esetében kiszámított elemhalmaz szerint rögzül, a C' pont helyzete pedig optimalizálás eredményeként adódik. Ezt programozástechnikailag úgy hajtottam végre, hogy R és Φ_0 előre meghatározott értékeit táblázatokból beolvasva, C' x és y koordinátáira építettem két egymásba ágyazott *for* ciklust. Az 5.13. ábrával összhangban a 21 nagyításérték esetében definiált 21 körív mentén 208 pont azonosítható a nyelvsont árnyéksávjában, így a 208 szögelfordulási érték figyelembevétele 208 individuális futási ciklust generált. Az összes optimalizációs ciklus futtatását követően az eredmények listájából kikerestem azt a $\{C', R, \Phi_0\}$ paraméterszettet, ami a legkisebb NND távolságot produkálta az optimalizációban részt vevő UH- és MRI-kontúrok között. Ennek megfelelően az optimalizáció végeredményét az 5.3. táblázat tartalmazza.

5.3. táblázat. Az *Optimalizáció 2.a* algoritmus futtatásával kapott eredmények.

R	Számolt Φ_0 (fok)	Számolt C' (px)	Optimalizált C' (px)	MIN (px)
0,395	4,32	(135; 253)	(160; 240)	28,64

Megfigyelhető, hogy C' számolt és optimalizált koordinátái nem esnek egybe, hiszen x irányban 25 pixelnyi növekedés, y irányban pedig 13 pixelnyi csökkenés jött létre, ami azt

eredményezi, hogy C' a nyelvcsont árnyéksávján kívülre kerül, és a gégefedő csúcspontjának C' optimalizált koordinátaival (5.10) alapján kiszámolt helyzete sem felel meg az elvárt pozíciónak. Ennek következtében az 5.10. ábrához hasonló elrendeződést kapunk.

5.5.2.2. Optimalizáció 2.b ♦ Az algoritmus második verziójában az R nagyítást változtatlanul az $I_R = [0,3; 0,4]$ intervallum 21 különböző értéke szerint választottam meg, a C' pont lehetséges pozícióit ez esetben is az 5.13. ábra 21 köríve mentén, a nyelvcsont árnyéksávjában lévő 208 pont alkotta, de ezúttal a Φ_0 szögelfordulásra végeztem optimalizációt, vagyis a *for* ciklust kizárólag erre a változóra alapoztam, miközben R és C' előre meghatározott értékeit táblázatokból olvastam be. A 208 individuális ciklus futtatása után az eredmények sokaságából kikerestem azt a $\{C', R, \Phi_0\}$ paraméterhármaszt, ami a legkisebb NND távolságot biztosítja az optimalizációnak alávetett UH- és MRI-kontúrok között. Az optimalizáció végeredményét az 5.4. táblázat rendszerezi.

5.4. táblázat. Az Optimalizáció 2.b algoritmus futtatásával kapott eredmények.

R	Számolt C' (px)	Számolt Φ_0 (fok)	Optimalizált Φ_0 (fok)	MIN (px)
0,355	(127; 233)	15,48	28,85	37,87

Látható, hogy az optimalizált szögelfordulás közel duplája a számított értéknek, ami jelentős különbségként könyvelhető el, és ennek nyomán megint az tapasztalható, hogy a gégefedő csúcspontjának valódi pozíciója a Φ_0 optimalizált értékével (5.10) alapján kiszámolt koordinátákkal nem kapható vissza. Így – bár C' eleve megfelelő módon történő rögzítése folytán a nyelvcsont árnyéksávjában marad – a gégefedő csúcspontja az 5.10. ábrához hasonló mértékű eltérést mutat a korrekt helyzetéhez képest.

Az 5.5.2.1. és 5.5.2.2. paragrafusokban tárgyalt optimalizációs eljárásokkal tehát szintén nem teljesíthetők az 5.3. alfejezet feltételei, ezért további megfontolások bevezetésével más szemszögből próbáltam meg megoldani a problémát.

5.5.3. Optimalizáció 3

Miután a korábbi alfejezetekben kifejtett optimalizációs módszereknek éppen az volt az egyik gyenge pontja, hogy nem voltak képesek reprodukálni a gégefedő csúcspontjának helyes pozícióját, így a továbbiakban olyan módon szerkesztettem meg a programokat, hogy az optimalizációs algoritmusokba eleve beépüljön a gégefedő csúcspontjának (5.1) által rögzített helyzete az (5.2)-(5.5) transzformációs formulák közvetítésével. Ez azt jelenti, hogy az (5.2)-(5.5) összefüggések felhasználásával matematikai feltételek mentén kaphatjuk meg a $\{C', R, \Phi_0\}$ paraméterhalmaz elemeit. Ebben a megközelítésben is két lehetséges utat jártam végig, amiket a következő két paragrafusban taglalok.

5.5.3.1. Optimalizáció 3.a ♦ Az optimalizáció során a C' pont x és y koordinátái egymásba ágyazott *for* ciklusok mentén futnak, miközben az R nagyítás és a Φ_0 szögelfordulás az (5.2)-(5.5) összefüggések révén adódik a ciklusváltozók aktuális értékei szerint, ami az

(5.3)-(5.5) egyenletek kombinálásával az

$$R = \frac{\sqrt{(G'_x - C'_x)^2 + (C'_y - G'_y)^2}}{R_1},$$

$$\Phi_0 = \phi_1 + \arctan\left(\frac{G'_x - C'_x}{C'_y - G'_y}\right) \quad (5.13)$$

kifejezésekre vezet, ahol G'_x , G'_y , R_1 , illetve ϕ_1 értékei (5.2) szerint rögzülnek. Másképpen megfogalmazva az optimalizáció szabadsági foka C' x és y koordinátáinak szintjén realizálódik, R és Φ_0 értékeit pedig az (5.13) geometriai kondíciók szabják meg az optimalizált C'_x és C'_y értékeken keresztül. Az algoritmus eredményeit az 5.5. táblázat mutatja be.

5.5. táblázat. Az Optimalizáció 3.a algoritmus futtatásával kapott eredmények.

R	Φ_0 (fok)	C' (px)	MIN (px)
0,20	44,08	(141; 183)	31,46

5.5.3.2. Optimalizáció 3.b ♦ Az optimalizáció során az R nagyítás és a Φ_0 szögelfordulás egymásba ágyazott *for* ciklusok mentén fut, miközben a C' pont x és y koordinátáit az (5.10) relációk definiálják a ciklusváltozók aktuális értékei szerint, ami (5.10) átrendezésével a

$$C'_x = G'_x - R R_1 \frac{\tan(\Phi_0 - \phi_1)}{\sqrt{1 + \tan^2(\Phi_0 - \phi_1)}},$$

$$C'_y = G'_y + R R_1 \frac{1}{\sqrt{1 + \tan^2(\Phi_0 - \phi_1)}} \quad (5.14)$$

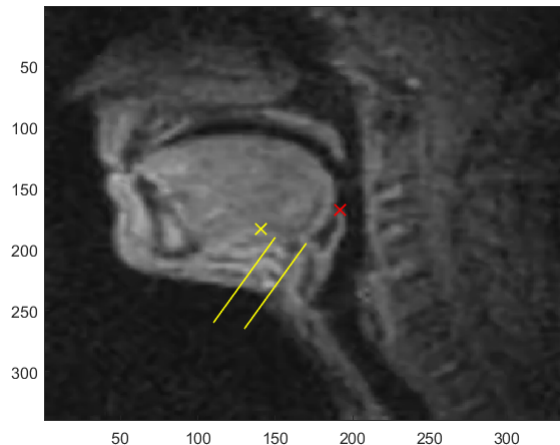
formulákat generálja, ahol G'_x , G'_y , R_1 , illetve ϕ_1 értékeit továbbra is (5.2) adja meg. Jelen esetben tehát az optimalizáció szabadsági fokát R és Φ_0 biztosítja, C' x és y koordinátáit pedig az (5.14) feltételek határozzák meg az optimalizált R és Φ_0 értékeken keresztül. Az algoritmus eredményeit az 5.6. táblázat prezentálja.

5.6. táblázat. Az Optimalizáció 3.b algoritmus futtatásával kapott eredmények.

R	Φ_0 (fok)	C' (px)	MIN (px)
0,21	43,58	(140,46; 183,65)	31,44

Az 5.5. és 5.6. táblázatban foglalt eredmények összehasonlításával megállapítható, hogy a két megközelítés minden paraméter esetében nagyon hasonló értékeket produkál. Miután a gégefedő csúcspontja eleve az anatómiailag helyes pozícióba volt elhelyezve,

az optimalizációs algoritmusok kimeneteleinek ellenőrzése ezúttal a C' pont helyzetének tesztelésére korlátozódik. Az 5.5. és 5.6. táblázatok tanúsága szerint a C'_x és C'_y koordináták mindössze 1-1 pixelnyi eltérést mutatnak, tehát lényegében ugyanazon pozícióról van szó, ami az 5.14. ábra sárga keresztjele szerint a nyelvcsont árnyéksávján kívül, a nyelvszövet belsejében lokalizálódik.



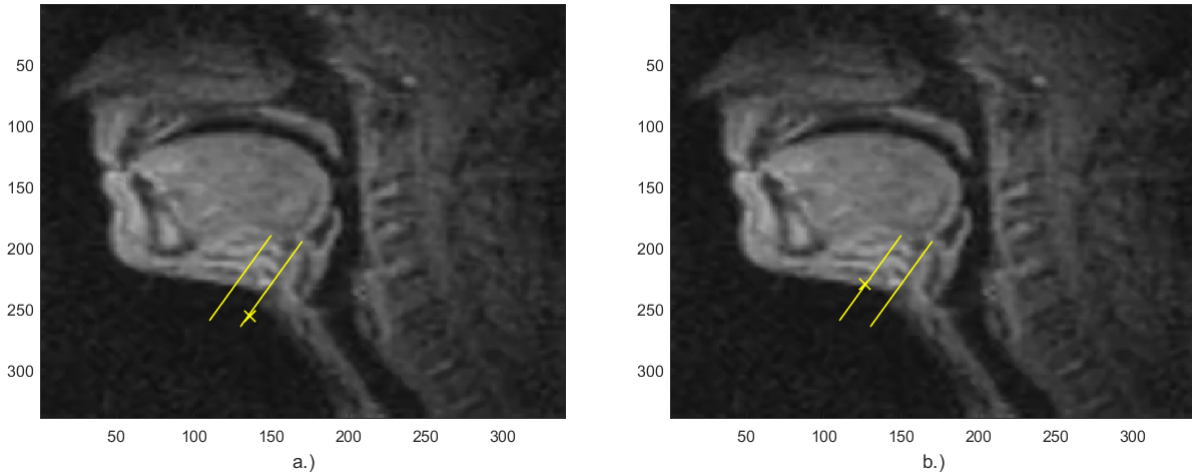
5.14. ábra. Az optimalizáció eredményeképpen kapott C' (sárga) pont helyzete a nyelvcsont árnyéksávjához (sárga), illetve a gégefedő csúcspontjának valódi pozíciójához (piros) viszonyítva. (Az ábra tengelyeinek mértékegysége pixel.)

Látható, hogy sajnos az 5.5.3.1. és 5.5.3.2. paragrafusokban kifejtett optimalizációs algoritmusok ismét túlmutatnak az 5.3. alfejezet elvárásain, így újabb nézőpontok mentén kellett továbbhaladnom.

5.5.4. Optimalizáció 4

A korábbi alfejezetek tapasztalatai arra a felismerésre vezettek, hogy a $\{C', R, \Phi_0\}$ mennyiségek optimalizációja nem ad hitelesíthető eredményeket abban az esetben, ha egyidejűleg több paraméter idealizálása révén szeretnénk kialakítani a $\{C', R, \Phi_0\}$ halmaz elemeinek legkedvezőbb értékeit. Megfigyelhettük azt is, hogy bizonyos geometriai és anatómiai megszorítások kijelölése sem képes reális mederben tartani az optimalizáció szabadsági fokait. Ennélfogva úgy próbáltam javítani a módszeren, hogy kizárólag a C' pont egyik koordinátáját definiáltam a *for* ciklus változójaként, miközben a másik koordinátát (5.14) megfelelő relációjának alkalmazásával korlátoztam. Emellett a gégefedő csúcspontjának koordinátáit továbbra is állandónak tekintettem (5.1) szerint, az R nagyításra vonatkozóan fenntartottam az (5.3)-(5.4) összefüggéseket, a Φ_0 szögelfordulást pedig egy adott intervallumon belül változtatva és az algoritmus futása során manuálisan rögzítve tanulmányoztam a C' pont optimalizált helyzetének helyességét. A Φ_0 szögelfordulás lehetséges értékeit az $I_{\Phi_0} = [0,05; 0,3]$ intervallum 0,01-os lépésközzel megválasztott elemei mentén jelöltem ki, ami fokban kifejezve az $I_{\Phi_0} = [2,86^\circ, 17,19^\circ]$ intervallumot definiálja 0,57°-os lépésközzel, és először a C'_y , majd a C'_x koordinátára végeztem el az optimalizációt. Az I_{Φ_0} intervallum kezdő- és végpontját vizuális úton határoztam meg úgy, hogy a két érték beállításakor a C' pont két olyan szélsőséges helyzetben legyen, ami a nyelvcsont árnyéksávjának határvonalán realizálható az 5.15. ábra tanúsága szerint, ahol a sárga keresztjel pozicionálja

a C' pontot. Az 5.15.a és 5.15.b ábrákon C' alsó és felső határhelyzete a $\Phi_0 = 0,05$ és a $\Phi_0 = 0,3$ értékeknek feleltethető meg. C' 5.15. ábrán szemléltetett szélső pozíciói a C'_y és C'_x koordinátákra írt optimalizációs algoritmusok esetében egy-egy pixelnyi eltéréstől eltekintve lényegében ugyanazon pontokban rögzülnek.



5.15. ábra. A C' pont két szélső helyzete a $\Phi_0 = 0,05$ (a.) és a $\Phi_0 = 0,3$ (b.) szögelfordulási értékek esetén. (Az ábra tengelyeinek mértékegysége pixel.)

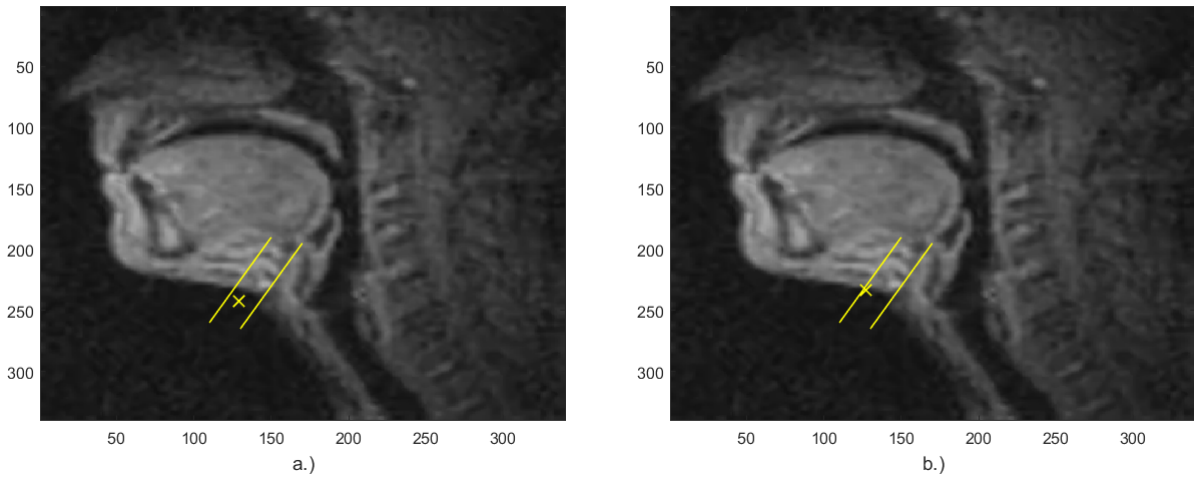
Az optimalizáció eredményeit az 5.7. táblázatpár sorakoztatja fel. A bal oldali táblázat a C'_y , a jobb oldali táblázat pedig a C'_x koordinátára megvalósított optimalizáció esetében érvényes.

5.7. táblázat. Az *Optimalizáció 4* algoritmus futtatásával kapott eredmények.

Φ_0 (rad)	R	C'_x (px)	C'_y (px)	MIN (px)
0.05	0.4012	136.1734	256	106.58
0.06	0.3992	135.5798	255	104.49
0.07	0.4019	134.3358	255	102.44
0.08	0.3999	133.7525	254	100.44
0.09	0.3980	133.1814	253	98.46
0.10	0.3961	132.6224	252	96.46
0.11	0.3941	132.0757	251	94.43
0.12	0.3922	131.5411	250	92.39
0.13	0.3904	131.0188	249	90.36
0.14	0.3885	130.5086	248	88.34
0.15	0.3866	130.0108	247	86.34
0.16	0.3848	129.5252	246	84.38
0.17	0.3829	129.0519	245	82.54
0.18	0.3811	128.5910	244	80.82
0.19	0.3742	128.9939	242	79.18
0.20	0.3723	128.5753	241	77.54
0.21	0.3705	128.1695	240	75.10
0.22	0.3686	127.7766	239	74.54
0.23	0.3615	128.3197	237	73.16
0.24	0.3596	127.9716	236	71.73
0.25	0.3577	127.6371	235	70.41
0.26	0.3558	127.3164	234	69.16
0.27	0.3539	127.0094	233	67.94
0.28	0.3520	126.7165	232	66.76
0.29	0.3501	126.4377	231	65.62
0.30	0.3425	127.2350	229	64.52

Φ_0 (rad)	R	C'_x (px)	C'_y (px)	MIN (px)
0.05	0.4025	136	256.2733	106.60
0.06	0.4033	135	255.8940	104.53
0.07	0.3972	135	253.9979	102.55
0.08	0.3982	134	253.6346	100.46
0.09	0.3992	133	253.2621	98.47
0.10	0.3935	133	251.4659	96.56
0.11	0.3946	132	251.1048	94.43
0.12	0.3958	131	250.7339	92.55
0.13	0.3905	131	249.0249	90.36
0.14	0.3917	130	248.6617	88.50
0.15	0.3867	130	247.0137	86.34
0.16	0.3818	130	245.4072	84.51
0.17	0.3832	129	245.0635	82.55
0.18	0.3786	129	243.5098	80.80
0.19	0.3742	129	241.9929	79.18
0.20	0.3757	128	241.6621	77.61
0.21	0.3714	128	240.1912	76.02
0.22	0.3673	128	238.7531	74.51
0.23	0.3633	128	237.3464	73.08
0.24	0.3594	128	235.9699	71.74
0.25	0.3557	128	234.6223	70.49
0.26	0.3575	127	234.3228	69.22
0.27	0.3539	127	233.0094	67.95
0.28	0.3505	127	231.7221	66.73
0.29	0.3471	127	230.4597	65.56
0.30	0.3438	127	229.2214	64.43

Összehasonlítva a két táblázat adott Φ_0 értékekhez tartozó sorait, megállapítható, hogy az összes paraméter számértékei néhány századnyi vagy néhány ezrednyi eltéréstől eltekintve gyakorlatilag megegyeznek, az $\{R, C'_x, C'_y\}$ paraméterek azonban csak a sárgával kiemelt $\Phi_0 = 0,19 = 10,89^\circ$, illetve $\Phi_0 = 0,27 = 15,47^\circ$ szögelfordulási értékek esetében mutatnak tökéletes egybeesést, amennyiben a C'_x és C'_y oszlopaiban lévő adatokat egész értékekre kerekítjük. Ezen numerikus jellemzőkkel szembesülve logikusnak tűnt az említett egybeesések alapján kiválasztani az optimalizáció által generált adathalmazból a leginkább elfogadható megoldást, hiszen az az eredmény tekinthető a leghitelesebbnek, amit mindkét algoritmus egzakt formában megerősít. Innentől kezdve tehát azt kellett eldöntenem, hogy a $\Phi_0 = 0,19$ vagy a $\Phi_0 = 0,27$ szögelfordulási értékhez rendelt paraméterhalmaz produkálja-e a jobb eredményt. Ezt a kérdést kétféle aspektusból közelítettem meg. Egyrészt mérhető lehet az, hogy az 5.7. táblázatpár utolsó oszlopainak tanúsága szerint a $\Phi_0 = 0,27$ érték esetében kisebb az optimalizációban résztvevő UH- és MRI-kontúrok között mért NND távolság minimuma, ami kétségtelenül hatékonyabb eredményként értékelhető. Másrészt azonban azt is érdemes megvizsgálni, hogy a kapott C' pont hol helyezkedik el a nyelvcsont árnyéksávján belül, így a két nézőpont együttes mérlegelésével kialakítható a végső eredmény.



5.16. ábra. A $\Phi_0 = 0,19$ (a.), illetve a $\Phi_0 = 0,27$ (b.) szögelfordulási értékek beállításakor kapott C' pont pozíciói. (Az ábra tengelyeinek mértékegysége pixel.)

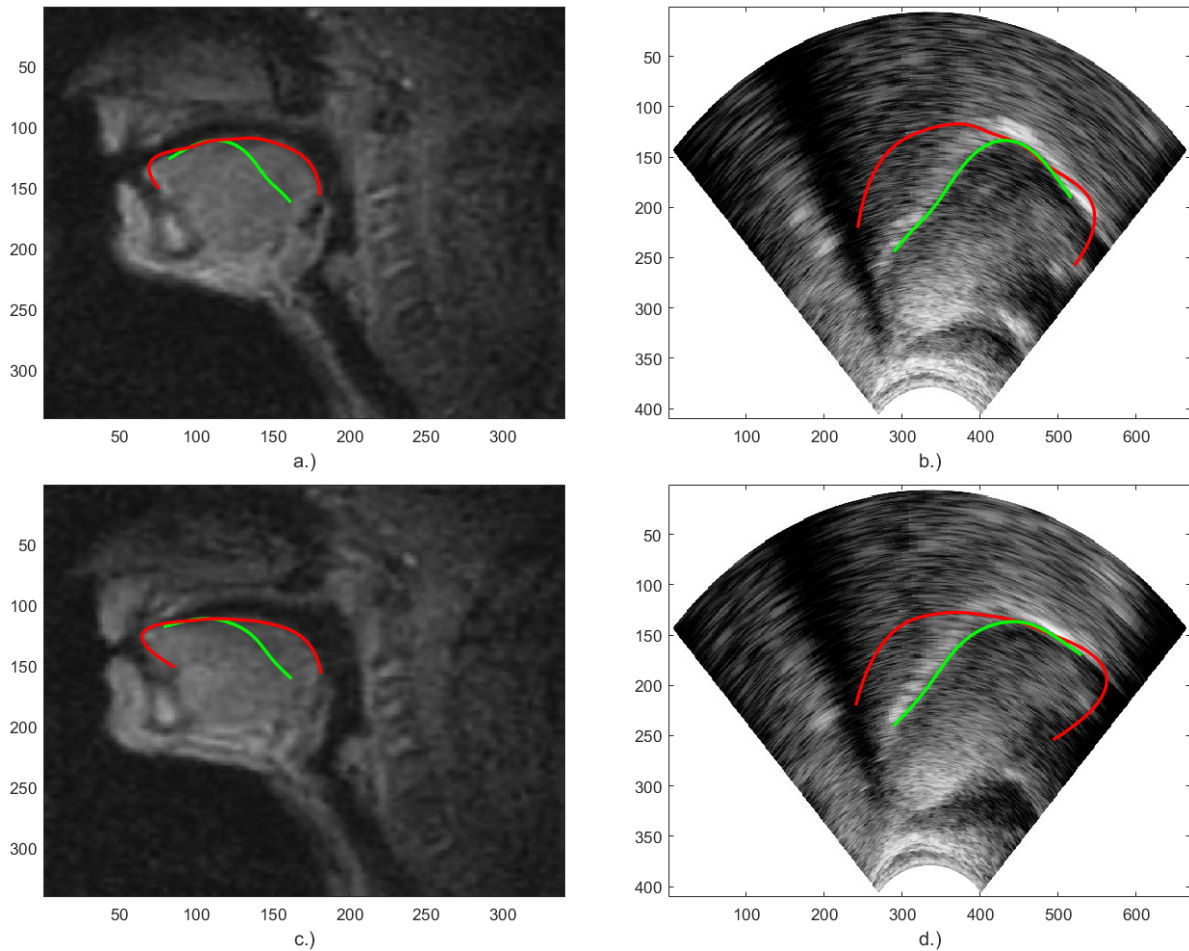
Az 5.16.a és 5.16.b ábrák rendre a $\Phi_0 = 0,19$, illetve a $\Phi_0 = 0,27$ szögelfordulási értékek beállításakor adódó C' pont pozícióit demonstrálják. Miután az optimalizáció során mindig azt tartottam szem előtt, hogy az UH- és MRI-felvételek anatómiai és geometriai elemeinek pozíciói egymásnak kölcsönösen egyértelmű transzformáltjai legyenek, mindenképpen kontrollálni kell, hogy az MRI-keretek optimalizált C' pontja hogyan viszonyul az UH-keretek megfelelő pontjához. Az 5.5. és 5.16. ábrák összevetésekor megfigyelhetjük, hogy az UH-felvétel C pontjának és az MRI-felvétel C' pontjának relatív helyzete a $\Phi_0 = 0,19$ érték esetében nagyon hasonló, hiszen ekkor C és C' a nyelvcsont árnyéksávjának közepe tájékán helyezkedik el, míg a $\Phi_0 = 0,27$ érték megválasztásakor C' egészen közel van az árnyéksáv felső határához. Ezen észrevétel alapján a $\Phi_0 = 0,19$ értékhez tartozó paraméterhalmazt preferáltam, annak ellenére, hogy ekkor – visszautalva az 5.7. táblázatra – nagyobb a minimalizált NND távolság. Az optimalizáció végső eredményét tehát az 5.8. táblázat emeli ki.

5.8. táblázat. Az *Optimalizáció 4* algoritmus futtatásával kapott legvégső eredmények.

Φ_0 (fok)	R	C'_x (px)	C'_y (px)	MIN (px)
10,89	0,37	129	242	79,18

5.5.5. Az optimalizáció eredményeinek vizuális validálása

Az 5.5.4. alfejezetben kapott eredmények alátámasztása céljából megvizsgáltam, hogy az optimalizációnak alávetett és párba rendelt UH- és MRI-kontúrok hogyan viszonyulnak egymáshoz, azaz a szájpaddockontúrok és az adott hanghoz tartozó nyelvkontúrok milyen mértékben illeszkednek. A vizuális ellenőrzés során az volt a mérvado irányelv, hogy az egymásnak megfeleltetett UH- és MRI-kontúrok a szájüregi régió releváns tartományán belül maradjanak, és szubjektív megítélés alapján elfogadható legyen a két görbe közötti átfedés jósága.

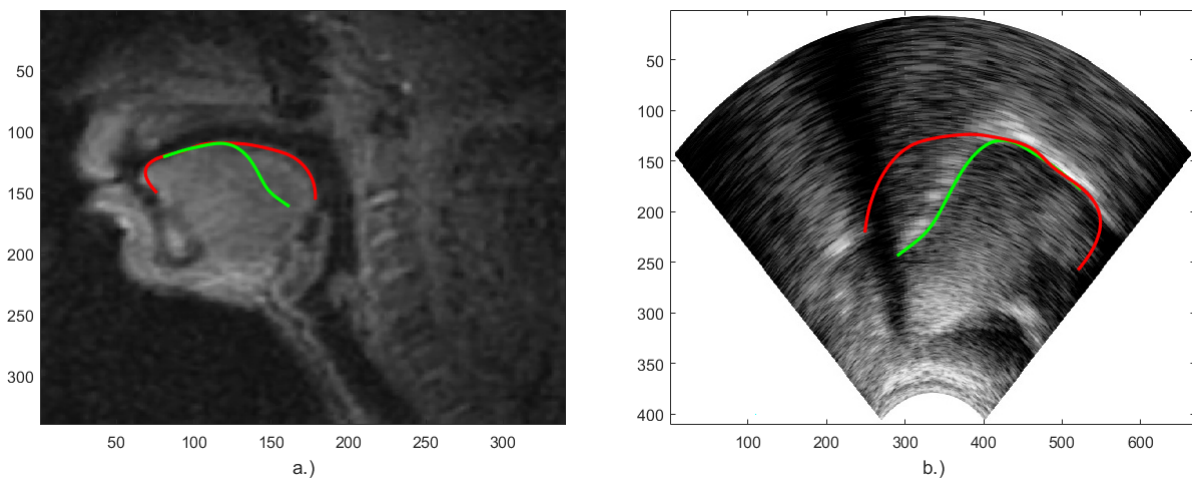


5.17. ábra. Az optimalizáció vizuális eredményei az *e* (a. és b.) és *t* (c. és d.) hangokhoz tartozó UH- (zöld) és MRI-nyelvkontúrok (piros) esetében. (Az ábra tengelyeinek mértékegysége pixel.)

Ennek nyomán az 5.17. ábra példaként illusztrálja a grafikus eredmények egy részét, ahol

az UH-kontúrok zölddel, az MRI-kontúrok pedig pirossal vannak megrajzolva. Az 5.17.a és 5.17.c ábrákon az adott MRI-keretre vetített UH-nyelvkontúr látható az e és t hangok esetében. Ez az ábrázolás megfelel az $UH \rightarrow MRI$ irányban megvalósított transzformációnak és optimalizációnak (5.4. alfejezet). Az 5.17.b és 5.17.d ábrák az ezzel ellentétes $MRI \rightarrow UH$ irányú projekciót jelenítik meg ugyanazon e és t hangok esetében, amennyiben a transzformációt az 5.8. táblázat paramétereinek alkalmazásával hajtjuk végre. Az ábrák alapján nyugtázható, hogy a transzformált nyelvkontúrok nem lépnek ki a nyelv által definiált zónából, és ilyen értelemben véve elfogadható módon közelítik a viszonyítási görbeként szolgáló eredeti nyelvkontúrt. A tökéletes illeszkedés természetesen nem szerepelhet szigorú elvárásként, hiszen különböző nemű, anyanyelvű és anatómiai berendezkedésű beszélők felvételeivel dolgoztam, ezért kielégítőnek találtam a fentebb jelzett kritérium teljesülését, így az optimalizáció kimenetelét vizuálisan reálisnak tekintem.

Eredményeim további megerősítése végett olyan UH- és MRI-nyelvkontúrpárok illeszkedését is tanulmányoztam, melyek nem szerepeltek a kijelölt $\{a, e, o, k, s, t\}$ hanghalmaz elemeihez tartozó nyelvkontúrok között. A görbék egymásra vetítését ez esetben is az 5.8. táblázat paramétereinek beállításával vittem véghez, és azt tapasztaltam, hogy a görbepárok közötti egyezés kvalitatíve nagyon hasonló az 5.17. ábrán bemutatott mintákhoz. Ezt igazolja az 5.18. ábra, ami az $\ddot{o}$ hang esetében előálló UH- és MRI-nyelvkontúrokat példázza hasonló szellemben az 5.17. ábrához. Ez a vizsgálat tehát azt tanúsítja, hogy az optimalizáció eredményei a figyelembe vett hanghalmazon kívül eső hangokra is sikeresen alkalmazhatók.

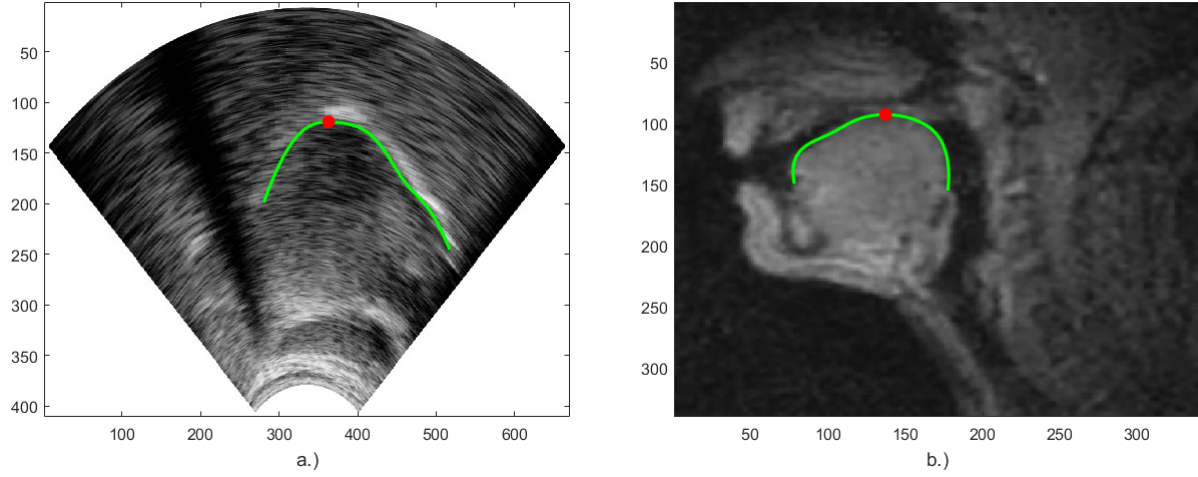


5.18. ábra. Az optimalizáció vizuális eredményei az $\ddot{o}$ hanghoz tartozó UH- (zöld) és MRI-nyelvkontúrok (piros) esetében. (Az ábra tengelyeinek mértékegysége pixel.)

5.5.6. Az optimalizáció eredményeinek numerikus validálása

Az 5.5.4. alfejezetben prezentált eredményeket a kvalitatív validálás mellett kvantitatív módon is megpróbáltam verifikálni. Ennek első lépéseként az R nagyítást határoztam meg. A számításaimhoz a 40 UH-bemondásban található összes k hanghoz rendelt keretet és az aka hangsort tartalmazó MRI-bemondás k hangjához tartozó keretblokk középső keretét vettem alapul. A jelzett UH- és MRI-keretek mindegyikén megkerestem az aktuális nyelvkontúr legmagasabb pozícióban lévő pontját, ami a k hang esetében a nyelvhat középső

tartományában lokalizálható. Ezt tükrözi az 5.19. ábra, melynek a.) része egy tetszőlegesen kiválasztott UH-kereten, b.) része pedig az MRI-kereten azonosítható nyelvkontúr pirossal megjelölt maximumpontját emeli ki.



5.19. ábra. Egy tetszőlegesen kiválasztott UH-kereten (a.), illetve az MRI-kereten (b.) felvett nyelvkontúr maximumpontja a k hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)

Ezt követően minden UH-kereten és az MRI-kereten is megmértem a nyelvkontúr maximumpontja és a nyugalomban lévő gégefedő G' csúcspontja közötti távolságot, majd (5.4) értelmében képeztem az MRI-kereten mért távolság, illetve a felhasznált 489 darab UH-kereten meghatározott távolságok hányadosát, amivel 489 különböző nagyságértékhez jutottam. Végül kiszámítottam a 489 paraméter átlagát, ami az

$$R = 0,3736 \quad (5.15)$$

értékre vezetett. Visszalépve az 5.2. ábra geometriájához, (5.4) átrendezésével megadtam az

$$R_2 = R R_1 = 0,3736 \cdot 259,7306 = 97,0354 \quad (5.16)$$

sugarat, majd (5.3) első egyenlőségének felhasználásával, és az 5.8. táblázatban foglalt optimalizált C'_x koordináta helyettesítésével kiszámítottam az

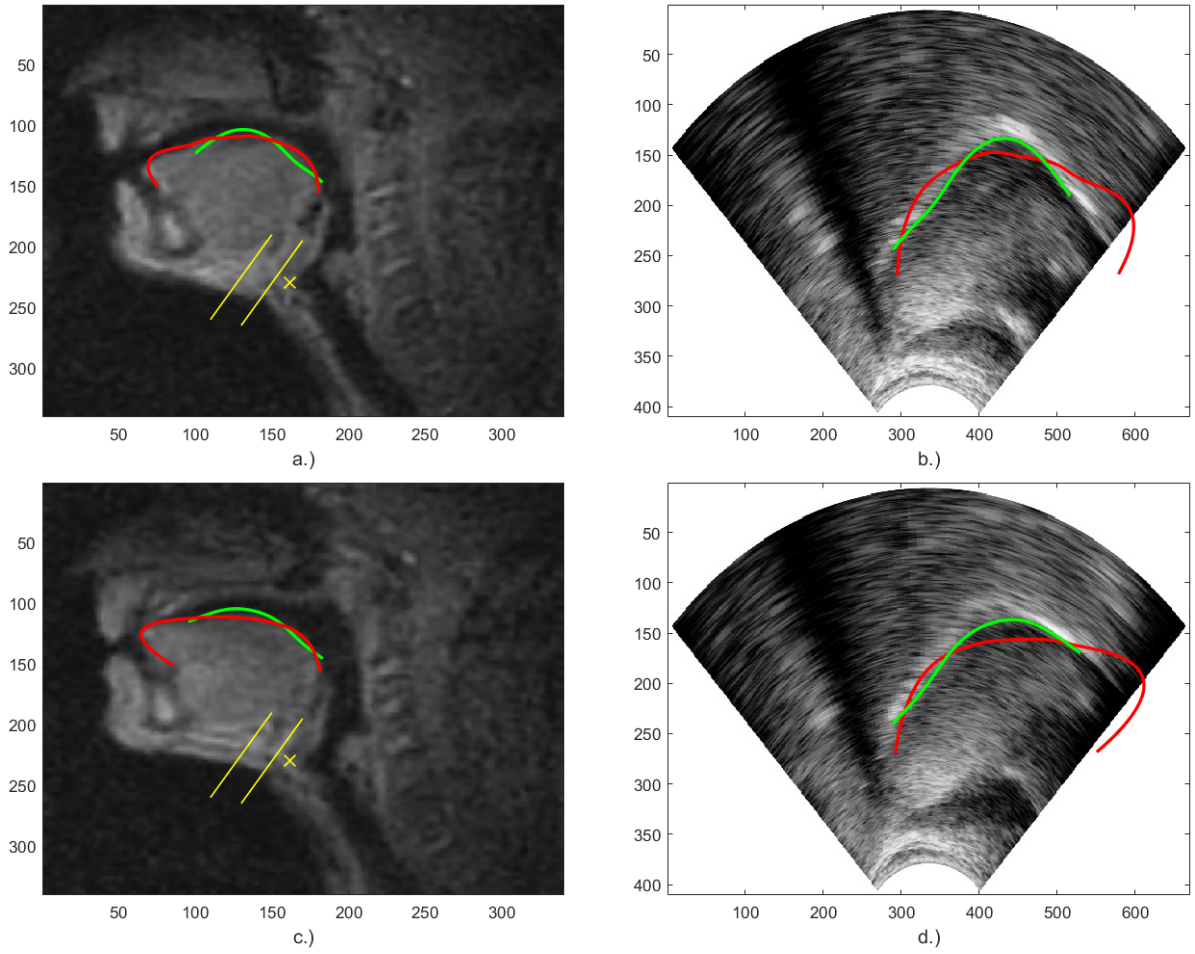
$$x_2 = G'_x - C'_x = 192 - 129 = 63 \quad (5.17)$$

távolságot. Ezt követően az (5.16) és (5.17) értékek segítségével trigonometrikus úton meghatároztam a

$$\phi_2 = \arcsin\left(\frac{x_2}{R_2}\right) = \arcsin\left(\frac{63}{97,0354}\right) = 40,4849^\circ \quad (5.18)$$

szöget. Végül (5.5) szerint megkaptam a

$$\Phi_0 = \phi_1 + \phi_2 = -29,5260^\circ + 40,4849^\circ = 10,9589^\circ \quad (5.19)$$



5.20. ábra. A kevésbé szigorú optimalizáció vizuális eredményei az e (a. és b.) és t (c. és d.) hangokhoz tartozó UH- (zöld) és MRI-nyelvkontúrok (piros) esetében. (Az ábra tengelyeinek mértékegysége pixel.)

szögelfordulást is. Az (5.15) és (5.19) numerikus adatok egyértelműen reprodukálják az 5.8. táblázat megfelelő paramétereit, ami azt bizonyítja, hogy az optimalizáció eredményei mérvadó módon elfogadhatók.

Ezen a ponton fontosnak tartom rámutatni arra, hogy az optimalizációs eljárással jobb eredményeket is el lehet érni abban az esetben, ha nem ragaszkodunk mereven az 5.3. alfejezetben kiszabott anatómiai és geometriai megfontolásokhoz. Abból a célból, hogy ezt az állítást alátámasszam, az optimalizációt olyan körülmények között is végrehajtottam, amikor a $\{C', R, \Phi_0\}$ paraméterhalmaz értékeit nagyobb szabadsági fokok mentén kereshetjük. Ezt olyan programkonstrukcióban valósítottam meg, hogy a nagyítást az 5.8. táblázat, illetve az (5.15) számítás eredményeinek figyelembevételével az $R = 0,37$ érték szerint választottam meg, azonban a transzformáció középpontjának $\{C'_x, C'_y\}$ koordinátáit, valamint a szögelfordulást szabadon futtattam mindenféle korlátozás vagy extra feltétel nélkül. Az optimalizáció során a $C'_x = 162$, $C'_y = 230$, $\Phi_0 = 0,05$ numerikus értékeket kaptam, ami vizuálisan sokkal meggyőzőbb illeszkedéshez vezetett az UH- és MRI-kontúrok viszonylatában, ahogyan azt az 5.20. ábra is igazolja az e és t hangok esetében. Az 5.20. ábra blokkjai egyenként megfeleltethetők az 5.17. ábra blokkjainak, és megfigyelhetjük, hogy valóban szembetűnő a különbség, hiszen az 5.20. ábra esetében nagyobb mértékű

és realisabb az átfedés a görbepárok között. Az 5.20.a és 5.20.c ábrákon sárga kereszttel megjelöltem a C' pont optimalizált pozícióját is, ami ezúttal a nyelvcsont árnyéksávján kívül található, tehát ez a megközelítés nem biztosítja azt a helyzetet, ami az 5.5. ábra szerint teljesül az UH-felvételek esetében. Ez a vizsgálat végső soron azt mutatja, hogy az 5.17. ábra görbepárjainak kevésbé jó illeszkedése az 5.3. alfejezetben megkövetelt kritériumok érvényesítésének következménye. Mindezek ellenére mégis ragaszkodtam az 5.3. alfejezetben megkövetelt korlátozásokhoz, mivel a kutatómunkát az motiválta, hogy különböző beszélőktől származó felvételeken tudjam a legjobb illesztést létrehozni a kialakított geometriai keretrendszer viszonyítási pontjainak egységes kezelésével.

5.6. T1 tézis

A kétdimenziós UH- és MRI-felvételek radiális és négyszöges struktúrái egymásba ágyazhatók olyan geometriai transzformációk révén, melyek megvalósítják a felvételek nyelv- és szájpaddocktúrajai által definiált geometriai tartományok kétirányú konverzióját: a radiális tartomány skálázását, illetve a szögtartomány forgatását. Az operációkhoz kapcsolódó radiális nagyítás, szögelfordulás, valamint a transzformációk középpontjának optimalizálásával elérhető az UH- és MRI-felvételek anatómiai tartományai közötti legjobb átfedés. (Kapcsolódó saját publikációk: [S1],[S2],[S4],[S5],[S9],[S11])

6. UH- és MRI-nyelvkontúrokkal megvalósított gépi tanítás

Kutatómunkám második terjedelmesebb szakaszában továbbra is az UH- és MRI-felvételek összehangolását céloztam meg, melynek kellékei a 3. fejezetben bevezetett nyelvkontúrok voltak. Ezúttal azonban az analitikus geometriai szemléleten és az abból adódó egzakt matematikai transzformációkon alapuló optimalizációs megközelítést mesterséges intelligencia alkalmazásával váltottam fel. Ez azt jelenti, hogy a két forrás nyelvkontúrait gépi tanulóalgoritmusok segítségével kapcsoltam össze, melynek során a neurális hálózatot eltérő konstrukciók szerint alkottam meg, és a kontúrok alaki sajátosságait két különböző jellegű paraméterrel vettem figyelembe. A gépi tanítást a rendszer szerkezete, valamint bemeneti és kimeneti adatainak száma és típusa tekintetében is többféle konfigurációban valósítottam meg. A következő alfejezetekben részletesen bemutatom a tanításhoz felhasznált hangkészletet, a nyelvkontúrok jellemzésére értelmezett paramétereket, a tanulóalgoritmus fajtáját, illetve a különböző rendszerkonfigurációkban kapott eredményeket. A gépi tanulóalgoritmusok programkódjai az alábbi linken tekinthetők meg:

https://drive.google.com/drive/folders/1UTGLfxObIDkqanai0zSS_IJMRN_kpUdM?usp=sharing

6.1. A gépi tanítás eszközei

Abból a célból, hogy elemzéseim következetesek legyenek, az optimalizációs fázishoz hasonlóan a gépi tanítás során is az $\{a, e, o, k, s, t\}$ hanghalmazt vettem alapul, így változatlanul ugyanazok a nyelvállások szerepelhettek a vizsgálataimban. Ez esetben azonban a gépi tanulás hatékonyságának növelése érdekében biztosítanom kellett a tanítóalakzatok

elegendően nagy számát, ami azt vonta maga után, hogy az UH-nyelvkontúrok sokaságán kezdetben egyáltalán nem végeztem semmiféle szelekciót, azaz a 40 különböző bemondást tartalmazó UH-csomagból kiemeltem az 1 084 *a*-kontúrt, 386 *e*-kontúrt, 853 *o*-kontúrt, 489 *k*-kontúrt, 205 *s*-kontúrt és 1 062 *t*-kontúrt. Ez összesen egy 4 079 elemből álló tömböt eredményezett. Az adott hanghoz rendelt görbesereg átlagához legközelebb eső kontúr helyett tehát az összes görbét felhasználtam. Ezzel nemcsak a tanítóalakzatok viszonylag nagy számát értem el, hanem a mintázatok változatosságáról is gondoskodtam, hiszen az adott hanghoz tartozó UH-nyelvkontúrok meglehetősen sokféle variációban ölthetnek alakot (pl. 5.9. ábra), ami a neurális hálózatok trenírozásában is nyomós szempont lehet. Miután az MRI-csomag adott hanghoz rendelt kereteinek száma sokkal kisebb, mint az UH-csomag ugyanazon hangjához rendelt keretek száma, így alaphelyzetben nem garantálható a két adathalmaz dimenziójának egyezősége, ami viszont elengedhetetlen az alkalmazott neurális hálózatokban. Ezért az MRI-nyelvkontúrok esetében az 5.4. alfejezetben ismertetett kiválasztási mechanizmust alkalmaztam, vagyis az adott hanghoz rendelt nyelvkontúrt a releváns keretek csoportjának középső elemén rögzített görbeként definiáltam, majd a kijelölt kontúrt egyszerűen megismételtem a kívánt érték szerint (pl. az *a* hang esetében 1 084-szer).

Fontos megjegyezni azt is, hogy a 4. fejezetben megállapított UH- és MRI-szájpadkontúrokat nem csatoltam hozzá az $\{a, e, o, k, s, t\}$ hangkészlethez, mivel lényegében egy-egy rögzített kontúrt kellene egymással párba állítani, így egyedülként szereplő elemekként, alaki változatosság híján ezen görbék nem kompatibilisek a neurális hálózatok fő követelményeivel.

6.2. A nyelvkontúrok paraméterei

6.2.1. Tanítópontok

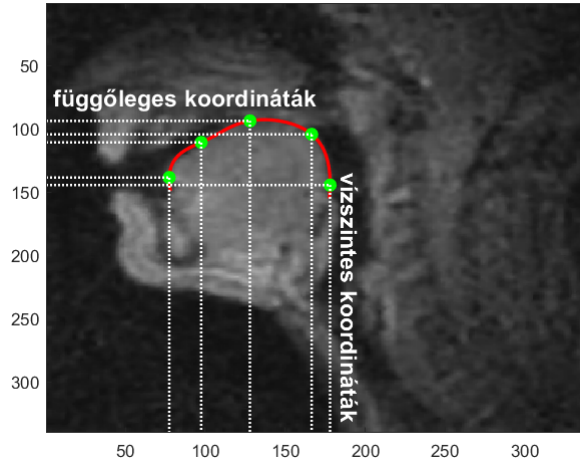
A nyelvkontúrok parametrizálásának egyik legegyszerűbb módja az, ha kijelölünk egy véges számú elemből álló diszkrét pontsorozatot a görbe mentén. Ekkor a p számú tanítópont vízszintes koordinátáinak

$$P_x = \{P_{x_1}, P_{x_2}, \dots, P_{x_p}\} \quad (6.1)$$

halmaz a nyelv előre-hátra irányú mozgásának tanulmányozásában lehet releváns, függőleges koordinátáinak

$$P_y = \{P_{y_1}, P_{y_2}, \dots, P_{y_p}\} \quad (6.2)$$

csoportja pedig jó jelöltekkel szolgálhat a nyelv fel-le irányú mozgásának vizsgálatában. A tanítópontokat rendkívül könnyű hozzárendelni a nyelvkontúrokhoz, de mivel a pontok kiválasztása diszkrét, a tanítópontok koordinátái csak lokális közelítést adhatnak, a görbe globális tulajdonságai azonban nem tárulnak fel. A 6.1. ábra $p = 5$ tanítópont egy lehetséges elrendezését illusztrálja egy adott MRI-nyelvkontúr mentén, ahol a tanítópontok a 392 pontból álló görbe teljes szögtartományának 10, 30, 50, 70 és 90 %-ánál vannak rögzítve a görbe mentén jobbról balra mérve.



6.1. ábra. Egy adott MRI-nyelvkontúr (piros) mentén speciálisan megválasztott 5 tanítópont (zöld) a vízszintesen és függőlegesen levetített koordináták bejelölésével. (Az ábra tengelyeinek mértékegysége pixel.)

6.2.2. DCT-együtthatók

A nyelvkontúrok kezelésének egy másik lehetséges módja a diszkrét koszinusztranszformáció (Discrete Cosine Transform – DCT) alkalmazása, amely a matematikai transzformációknak egy speciális válfaja [65]. A DCT segítségével egy N elemű valós $\{x_1, x_2, \dots, x_N\}$ adathalmazt egy ugyanolyan elemszámú, szintén valós $\{X_1, X_2, \dots, X_N\}$ értékhalmaszra konvertálhatunk az

$$X_k = \sqrt{\frac{2}{N}} \sum_{n=1}^N x_n \frac{1}{\sqrt{1 + \delta_{k1}}} \cos\left(\frac{\pi}{2N}(2n-1)(k-1)\right) \quad (6.3)$$

formula szerint, ahol $k = 1, 2, \dots, N$. (6.3) alapján az inverz operáció

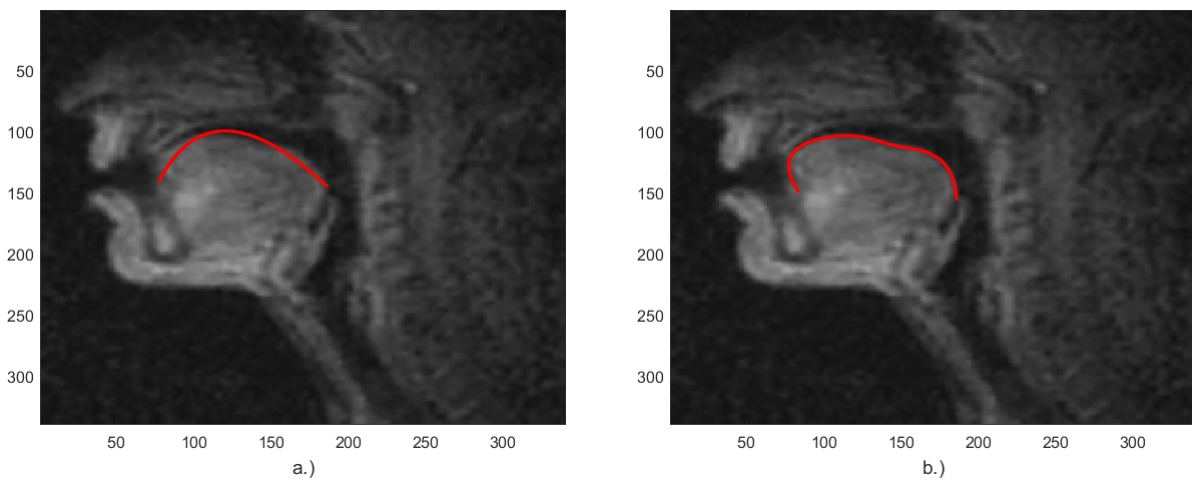
$$x_n = \sqrt{\frac{2}{N}} \sum_{k=1}^N X_k \frac{1}{\sqrt{1 + \delta_{k1}}} \cos\left(\frac{\pi}{2N}(2n-1)(k-1)\right) \quad (6.4)$$

alakban írható fel, mellyel az eredeti adattömb kapható vissza. A (6.3) és (6.4) összefüggésekben δ_{k1} a Kronecker-deltát jelöli a

$$\delta_{k1} = \begin{cases} 0, & k \neq 1, \\ 1, & k = 1 \end{cases} \quad (6.5)$$

definíciónak megfelelően. Esetünkben az $\{x_1, x_2, \dots, x_N\}$ paraméterek a nyelvkontúrok pontjainak vízszintes vagy függőleges koordinátáit testesíthetik meg, az $\{X_1, X_2, \dots, X_N\}$ értékek pedig az adott görbéhez rendelt DCT-együtthatókat adják meg. A (6.3) és (6.4) műveletekből kiindulva látható, hogy habár a DCT a tanítópontok meghatározásához hasonlóan egy diszkrét matematikai manipuláció, a transzformációban azonban a folytonosnak tekintett nyelvkontúrok (kellően sűrűn követik egymást a pontok, két szomszédos pont vízszintes vagy függőleges koordinátái közt többnyire egy pixelnél kisebb a különbség) összes pontja részt vesz. Következésképpen a DCT-együtthatók használatával globális

információt nyerhetünk ki a nyelvkontúrokból anélkül, hogy individuális pontokat specifikálnánk a görbe mentén. A DCT inverz műveletével együtt kiválóan alkalmas görbék simítására, aminek igen nagy jelentősége van az UH- és MRI-felvételekre illesztett nyelvkontúrok egyenetlenségeinek kiküszöbölésében is (3. fejezet). Ennélfogva a DCT-együtthatók fontos információkat foglalnak magukba a nyelvkontúr alaki tulajdonságainak vonatkozásában. A (6.4) inverz transzformációban szereplő DCT-együtthatók száma igen releváns faktor a görbék simításában, mivel meghatározza a nyelvkontúr alapvető formáját. A 6.2. ábra egy s hanghoz tartozó MRI-kereten megrajzolt nyelvkontúrokat vizualizál abban az esetben, amikor a görbe simítása az első 4, illetve 8 DCT-együtthatóval történt az a.), illetve b.) blokkokon. Az ábra tanúsága szerint nyilvánvaló, hogy 4 DCT-együttható nem elegendő ahhoz, hogy egy jól illeszkedő görbét produkáljunk, 8 DCT-együttható használatával azonban a kontúr megfelelően követi a nyelvfelszín alakját [66].



6.2. ábra. Az s hanghoz tartozó MRI-nyelvkontúrok az inverz transzformáció első 4 (a.), illetve 8 (b.) DCT-együtthatójával történő simítást követően. (Az ábra tengelyeinek mértékegysége pixel.)

6.3. A gépi tanítás módszere

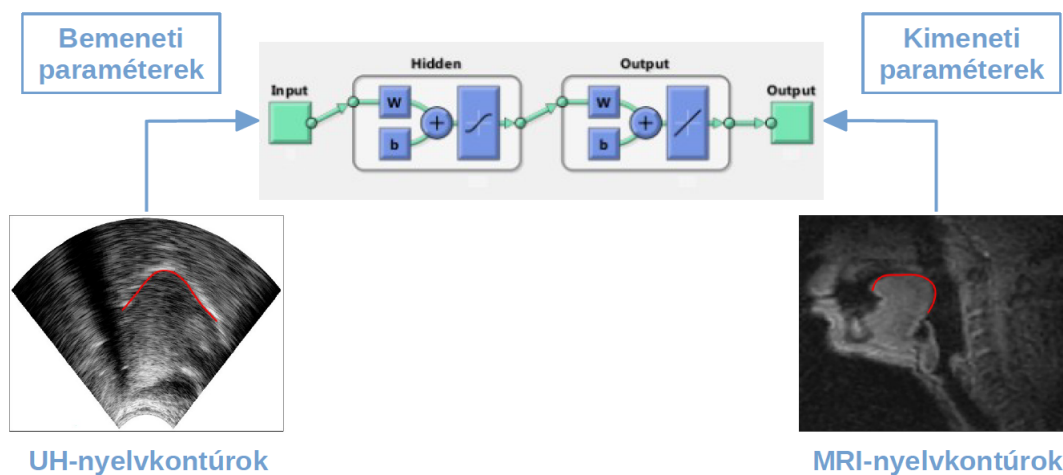
A gépi tanítást szintén a MATLAB szoftver segítségével valósítottam meg, és az előrecsatolt neurális hálózat súlyfaktorait a skálázott konjugált gradiens (Scaled Conjugate Gradient – SCG) módszer alkalmazásával határoztam meg, ami egy felügyelt tanulóalgoritmus [67]. Ezen optimalizációs eljárás a problémához rendelt egyenletrendszert a bemeneti paraméterek ismeretében iterációs módszerrel oldja meg, miközben az eljárással számított kimeneti paraméterek értékei konvergálnak az előírt értékekhez. A módszer előnye, hogy az iterációs algoritmus lépésközeinek számát minimalizálva elég gyors konvergencia biztosítható, így a gépi tanítás viszonylag rövid idő alatt véghezvihető. Az iterációs lépések olyan irány mentén valósulnak meg, ami gyorsabb konvergenciát biztosít, mint a legmeredekebb ereszkedésnek megfelelő legnegatívabb gradiens, miközben megőrzi a korábbi lépésekben kapott hibaminimalizációt. A teljes adathalmaz 70 %-a vesz részt a tanításban, a megmaradó 15-15 % pedig a validálásra és a tesztelésre van szánva. A tanítás akkor fejeződik be, ha elérjük a tanítási körök maximális számát (1000), vagy meghaladjuk a maximális időmennyiséget (végtelen), vagy az átlagos négyzetes hiba által definiált teljesítmény a

célértékre (0) minimalizálódik, vagy a teljesítménygradiens a minimum (10^{-6}) alá esik, vagy a validációs teljesítmény eléri a validációs hibák maximumát (6) a legutolsó változást követően. A zárójelekben az adott mutató alapértelmezett értékei vannak feltüntetve.

A tanulóalgoritmusok megszerkesztése során minden esetben az SCG módszert választottam, ugyanis a vizsgálataimnak nem volt célja a tanítási módszer optimalizálása, ezért másfajta eljárásokkal nem próbálkoztam.

6.4. A neurális hálózatok konstrukciója

A neurális hálózatok bemenetét az UH-nyelvkontúrokból kinyert adatokkal gerjesztettem, a kimeneten pedig az MRI-nyelvkontúrokból származtatott paramétereket állítottam be, tehát a rendszer végső soron a különböző beszédhangokhoz tartozó MRI-nyelvkontúrok alakjait tanulja be a tanítómintázatként alkalmazott UH-nyelvkontúrok alapján. Ezt a tanítási koncepciót az összes kialakított rendszerkonfiguráció esetében fenntartottam, melynek oka az UH- és MRI-felvételek 2. fejezetben kifejtett, eltérő vizuális tulajdonságaira vezethető vissza. Ennek megfelelően a nyelvcsont és az állcsont árnyékoló hatása miatt az UH-kereteken kevesebb információ érhető el a nyelv alakjáról az MRI-keretekhez viszonyítva, így a rendszernek egy bővebb adathalmazt kell generálnia a kimeneten egy szűkebb bemeneti paramétercsoportból kiindulva. Ez természetesen több kihívást tartogathat és talán érdekesebb lehet, mintha UH-nyelvkontúrokat tanítanánk MRI-nyelvkontúrok segítségével. A neurális hálózatok szerkezeti sémáját a 6.3. ábra demonstrálja, ahol a grafikus szimbólumok arra is utalnak, hogy a rejtett réteg hiperbolikus tangens szigmoid, a kimeneti réteg pedig lineáris aktivációs függvényrel működik.



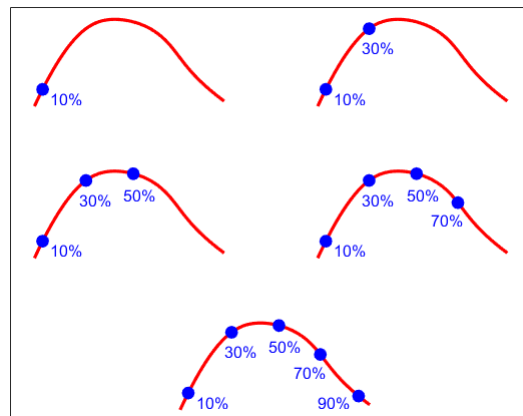
6.3. ábra. A megszerkesztett neurális hálózatok alapvető struktúrája.

Variálva a neurális hálózat bemeneti és kimeneti paramétereinek típusát és számát, illetve a rejtett rétegek és neuronok számát, különböző rendszerkonfigurációkat konstruáltam, és a kapott eredményeket kvalitatív és kvantitatív módon is analizáltam, kiválasztva a legjobb eredményt hozó rendszerbeállítást. Hangsúlyozom, hogy a tanítási feladat elvégzése során nem volt kitűzött törekvés a neurális hálózat belső paramétereinek optimalizációja, mindössze felvettem néhány lehetséges rendszerstruktúrát, és elemeztem a gépi tanítás hatékonyságát és jóságát. A tanulmányozott rendszerkonfigurációk által produkált eredményeket a következő alfejezetek foglalják magukba.

6.4.1. Tanítás 1

A neurális hálózatban egy rejtett réteget helyeztem el, ami csupán 10 neuront tartalmaz, mivel első közelítésben azt szerettem volna tesztelni, hogy egy nagyon egyszerű hálózat milyen teljesítménnyel képes egy viszonylag bonyolult tanulási feladat végrehajtására. A rendszer tanítását kétféle verzióban valósítottam meg.

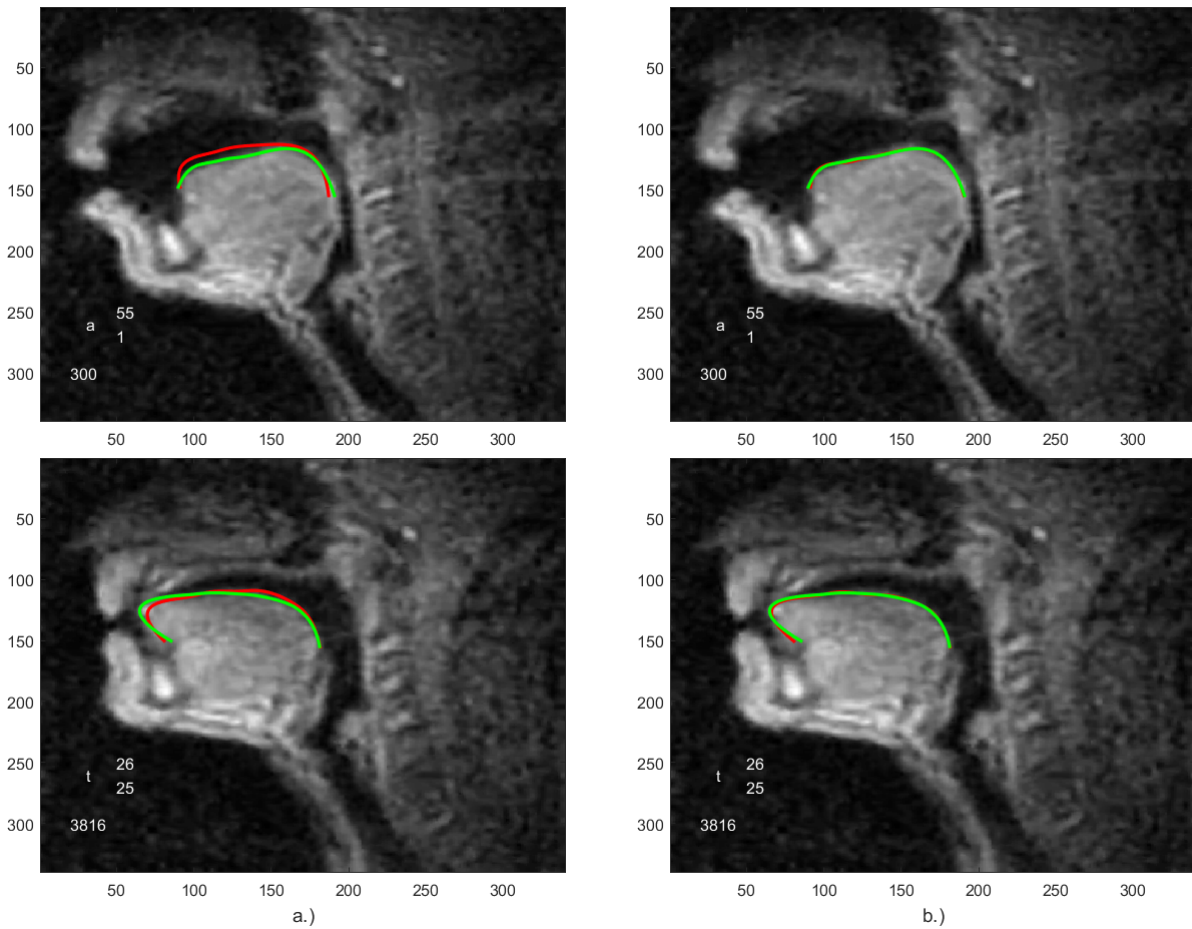
6.4.1.1. V_1 verzió ♦ A hálózat bemenetét az UH-nyelvkontúrokból eredeztetett tanítópontokkal stimuláltam úgy, hogy a pontok számát 1-től 5-ig változtattam. Az öt tanítópontot a görbék teljes szögtartományának 10, 30, 50, 70 és 90 %-ánál rögzítettem, és öt különböző bemeneti kompozíciót hoztam létre a 6.4. ábra által vázolt módon. A rendszert 1, 2, 3, 4 vagy 5 kontúrponthallal tanítottam azzal a stratégiával, hogy a 10 %-os pozíciónál lévő pontból indulva mindig hozzáadtam a soron következő pontot az adott beállításban figyelembe vett tanítópontok halmazához. Ennek nyomán alakulnak ki a 6.4. ábra egymást követő blokkjai, ahol a nyelv hátsó része a görbe bal oldalánál, a nyelvhegy felőli része pedig a görbe jobb oldalánál található. A tanításhoz a tanítópontoknak csak a (6.2) szerint bevezetett függőleges koordinátáit használtam fel. A hálózat kimenetén az MRI-nyelvkontúrok első 10 DCT-együtthatóját írtam elő. Ebben a verzióban tehát csak a bemeneti paraméterek számát módosítottam, miközben a kimeneti paraméterek számát fixen tartottam. Következésképpen a bemeneti adathalmaz dimenziója $4079 \times m$, ahol m a tanítópontok számát jelöli $m = 1, 2, 3, 4, 5$ szerint, a kimeneti adattömb pedig 4079×10 méretű a 10 DCT-együtthatóval összhangban.



6.4. ábra. A tanítópontok sematikus elrendezése a neurális hálózat bemenetének öt különböző beállításában.

Az eredmények kvalitatív értékeléséhez vizuálisan tanulmányoztam a betanított és az eredetileg illesztett MRI-nyelvkontúrok egymáshoz viszonyított helyzetét és alakját az $\{a, e, o, k, s, t\}$ készlet összes hangjának esetében. Azt találtam, hogy a két MRI-görbe közötti egyezés mértéke javul, ha növeljük az UH-nyelvkontúrok tanítópontjainak számát. A megállapításaimat alátámasztó példákat prezentál a 6.5. ábra az a és t hangok esetében, ahol a betanított görbék pirossal, az illesztett nyelvkontúrok pedig zölddel vannak megjelölve. A 6.5.a ábra az 1 tanítóponttal, a 6.5.b ábra pedig az 5 tanítóponttal végzett gépi tanítás eredményeit szemlélteti. Látható, hogy 1 tanítópont alkalmazása esetén

jól kivehető különbség mutatkozik, 5 tanítópont esetén azonban szinte tökéletes egyezés adódik a két MRI-görbe között.



6.5. ábra. 1 (a.) és 5 (b.) tanítóponttal végzett tanítás eredményeként kapott görbék (piros) az eredetileg illesztett MRI-nyelvkontúrhoz (zöld) viszonyítva az *a* és *t* hangok esetében. (Az ábra tengelyeinek mértékegysége pixel.)

Az eredmények kvantitatív jellemzéséhez az (5.8) NND mértéket használtam fel. (5.8) alkalmazásával kiszámítottam a betanított és illesztett MRI-nyelvkontúrok közötti NND távolságokat az $\{a, e, o, k, s, t\}$ halmaz összes hangjának összes nyelvkontúrpárjára, majd megadtam a kapott NND távolságok átlagát mind az öt rendszerkonfiguráció esetében.

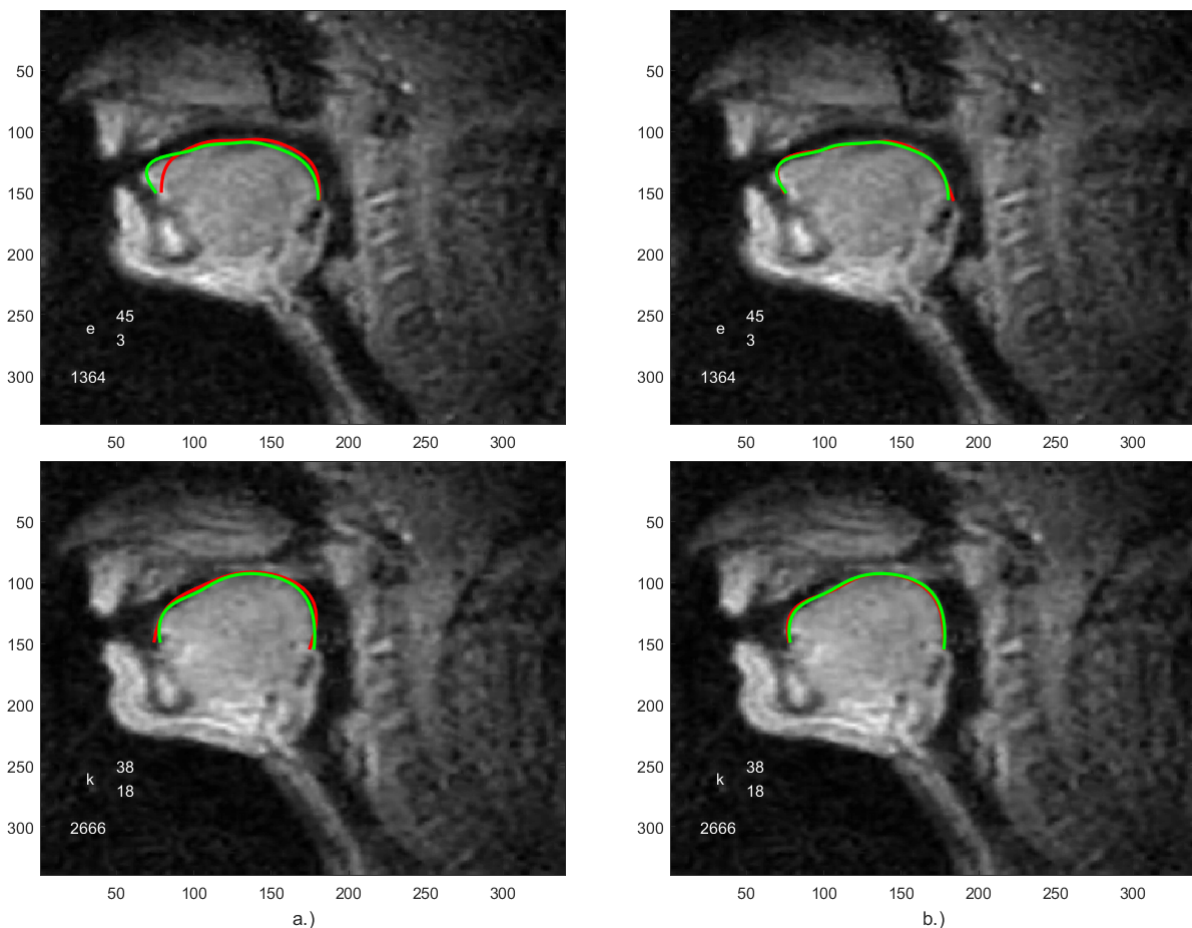
6.1. táblázat. A V_1 verzió alkalmazásával kapott eredmények.

	1 pont	2 pont	3 pont	4 pont	5 pont
NND (px)	5,66	5,44	4,87	4,02	3,76

Eredményeimet a 6.1. táblázat rendszerezi, és ez alapján kijelenthető, hogy a kapott numerikus adatok megerősítik a kvalitatív úton nyert tapasztalataimat, hiszen a tanítópontok

számának növelésével csökken az NND távolság, és 5 tanítópont esetén áll elő a legkisebb érték. Így a végső konklúzió az, hogy az 5 UH-kontúrponttal megvalósított tanítás produkálja a legjobb egyezést a betanított és illesztett MRI-nyelvkontúrok között.

6.4.1.2. V_2 verzió ♦ A hálózat bemenetén az UH-nyelvkontúrokból származó 5 tanítópontot szabtam ki, melyek helyzetét a 6.4. ábra utolsó rajza szerint választottam meg a görbe teljes szögtartományának 10, 30, 50, 70 és 90 %-ánál. A tanítást ezúttal is a tanítópontok függőleges koordinátaival végeztem el. A hálózat kimeneti adathalmazát az MRI-nyelvkontúrok első 5, 6, 7, 8, 9 vagy 10 DCT-együtthatójának formájában nyertem. Ebben a verzióban tehát a bemeneti paraméterek számát rögzítettem, és a kimeneti paraméterek számát variáltam. Ebből adódóan a bemeneten egy 4079×5 méretű adattömb táplálja az 5 tanítópontnak megfelelően, a kimeneten pedig egy $4079 \times n$ dimenziójú adatblokk jelenik meg, ahol n a DCT-együtthatók számát jelöli $n = 5, 6, 7, 8, 9, 10$ szerint.



6.6. ábra. 5 (a.) és 10 (b.) DCT-együtthatóval végzett tanítás eredményeként kapott görbék (piros) az eredetileg illesztett MRI-nyelvkontúrhoz (zöld) viszonyítva az e és k hangok esetében. (Az ábra tengelyeinek mértékegysége pixel.)

Az eredmények kvalitatív analízisét ez esetben is a betanított és illesztett MRI-nyelvkontúrok vizuális összehasonlításával hajtottam végre az $\{a, e, o, k, s, t\}$ szett összes hangjára vonatkozóan, és arra jutottam, hogy a DCT-együtthatók számának növelésével fokozódik az átfedés mértéke a két MRI-görbe között. Erről tanúskodik a 6.6. ábra, amely

az e és k hangok pirossal, illetve zölddel megrajzolt betanított, illetve illesztett nyelvkontúrjait ábrázolja. A 6.6.a ábra az 5 DCT-együtthatóval, a 6.6.b ábra pedig a 10 DCT-együtthatóval megvalósított gépi tanítás eredményeit jeleníti meg. Megfigyelhető, hogy 10 DCT-együttható alkalmazása sokkal előnyösebb az 5 DCT-együtthatóval szemben, hiszen kis híján kifogástalan egybevághóságot produkál a két MRI-görbe.

Az eredmények kvantitatív minősítéséhez újfent az (5.8) NND mértékre támaszkodtam, és kiszámítottam a betanított és illesztett MRI-nyelvkontúrok közötti NND távolságokat az $\{a, e, o, k, s, t\}$ halmaz összes hangjának összes nyelvkontúrpárjára, majd megadtam a kapott NND távolságok átlagát mind a hat rendszerkonfiguráció esetében. A kimeneteket a 6.2. táblázat tömöríti, ami arra mutat, hogy az eredmények kvalitatív megítélése igazolást nyer, hiszen a DCT-együtthatók számának növelésével csökken az NND távolság. Az is megállapítható, hogy a 9 és 10 DCT-együtthatóval végzett tanítás során kapott NND távolságok között árnyalatnyi különbség áll elő, ami végül is a 9 DCT-együttható javára billenti a minimalizációt. Ezen numerikus finomítás arra vezet, hogy az UH-nyelvkontúrok első 9 DCT-együtthatójával kivitelezett tanítás generálja a legbiztosabb egyezést a betanított és illesztett MRI-nyelvkontúrok között.

6.2. táblázat. A V_2 verzió alkalmazásával kapott eredmények.

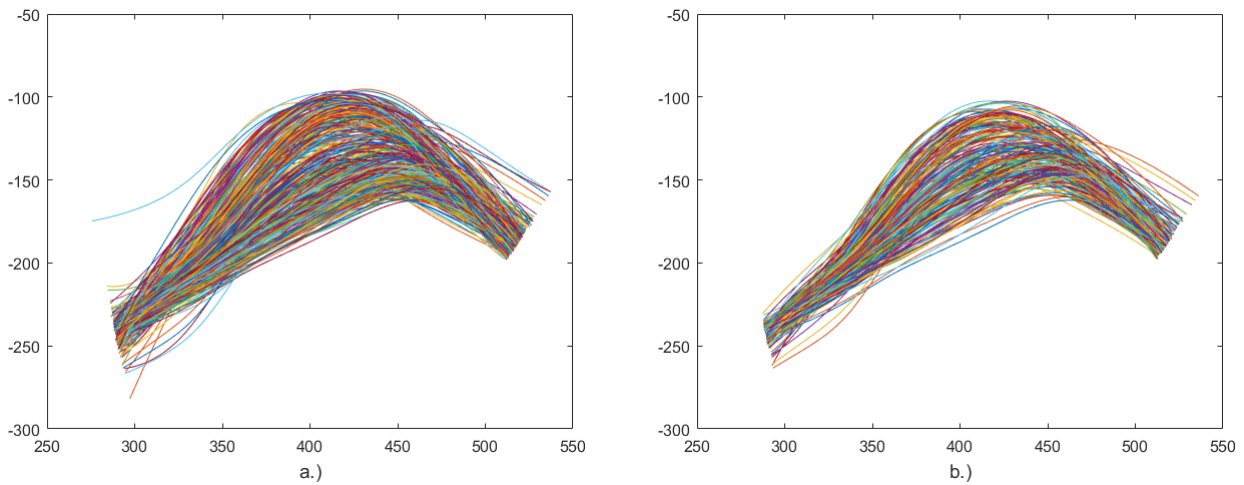
	5 DCT	6 DCT	7 DCT	8 DCT	9 DCT	10 DCT
NND (px)	4,17	4,15	4,03	4,07	3,72	3,76

A 6.4.1. alfejezetben ismertetett tanulóalgoritmusok bevetése során a 4.1. alfejezetben említett természetes hibákkal terhelt UH-nyelvkontúrokat is tartalmazó görbesereget alkalmaztam, tehát a tanításban gyengébb minőségű, pontatlan nyelvkontúrok is részt vettek. Így az eredmények finomítása végett további feladatként tűztem ki a természetes hibákból származó fals nyelvkontúrok szigorú kiszűrését, aminek az volt a célja, hogy kiiktassam a tanítóalakzatok közül az eleve tökéletlen nyelvkontúrokat. A görbék vizsgálatát és válogatását egy ötfokozatú szűrés keretében vittem véghez. A szelekció öt lépését a következő alfejezetben tárgyalom.

6.4.2. Az UH-nyelvkontúrok szelekciója

6.4.2.1. A hangátmenetekhez tartozó UH-nyelvkontúrok kiszűrése ♦ Miután az UH-nyelvkontúrokból származnak a hálózat tanításához szükséges bemeneti paraméterek, rendkívül nagy hangsúlyt kap a kontúrkövető algoritmus által produkált nyelvkontúrok jósága, pontossága és hitelessége. A felvételek alapvető jellemzője, hogy a képkockák az adott bemondásban szereplő hangok szerint vannak szegmentálva úgy, hogy egy hanghoz több képkocka rendelhető hozzá. A hangátmenetek azonban nincsenek megjelölve a szegmentált adathalmazban, ennél fogva az adott hanghoz tartozó keretek magukba foglalják a nyelvnek a koartikuláció közben kialakuló átmeneti pozícióit is. Abból a célból, hogy redukáljam a hangátmenetek nyelvállásra gyakorolt hatásait, és a lehető legjobb mértékben csak az adott hangot jellemző, ún. kulcs nyelvkontúrookra szorítkozhassak, azt a taktikát követtem, hogy minden hang esetében negligáltam 2-2 képkockát a tömb elején és végén. Ily módon jó közelítéssel kiszűrtem az átmeneti keretek nagy részét. A 6.7. ábra az

átmeneti keretek kiszűrése előtti és utáni nyelvkontúr-sokaság közti vizuális különbséget érzékelteti az *e* hang esetében.

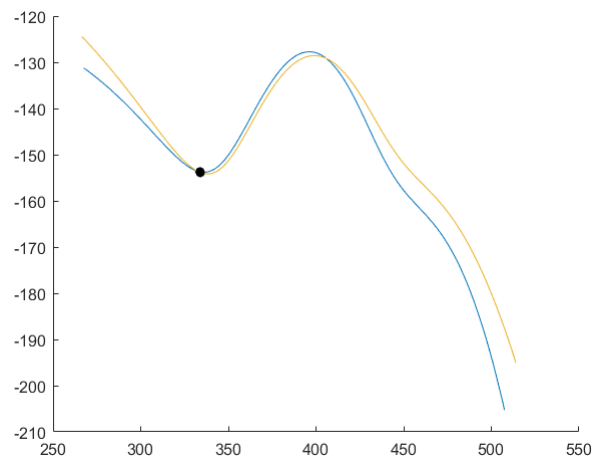


6.7. ábra. Az átmeneti keretek kiszűrése előtti (a.) és utáni (b.) UH-nyelvkontúrok görbeserege az *e* hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)

6.4.2.2. A negatív iránytangensű UH-nyelvkontúrok kiszűrése ♦ A nyelvkontúr-követés sok esetben olyan görbét generál, melyek a nyelv hátsó részénél irreális módon visszahajlanak, aminek következtében a görbék ezen szakasza negatív értékű meredekséggel jellemezhető. Egy görbe meredeksége matematikailag az iránytangenssel írható le, ami folytonos értelmezési tartomány és értékészlet esetén a görbe elsőrendű differenciáhányadosa, azaz deriváltja. Jelen esetben azonban – bár a nyelvkontúrokat folytonos görbék reprezentálják – a nyelvkontúrok vízszintes és függőleges értékhalmaza diszkrét pontok sokaságából tevődik össze, ezért a derivált helyett a görbék elsőrendű differenciáhányadosai értelmezhetők, és ezek alkotják a vizsgálandó mennyiségeket. Ennek megfelelően minden beszédhang esetében kiszámítottam az adott hanghoz tartozó összes nyelvkontúr összes szomszédos pontpárjára vonatkozó elsőrendű differenciáhányadosokat a

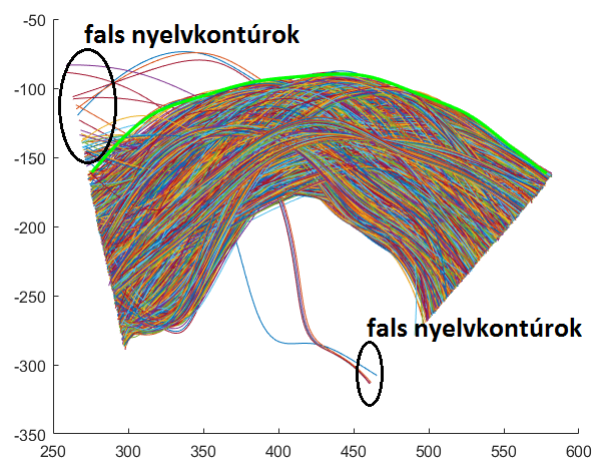
$$D_{k,k+1} = \frac{y_{k+1} - y_k}{x_{k+1} - x_k} \quad (6.6)$$

összefüggés felállításával, melyben a hányadosképzés végigfut a nyelvkontúrok teljes hosszán, amit 177 pont alkot, így $k = 1, 2, \dots, 176$. Mivel a görbék kritikus visszahajlása a nyelvkontúrok hátsó tartományában észlelhető, a differenciáhányadosokat a nyelvkontúrok első 50 pontjának kiválasztásával vettem elemzés alá, és az ezen tartományon negatív értékű iránytangenssel rendelkező görbéket kizártam. Negatív iránytangensű nyelvkontúrokat mutat be az *o* hang esetében a 6.8. ábra, ahol a görbék balról számított 50. pontja fekete ponttal van megjelölve.



6.8. ábra. Negatív iránytangensű fals UH-nyelvkontúrok az *o* hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)

6.4.2.3. A szájpadkontúron túlcsorduló UH-nyelvkontúrok kiszűrése ♦ Mint ahogyan azt a 2. fejezetben is taglaltam, az UH-felvételeken közvetlen módon nem detektálható a szájpad kontúrvonala, viszont az artikuláció tanulmányozása során sok esetben nélkülözhetetlen támpontként kell, hogy szolgáljon. Kiváltképp igaz ez akkor, amikor a nyelv alakja és különböző pozíciói kulcsfontosságú szerepet játszanak a beszéd analízisében, hiszen a nyelvállást alapvetően a szájpadkontúrhoz viszonyítva adhatjuk meg. Ezért a 4. fejezetben megszerkesztett UH-szájpadkontúr felhasználásával végeztem el a görbék válogatását. A 6.9. ábrán megjelenő szájpadkontúr zöld vonallal van rávetítve a 40 UH-bemondás összes nyelvkontúrjának halmazára, és feketével bekarikázva tűnnek fel a falsnak tekintett nyelvkontúrok. Az ábra felső részén láthatók azok a görbék, melyek túllépnek a szájpadkontúr határvonalán.



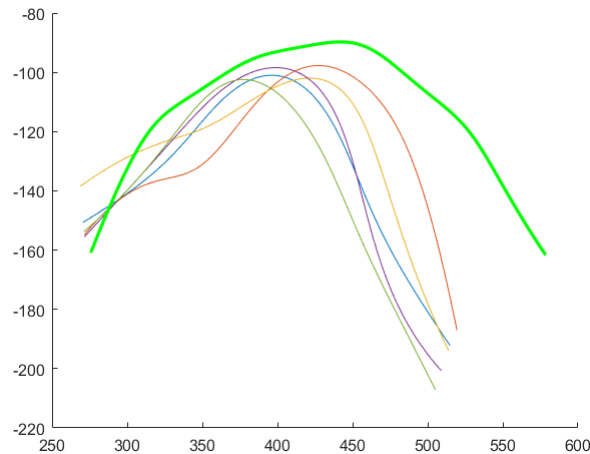
6.9. ábra. A szájpadkontúr zöld görbéje a 40 UH-bemondás összes nyelvkontúrjára rávetítve a fals nyelvkontúrok megjelölésével. (Az ábra tengelyeinek mértékegysége pixel.)

A negatív iránytangensű nyelvkontúrok kiszűrése után azokat a görbéket távolítottam el, melyek az előbb említett szájpadkontúron túlhaladnak. Ezt egy egyszerű feltétel kiszűrésével értem el, miszerint ha az adott G görbe balról számított első pontjának függőleges

koordinátája a 6.9. ábra szerinti elrendezésben magasabb pozícióban található, mint az SZ szájpaddockontúr első pontjának függőleges koordinátája, azaz ha

$$|G(1; y)| > |SZ(1; y)|, \quad (6.7)$$

akkor a görbétől eltekintek. Miután a 6.9. ábra függőleges tengelye negatív skálázású, a (6.7) reláció a koordináták abszolút értékeire teljesül. A szájpaddockontúron túlsorduló nyelvkontúrokat példáz a 6.10. ábra a k hang esetében.

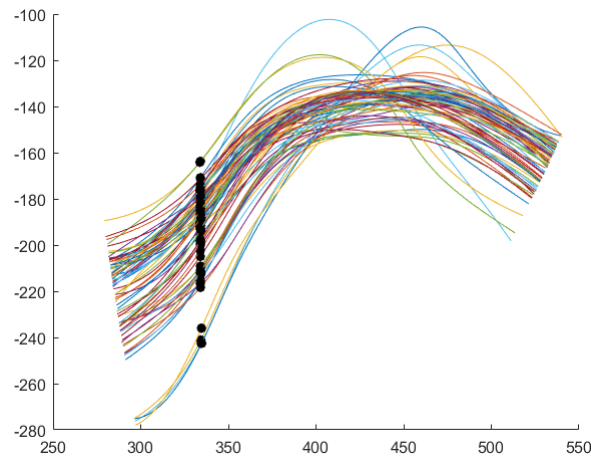


6.10. ábra. A szájpaddockontúron (zöld) túlsorduló fals UH-nyelvkontúrok a k hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)

6.4.2.4. A konvex UH-nyelvkontúrok kiszűrése ♦ A 6.8. ábra által is vázolt, igen meredek visszahajlást mutató negatív iránytangensű nyelvkontúrok mellett olyan görbék is keletkezhetnek a kontúrkövetés során, melyek kisebb mértékben ugyan, de még mindig deformáltnak tekinthetők, mert konvex szakaszokat tartalmaznak a nyelv hátsó részénél. Ezeken a konvex szakaszokon kis értékű pozitív iránytangenseket mérhetünk, így az ilyen típusú görbék selejtezése a

$$\min(D_{k,k+1}) < M_1 \quad (6.8)$$

feltétel kijelölésével hajtható végre, ami az adott hanghoz tartozó összes nyelvkontúr differenciahányadosainak minimumai közül kiszűri az M_1 konstans által felülről korlátozott értékeket. A differenciahányadosokat ez esetben a nyelvkontúrok első 40-50 pontjára vizsgáltam meg. A feltételt mindenképpen a legkisebb értékű iránytangensekre kell megfogalmaznunk, mivel a differenciahányadosok minimumainak környékén a leglaposabb a görbe, és éppen ezt a kedvezőtlen tulajdonságot kell kiküszöbölnünk a nyelvkontúrok válogatásakor. Az M_1 konstans értékét minden hang esetén tapasztalati úton állítottam be, ami jellemzően 0,5-0,6. Konvex szakaszokat tartalmazó nyelvkontúrokat vizualizál a t hang esetében a 6.11. ábra, ahol a görbék balról számított 40. pixeljét fekete pontok jelzik.



6.11. ábra. Konvex szakaszokat tartalmazó fals UH-nyelvkontúrok a *t* hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)

6.4.2.5. A kirívó UH-nyelvkontúrok kiszűrése ♦ A nyelvkontúrok halmazában néhány esetben felbukkannak olyan görbék, melyek szemmel láthatóan elütnek attól a görbeseregtől, amit az adott hang nyelvkontúrjainak nyalábja hoz létre. A nyelvkontúrok változatossága az UH-felvételeken abból adódik, hogy az egyes bemondások során a beszélő a hangokat különböző hangkörnyezetekben szólaltatja meg, ami a koartikuláció miatt háttással van az adott hang nyelvállására. Ez az effektus még akkor is megfigyelhető, miután a hangátmenetekhez tartozó keretek nagy részét jó közelítéssel elhanyagoljuk. A kirívó nyelvkontúrok kiszűrésekor a görbesereg A átlagát jelöltem ki viszonyítási alapként, és azokat a G görbéket vetettem el, melyek átlagtól mért távolságának legnagyobb abszolút értéke nagyobb egy meghatározott küszöbnél a

$$\max(|G - A|) > M_2 \quad (6.9)$$

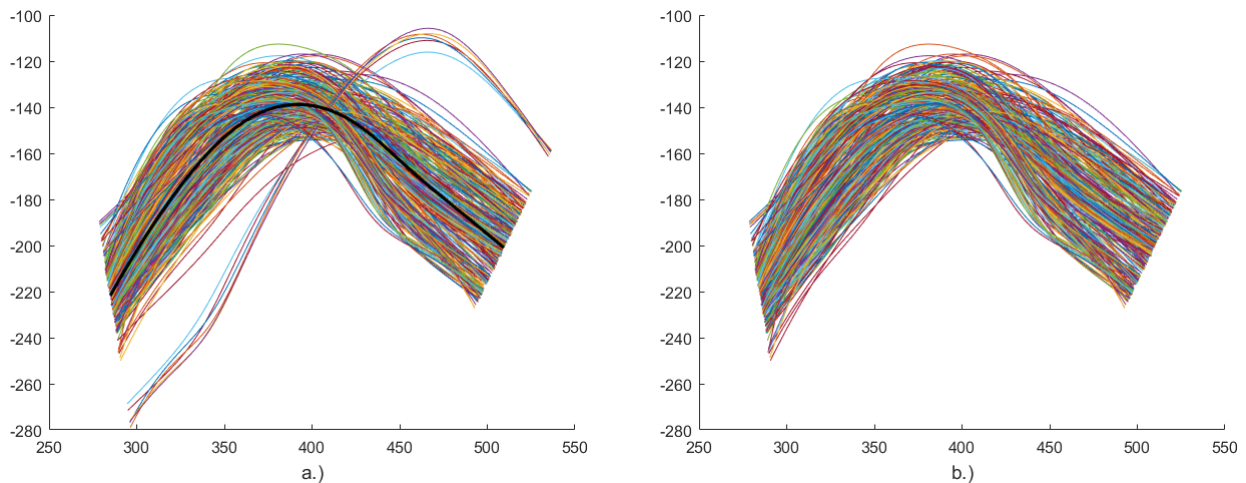
reláció szerint, azaz a görbének az átlaggörbétől legtávolabb eső pontja már kívül esik az M_2 korlát által rögzített tartományon. Az M_2 konstans értékét minden hang esetén tapasztalati úton állítottam be, ami jellemzően a $[30, 70]$ intervallumba esik. A nyelvkontúrok sokaságától elütő görbéket szemléltet az *a* hang esetében a 6.12. ábra. A 6.12.a ábra az átmeneti, a negatív iránytangensű, a szájpadkontúron túlsorduló, illetve a konvex nyelvkontúrokat már nem tartalmazó görbesereget demonstrálja a fekete vonallal megrajzolt átlag feltüntetésével, a 6.12.b ábrán pedig már a kirívó nyelvkontúroktól is mentes végső sokaság látható.

6.3. táblázat. Az összes és a redukált UH-nyelvkontúrok száma hangonkénti bontásban.

	a	e	o	k	s	t	Összesítve
Összes nyelvkontúrok száma	1 084	386	853	489	205	1 062	4 079
Redukált nyelvkontúrok száma	328	144	489	155	37	433	1 586

A fentebb vázolt öt szelekciós lépés nyomán egy szigorúan redukált nyelvkontúrhalmoz

állt elő, melynek hangonkénti elemszámáról a 6.3. táblázat ad tájékoztatást.



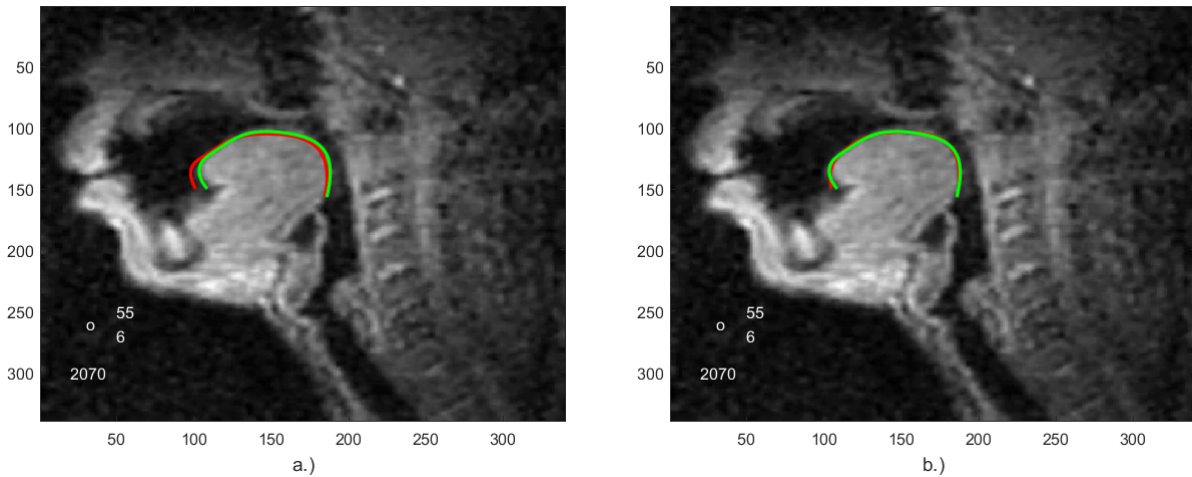
6.12. ábra. Az *a* hanghoz tartozó görbeseregek a kirívó nyelvkontúrokat tartalmazó (a.) és a kirívó nyelvkontúroktól mentes (b.) változatban. (Az ábra tengelyeinek mértékegysége pixel.)

Mindezekon túlmenően szem előtt kell tartanunk azt is, hogy a neurális hálózat tanítása során az MRI-nyelvkontúrok ugyanolyan lényeges szerepet töltenek be, mint az UH-nyelvkontúrok, hiszen ezeknek a görbéknek az alakja is érdemben befolyásolja a gépi tanulás eredményét. Ennek folytán ez esetben is rendkívül fontos a kontúrkövető algoritmus paramétereinek gondos beállítása, ami biztosítja az ideálisan illeszkedő, megbízható nyelvkontúrok előállítását. Az MRI-felvételek esetében azonban jóval kevesebb a potenciális hibaforrás, mivel – az UH-felvételekkel ellentétben – a képek kontrasztja kielégítő minőségű a pontos kontúrok megállapításához, és árnyékoló hatások sem zavarják meg a kontúrkövetés tartományát. Ebből következően nincs szükség olyan jellegű szelektálásra, mint amit az UH-nyelvkontúrok halmazában végeztem.

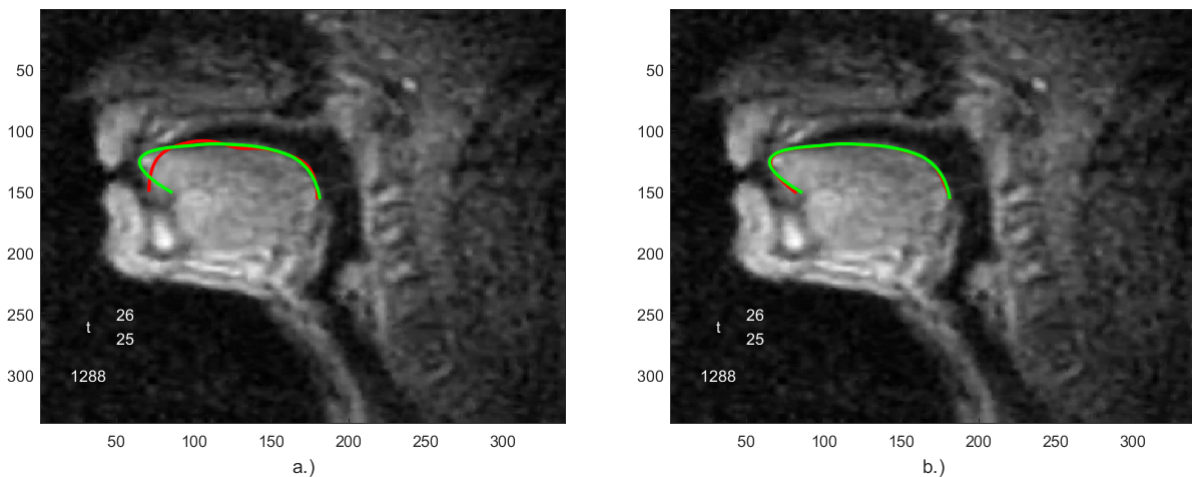
6.4.3. Tanítás 2

A neurális hálózat tanítását a 6.4.1.1. és 6.4.1.2. paragrafusokban kifejtett V_1 és V_2 verziók szerint megvalósítottam a redukált UH-nyelvkontúrok halmazával is úgy, hogy megtartottam az egyetlen rejtett réteget, melyben továbbra is 10 neuron foglalt helyet. Ekkor a 6.3. táblázattal összhangban a rendszer bemenete m tanítópont esetén $1\,586 \times m$ dimenziójú mátrixokkal van táplálva, a kimeneten pedig n DCT-együttható tekintetbe vételekor $1\,586 \times n$ méretű tömbök jelennek meg. Az eredmények kvalitatív és kvantitatív vizsgálatát ugyanolyan szempontok alapján végeztem el, mint ahogyan azt a 6.4.1. alfejezetben is vázoltam.

A redukált UH-nyelvkontúrok alkalmazásakor is kvalitatíve azt állapítottam meg, hogy a betanított és illesztett MRI-nyelvkontúrok alaki egyezősége javul a tanítópontok számának (V_1 verzió), illetve a DCT-együtthatók (V_2 verzió) számának növelésével. Észrevételeimet a 6.13. és 6.14. ábrák is bizonyítják, ahol a betanított görbék pirossal, az illesztett nyelvkontúrok pedig zölddel vannak berajzolva. A 6.13.a és 6.13.b ábrák a V_1 konfiguráció 1, illetve 5 tanítópontja esetében kapott eredmények láthatók az *o* hang nyelvkontúrája, a 6.14.a és 6.14.b ábrák pedig a V_2 konfiguráció 5, illetve 10 DCT-együtthatója esetében kapott eredményeket szemléltetik a *t* hang nyelvkontúrájának példáján keresztül.



6.13. ábra. A betanított (piros) és az illesztett (zöld) MRI-nyelvkontúrok 1 (a.), illetve 5 (b.) tanítópont alkalmazásával a redukált UH-nyelvkontúrok felhasználásával az *o* hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)



6.14. ábra. A betanított (piros) és az illesztett (zöld) MRI-nyelvkontúrok 5 (a.), illetve 10 (b.) DCT-együttható alkalmazásával a redukált UH-nyelvkontúrok felhasználásával a *t* hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)

A kvantitatív analízis kimeneteleit a 6.4. és 6.5. táblázatok összesítik. A 6.4. táblázatban a V_1 konfiguráció, míg a 6.5. táblázatban a V_2 konfiguráció eredményei követhetők nyomon. A számadatok alapján kijelenthető, hogy a kvantitatív eredmények a redukált UH-nyelvkontúrok esetében is megerősítik a kvalitatív úton nyert benyomásokat, azaz növelve a tanítópontok vagy a DCT-együtthatók számát csökkenő tendencia mutatkozik az átlagos NND távolságok értékében, és ezúttal is azt a következtetést vonhatjuk le, hogy az 5 tanítóponttal, illetve a 9 DCT-együtthatóval végzett tanítás a leghatékonyabb. Ezen túlmenően az is világosan látszik, hogy az átmeneti, a negatív iránytangensű, a szájpaddkontúron túlsorduló, a konvex, valamint a kirívó UH-nyelvkontúrok kizárása javulást idéz elő a betanított és illesztett MRI-nyelvkontúrok összeegyeztetésében, hiszen a 6.4. és 6.5. táblázatok minden oszlopában csökkenés tapasztalható a redukált nyelvkontúrok javára.

6.4. táblázat. A betanított és illesztett MRI-nyelvkontúrok között mért átlagos NND távolságok különböző számú tanítópontok alkalmazásával.

	1 pont	2 pont	3 pont	4 pont	5 pont
Összes nyelvkontúr	5,66	5,44	4,87	4,02	3,76
Redukált nyelvkontúrok	5,28	4,47	3,63	3,29	2,82

6.5. táblázat. A betanított és illesztett MRI-nyelvkontúrok között mért átlagos NND távolságok különböző számú DCT-együtthatók alkalmazásával.

	5 DCT	6 DCT	7 DCT	8 DCT	9 DCT	10 DCT
Összes nyelvkontúr	4,17	4,15	4,03	4,07	3,72	3,76
Redukált nyelvkontúrok	3,19	2,75	2,92	2,90	2,61	2,82

6.4.4. Tanítás 3

A 6.4.1. és 6.4.3. alfejezetekben használt egyszerű neurális hálózattal kivitelezett tanítást követően módosításokat eszközöltem a rendszer belső szerkezetére vonatkozóan, és változtattam a bemeneti és kimeneti paraméterek típusán és számán is. A rejtett rétegek és neuronok számának variálásával három különböző rendszerstruktúrát hoztam létre, mely szerint 1, 2 vagy 3 rejtett réteget értelmeztem rendre 10, 20-20 és 30-30-30 neuron beültetésével. A bemeneti és kimeneti adathalmazok szintjén négy különböző konfigurációt realizáltam a 6.6. táblázat kategóriáinak megfelelően. Ennek értelmében a hálózat bemenetén változó számú tanítópontokat vagy DCT-együtthatókat állítottam be az [1,5], illetve az [1,10] intervallumokból választott értékekkel, a kimeneti paramétereket pedig 5 tanítópont vagy 10 DCT-együttható formájában rögzítettem. Az 5 tanítópont helyzetét változatlanul a 6.4. ábra írja le. Az eddigiekhez hasonlóan a bemeneti adatokat az UH-nyelvkontúrokból, a kimeneti adatokat az MRI-nyelvkontúrokból eredeztettem. Ezzel a módszerrel annak tanulmányozását irányoztam elő, hogy a bemeneti paraméterek típusa és száma hogyan befolyásolja a fix kimeneti paraméterek betanításának sikerességét.

6.6. táblázat. A négy különböző rendszerkonfiguráció értelmezése a bemeneti és kimeneti paraméterek tükrében.

Konfiguráció	Bemenet (UH)	Kimenet (MRI)
C_1	1,2,3,4,5 tanítópont	10 DCT-együttható
C_2	1,2,3,...,10 DCT-együttható	10 DCT-együttható
C_3	1,2,3,4,5 tanítópont	5 tanítópont
C_4	1,2,3,...,10 DCT-együttható	5 tanítópont

A neurális hálózat fentebb részletezett strukturális átalakítása mellett a tanításban részt vevő beszédhangok halmazát is kibővítettem az $\{a, \acute{a}, e, \acute{e}, i, o, \ddot{o}, u, \ddot{u}, g, k, l, s, t\}$ hangcsoport figyelembevételével, melyeket az $\{\mathfrak{a}, \acute{a}, \varepsilon, \acute{e}, i, o, \emptyset, u, y, g, k, l, f, t\}$ IPA (International Phonetic Alphabet) szimbólumokkal láthatunk el. Ezen kiterjesztett hangkészlet elemeihez rendelt UH-nyelvkontúrok sokaságán a 6.4.2. alfejezet paragrafusai által előírt szisztéma szerint ezúttal is válogatást végeztem, és a szelekció eredményeképpen kapott hangonkénti görbék számát a 6.7. táblázat listázza.

6.7. táblázat. A redukált UH-nyelvkontúrok száma hangonkénti bontásban.

Hang	a	á	e	é	i	o	ö	u	ü	g	k	l	s	t	Összesítve
Nyelvkontúrok száma	395	29	186	27	35	535	73	59	43	15	189	195	52	520	2 353

A 6.6. és 6.7. táblázatok alapján érthető, hogy a rendszer bemeneti adatblokkjának dimenziója $2353 \times m$, ahol $m = 1,2,3,4,5$ vagy $m = 1,2,3, \dots, 10$, a kimeneti oldal adattömbjének mérete pedig 2353×5 vagy 2353×10 . Az eredmények kvalitatív és kvantitatív analízise ez esetben is a 6.4.1. alfejezet aspektusai szerint történt.

A kvalitatív elemzéseket a 6.6. táblázat összes konfigurációjában végrehajtottam, ami a következő konklúziókkal zárult:

1. A betanított és illesztett MRI-nyelvkontúrok közötti vizuális egyezés mértéke javul, ha a bemeneten növeljük a tanítópontok számát függetlenül a kimeneti paraméterek típusától.
2. A betanított és illesztett MRI-nyelvkontúrok közötti vizuális egyezés mértéke javul, ha a bemeneten növeljük a DCT-együtthatók számát függetlenül a kimeneti paraméterek típusától.
3. A betanított és illesztett MRI-nyelvkontúrok közötti vizuális egyezés mértéke javul, ha a kimeneten tanítópontok helyett DCT-együtthatókat alkalmazunk függetlenül a bemeneti paraméterek típusától.

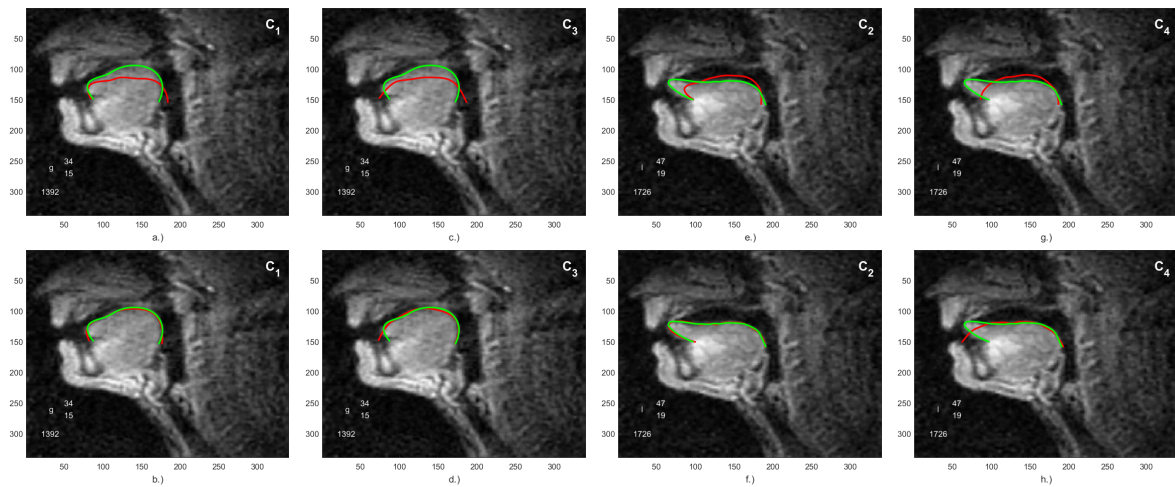
A kvantitatív kalkulációkat a 6.7. táblázat minden egyes beszédhangjára elvégeztem a 6.6. táblázat összes konfigurációjának esetében. A számításaim arra vezettek, hogy a C_1 és C_3 konfigurációk 5 tanítópont esetén produkálják a legkisebb NND távolságot, amikor a hálózatban 20-20 neuronnal felépített 2 rejtett réteg foglal helyet. Ezenkívül a C_2 és C_4 konfigurációkban 9 DCT-együttható esetén alakul ki a legkisebb NND mérték 30-30-30 neuront hordozó 3 rejtett réteg alkalmazásakor. Ezeket az eredményeket a 6.8. táblázat csoportosítja. A legjobb tanítóadatokhoz tartozó átlagos NND távolságok numerikus értékei (a táblázat által nem részletezett módon) alátámasztják a fentebb deklarált 1. és 2. konklúziót. Vegyük észre azt is, hogy a C_1 és C_3 és hasonlóan a C_2 és C_4 konfigurációk NND értékei megerősítik a fentebbi lista 3. állítását is, ami azzal magyarázható, hogy a tanítópontok diszkrét természetüknél fogva nem képesek előállítani a teljes nyelvkontúrt. A görbe egésze a tanítópontokból csak megfelelő interpolációs technikával alkotható meg, ezért kiegészítő lépésként az ún. módosított Akima-módszert (MAKIMA – Modified Akima) alkalmaztam a nyelvkontúrok hiányzó pontjainak rekonstruálására [68]. Interpolációval azonban soha nem hozhatunk létre olyan pontos nyelvkontúrokat, mint a

DCT-együtthatók segítségével, melyek globálisan jellemzik a görbéket. Ehhez kapcsolódóan érdemes felfigyelni arra is, hogy a 6.8. táblázat C_2 konfigurációjának NND távolsága adja a legkisebb értéket, ami rávilágít arra, hogy a felállított körülmények között az a tanítási módszer hozza a legkiválóbb eredményeket, melynek során DCT-együtthatókat tanul be a rendszer DCT-együtthatókból kiindulva.

6.8. táblázat. A legjobb tanítási eredmények a négy különböző rendszerkonfiguráció esetében.

Konfiguráció	A hálózat struktúrája	A legjobb tanítóadat	NND (px)
C_1	2 réteg + 20 neuron	5 tanítópont	3,48
C_2	3 réteg + 30 neuron	9 DCT-együttható	2,72
C_3	2 réteg + 20 neuron	5 tanítópont	4,10
C_4	3 réteg + 30 neuron	9 DCT-együttható	3,49

A vizuális és numerikus eredmények tükrében néhány példát illusztrál a 6.15. ábra. A 6.15.a, 6.15.b, 6.15.c és 6.15.d blokkok a C_1 és C_3 konfigurációk kimeneteleit hasonlítják össze a g hang esetében, míg a 6.15.e, 6.15.f, 6.15.g és 6.15.h blokkok a C_2 és C_4 konfigurációk hatásai között vannak párhuzamot az l hang példáján keresztül.



6.15. ábra. A betanított (piros) és az illesztett (zöld) MRI-nyelvkontúrok 1 (a. és c.), valamint 5 (b. és d.) tanítópont, illetve 5 (e. és g.), valamint 9 (f. és h.) DCT-együttható alkalmazásával a g és l hangok esetében. (Az ábra tengelyeinek mértékegysége pixel.)

A betanított görbék piros, az illesztett nyelvkontúrok pedig zöld vonallal vannak megjelölve. A 6.15.a-d ábrák az 1 (felső képek) és 5 (alsó képek) tanítóponttal generált eredményeket demonstrálják, amikor is 20-20 neuront tartalmazó 2 rejtett réteg található a hálózatban. A 6.15.e-h ábrák az 1 (felső képek) és 9 (alsó képek) DCT-együtthatóval produkált eredményeket mutatják 30-30-30 neuront magába foglaló 3 rejtett réteg elhelyezésekor. Megfigyelhető, hogy a bemeneti 5 tanítópontra, illetve a 9 DCT-együtthatóra alapozott tanítás sokkal jobb egyezést indukál a betanított és illesztett MRI-nyelvkontúrok között az 1 tanítópont, illetve az 1 DCT-együtthatóval megvalósított tanítással szemben.

Emellett az is érzékelhető, hogy a kimeneti tanítópontok ellenében a DCT-együtthatók betanításával jóval finomabb illeszkedés érhető el a két MRI-nyelvkontúr között.

Összességében véve tehát kijelenthető, hogy a kvalitatív és a kvantitatív következtetések kölcsönösen hitelesítik egymást, ami összhangban van a 6.4.1. és 6.4.3. alfejezetek kimeneteleivel is. A jelenlegi eredmények továbbá azt is igazolják, hogy a neurális hálózat tanítását nem befolyásolja kedvezőtlenül az adatközlők nemének és anyanyelvének különbözősége, hiszen a tanító- és a megtanulandó paraméterek típusának és számának megfelelő megválasztásával nagyon jó egybevághóság érhető el a betanított és illesztett MRI-nyelvkontúrok viszonylatában, amit a numerikus adatok is megszilárdítanak.

6.5. T2 tézis

A kétdimenziós UH- és MRI-felvételekre illesztett nyelvkontúrok gépi tanuló-algoritmusok segítségével összehangolhatók úgy, hogy UH-nyelvkontúrok paramétereire alapozva betaníthatók az MRI-nyelvkontúrok alakjai. A betanított és az automatikus kontúrkövető algoritmussal illesztett MRI-nyelvkontúrok közötti kvalitatív és kvantitatív egyezés mértéke javul, ha a kimeneten tanítópontok helyett DCT-együtthatókat alkalmazunk függetlenül a bemeneti paraméterek típusától. (Kapcsolódó saját publikációk: [S3],[S6],[S7],[S8],[S10])

7. Artikulációs beszédszintézis

Kutatómunkám harmadik fontosabb szakaszában gépi beszéd előállításával foglalkoztam, melynek kiindulópontját a korábbi fejezetekben felhasznált UH- és MRI-felvételek képezték. Az volt ugyanis az alapvető elképzelésem, hogy a beszédszintézist olyan vizuális információkra támaszkodva valósítsam meg, amik az említett kétdimenziós képi forrásokból közvetlenül vagy közvetve kinyerhetők. A szükséges geometriai adatok egy részét a 3. és 4. fejezetekben bevezetett nyelv- és szájpaddockontúrok segítségével származtattam úgy, hogy MATLAB-környezetben kidolgoztam egy algoritmust, melynek alkalmazásával a vokális traktusban dinamikus módon megmérhetők a szájpadd és a nyelvfelszín közötti szagittális radiális távolságok. Az adathalmaz másik részét a nyelvkontúrokból kivont DCT-együtthatók formájában határoztam meg. A kapott geometriai adatokat a gépi tanulás eszközei révén próbáltam meg összekapcsolni a beszédet jellemző különböző artikulációs paraméterekkel, melyeket az akusztikus csőmodell vagy a lineáris predikció elvének keretében értelmeztem. Ennek során önálló, kitartott beszédhangokat, illetve folyamatos beszédet kívántam produkálni MATLAB-környezetben. A következő alfejezetekben rövid áttekintést adok az akusztikus csőmodellről, illetve a lineáris predikcióról, majd ismertetem a dinamikus távolságmérés általam kidolgozott módszereit és a szintézis eredményeit. A távolságmérés programkódjai az alábbi linken tekinthetők meg:

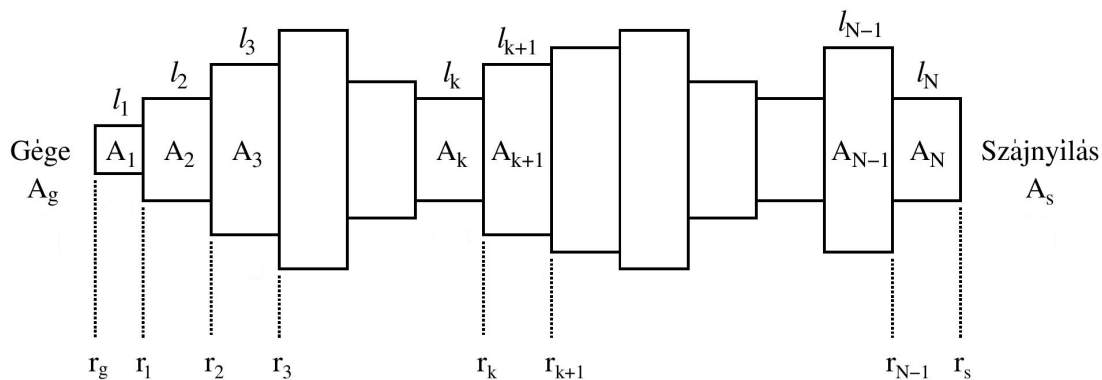
<https://drive.google.com/drive/folders/1rUg6m9Y9w6t5QWWiwHXCQ4y5LI4fYLZA?usp=sharing>

A beszédszintézis programkódjai az alábbi linken tekinthetők meg:

https://drive.google.com/drive/folders/1ko6PBjg9Gbt_UaFnhn5klT9jLu_qHoSA?usp=sharing

7.1. Az akusztikus csőmodell [69]

Az emberi beszédkeltés leképezésének egyik leggyakrabban alkalmazott és leghatékonyabb eszköze az akusztikus csőmodell. Ennek keretében a gégtől a száj- és orrnyílásig terjedő vokális traktust az egyik végén zárt, a másik végén nyílt rezonátorrendszerként kezeljük, mely akusztikai szűrőként működve egy több frekvenciakomponenst is tartalmazó hanghullámból csak adott frekvenciasávokba eső komponenseket enged át és sugároz ki a szabad térbe. Beszéd közben a tüdőből kipréselt levegő kváziperiodikus, turbulens vagy lökésszerű hullámmá alakulva gerjeszti a száj- és orrüreg által alkotott csatolt csőpárt. Az így létrejövő levegőáram a száj- és/vagy orrüregen keresztülhaladva az artikuláció során nyeri el végleges formáját, ami orális, nazális vagy nazorális artikulációt eredményez. A nazális régió határfalai rögzítve maradnak a beszédhangok képzésekor, így az orrüreg mérete és alakja állandónak tekinthető, aminek következtében az orrüreg fizikailag csupán egy passzív terelőcsatornát képez a levegőáram útjában. Ezzel szemben az orális régió térfogata dinamikusan változik az artikuláció során, hiszen a lágy szájpad, a nyelv, az állkapocs és az ajkak igen aktív és bonyolult módon összehangolt mozgást végeznek a kemény szájpad, a fogmeder és a fogak fix helyzetéhez viszonyítva. Ennek eredményeképpen a szájüreg keresztmetszete térben és időben fluktuál, amit matematikailag egy kétváltozós $A(x, t)$ függvénnyel írhatunk le, ahol x a vokális traktus tengelye mentén a gégtől mért pozíciót jelzi, t pedig az időparaméter. Miután a keresztmetszet térben és időben is folytonosan és véletlenszerűen változik a beszéd során, így nem tudjuk az $A(x, t)$ függvényt szabályos, zárt matematikai függvényalakokkal felruházni. Éppen ezért a folytonos toldalékcsovet egy térben szabályosan kvantált csővel közelítjük, azaz a vokális traktust felosztjuk N számú egymás után csatlakozó, egyenként állandó keresztmetszetű, rövid csőszakaszra, ahogyan azt a 7.1. ábra is szemlélteti. Az egyes csőszakaszok hosszúságát és keresztmetszeti paramétereit rendre az $l_1, l_2, l_3, \dots, l_k, l_{k+1}, \dots, l_{N-1}, l_N$, illetve az $A_1, A_2, A_3, \dots, A_k, A_{k+1}, \dots, A_{N-1}, A_N$ szimbólumok jelölik. A gégehez és a szájnyíláshoz két extra keresztmetszetet rendelünk A_g és A_s címkékkel. A 7.1. ábrán vázolt kép szerint gondolkodva azonnal felismerhető, hogy a szomszédos csőszakaszok csatlakozásánál történő ugrásszerű keresztmetszetváltásoknál a beszédhanghullámok visszaverődést szenvednek, amit fizikailag az $r_1, r_2, r_3, \dots, r_k, r_{k+1}, \dots, r_{N-1}$ reflexiós tényezőkkel írhatunk le. Ezt a halmazt kiegészítik a gégenél és a szájnyílásnál értelmezett r_g és r_s reflexiós tényezők.



7.1. ábra. A vokális traktus modellezése egyenként állandó keresztmetszetű, egyforma hosszúságú csőszakaszokkal.

A reflexiók tényezők és a keresztmetszetek között egyértelmű kapcsolat áll fenn, ami az

$$r_k = \frac{A_{k+1} - A_k}{A_{k+1} - A_k} \quad (7.1)$$

összefüggéssel fejezhető ki, ahol $k = 1, 2, \dots, N$. A fentebbiekkel egyetértésben a (7.1) reláció $r_N = r_s$ és $A_{N+1} = A_s$ figyelembevételével érvényes, emellett az is megjegyzendő, hogy az akusztikus csőmodell szerint $r_g = 1$ és $A_g = 1$. A modell tehát N csőszakasz alkalmazásakor $N+1$ reflexió tényezőt és $N+2$ keresztmetszetet definiál. A vokális traktus átlagos hossza a hangszalagoktól a szájnyílásig férfiak esetében 17 cm, nők esetében pedig 14 cm.

7.2. A lineáris predikció elve [69]

A lineáris predikció (LPC – Linear Predictive Coding) a digitális jelek feldolgozásának és becslésének igen széles körben elterjedt és sikeres eszköze, amely az időben lejátszódó jelenségek aktuális jellemzőjét empirikus információk felhasználásával állítja elő. Az eljárás ugyanis azon alapul, hogy a jel adott pillanatbeli értékét az azt megelőző időpillanatokhoz tartozó jelértékek segítségével megpróbáljuk előrejelezni, idegen szóval predikálni. A predikció abban az esetben lineáris, ha a becsült jel a becslés során alkalmazott értékek lineáris függvénye. Amennyiben az s jel mint idősor n -edik elemét az azt megelőző p darab minta alapján állapítjuk meg, akkor a jelzett lineáris viszony az

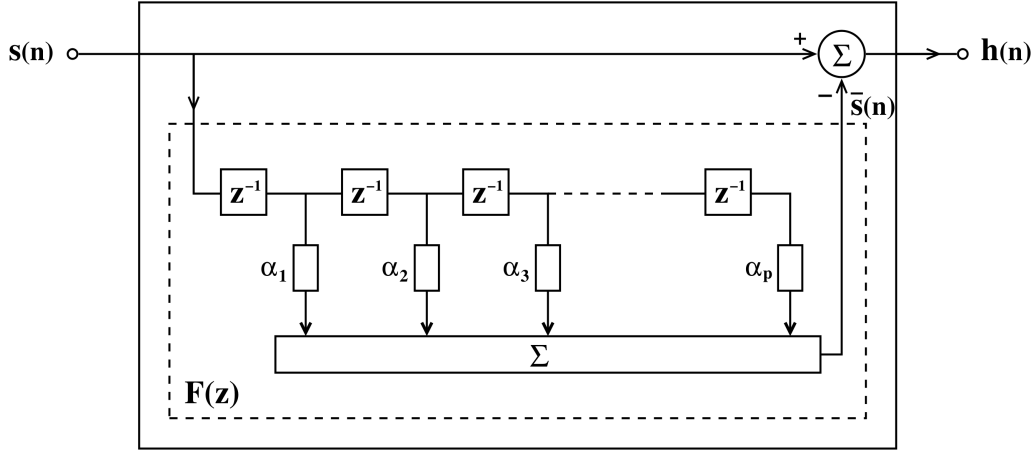
$$\bar{s}(n) = \sum_{i=1}^p \alpha_i s(n-i) \quad (7.2)$$

lineáris kombináció formájában jön létre, ahol az α_i koefficienseket predikciós együtthatóknak vagy LPC-együtthatóknak nevezzük, melyek a jelenségre vonatkozó előismeretet hordozzák. A p paraméter a predikció fokszámát adja meg. A predikciós együtthatókban foglalt tapasztalat általában nem elegendő az egzakt becsléshez. Az előrejelzés pontatlanságát (7.2) felhasználásával a

$$h(n) = s(n) - \bar{s}(n) = s(n) - \sum_{i=1}^p \alpha_i s(n-i) = \sum_{i=0}^p \beta_i s(n-i) \quad (7.3)$$

predikciós hibafüggvény jellemzi, melynek értéke mintáról mintára változik. (7.3) utolsó egyenlősége szerint a hibafüggvény egyetlen összegtagba tömöríthető, ezáltal a p darab α_i együttható $p+1$ számú β_i együtthatóvá transzformálható, és a (7.3) reláció tanulmányozásával rögtön leolvasható, hogy $\beta_0 = 1$, illetve $\beta_i = -\alpha_i$, ha $i = 1, 2, \dots, p$.

Rendszertechnikai szempontból az eredeti $s(n)$ jel, illetve a $h(n)$ hibajel úgy is felfogható, mint a predikciós koefficiensek által meghatározott fizikai rendszer bemeneti és kimeneti jele, ami a lineáris predikció 7.2. ábrán feltüntetett analízismodelljéhez vezet. Az elnevezés arra utal, hogy a kimeneten megjelenő hibajel analízise nyomán feltérképezhető az optimális, azaz a hibajel minimalizáló predikciós együtthatók csoportja. Ez azt jelenti, hogy a predikciós koefficiensek folyamatos változtatásával beállítható egy olyan konfiguráció, amikor a hibajel éppen eléri a minimális értékét, tehát éppen rátalálunk az ideális együtthatók halmazára, végrehajtva a legjobb predikciót.

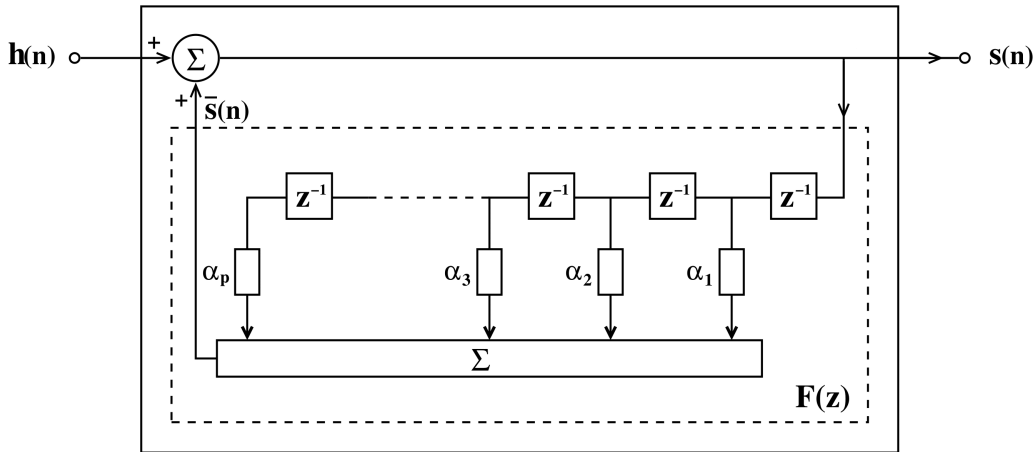


7.2. ábra. A lineáris predikció analízismodellje.

A 7.2. ábra szaggatott vonallal határolt,

$$F(z) = \sum_{i=1}^p \alpha_i z^{-i} \quad (7.4)$$

átviteli függvénnyel jellemzett moduljában a z^{-1} faktorok egy ütemnyi késleltetést valósítanak meg, és a (7.4) kifejezés z változójának soron következő $-i$ -edik hatványa i számú blokk összeszorozása révén állítható elő a késleltetők láncolata mentén. Az így kialakuló produktumot az aktuális leágazásnál található α_i koefficiens szorozza, végül a létrejövő tagokat a Σ tömb összesíti, ami a (7.4) összeget generálja. Az $s(n)$ gerjesztőjelből kivonva az $F(z)$ átviteli függvénnyel felruházott rendszer által előállított $\bar{s}(n)$ becsült jelet, a (7.3) különbségnek megfelelően a rendszer válaszként létrehozza a $h(n)$ predikciós hibát.



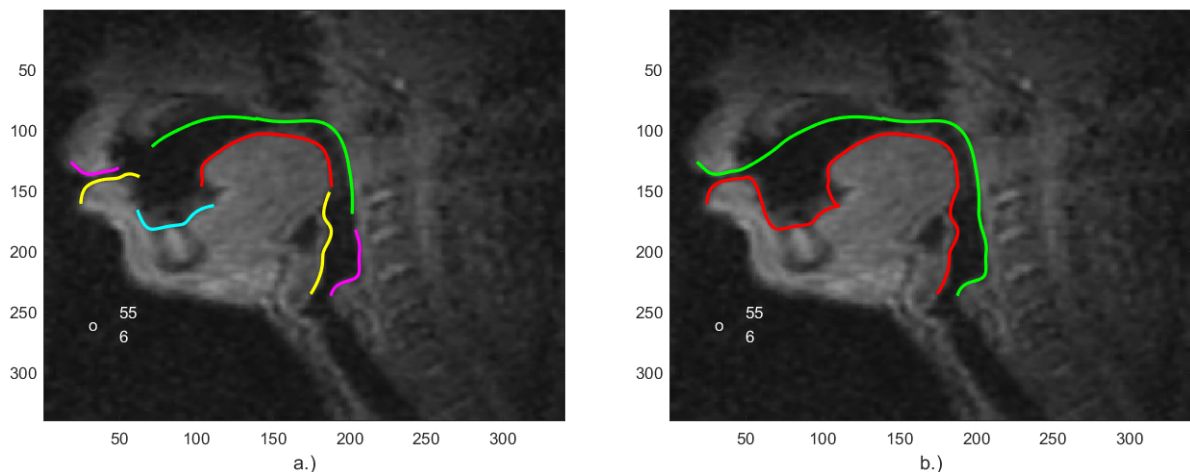
7.3. ábra. A lineáris predikció szintézismodellje.

Amennyiben a $h(n)$ hibafüggvény gerjesztőjelként szolgál, és az eredeti $s(n)$ jel a rendszer válaszként detektálható, akkor eljutunk a lineáris predikció szintézismodelljéhez, amit a 7.3. ábra szemléltet. A modell elnevezése azt tükrözi, hogy az $s(n)$ jel a rendszer kimenetén a bemenetre kapcsolt hibajelből szintetizálható. A 7.3. ábrán követhető, hogy a $h(n)$ hibajelnek, illetve az $F(z)$ átviteli függvénnyel ellátott modul által előállított $\bar{s}(n)$

becsült jelnek az összege a rendszer kimenetén az $s(n)$ jelet produkálja. Figyeljük meg, hogy a jel egy része a kimenetről visszacsatolódik és a hibajelhez hozzáadódva visszatér önmagába.

7.3. Az MRI-kontúrok kiterjesztése

A vizuális adatok UH- és MRI-felvételekből történő kinyeréséhez olyan anatómiai kontúrvonalakra volt szükségem, melyek a lehető legteljesebb mértékben képesek lefedni a vokális traktus látótérbe eső tartományait. Ez az elvárás az UH-felvételek esetében már eddig is maximálisan teljesült, hiszen a 3. és 4. fejezetek szerint megkonstruált nyelv- és szájpadkontúrokon kívül nem detektálhatók további görbék a vokális traktusban. Ennek technikailag az az akadálya, hogy a részlegesen megjelenő szájüregen kívül az artikulációs csatorna többi része nem hozzáférhető az UH-letapogatás számára. Ezzel szemben az MRI-felvételek esetében a 3. és 4. fejezetekben megállapított nyelv- és szájpadkontúrokkal még korántsem tekinthető komplettnek a vokális traktus anatómiai kontúrjainak azonosítása, mivel ezek a görbék a gége és a szájnyílás irányába is kiterjeszthetők. Ezért a 3. fejezetben részletezett kontúrkövető algoritmust adaptáltam a gégefedő, a hátsó garatfal, a nyelv alatti szublingvális terület, valamint az alsó és felső ajkak régióiban található határvonalak megrajzolására. Ennek során a 3. fejezetben bemutatott programlépések szerint jártam el, az algoritmus paramétereinek helyes beállítását azonban minden szervi tartomány esetében hosszas kísérletezéssel, egyedi módon kellett kitapasztalnom. Ez azt jelentette, hogy más-más helyre kellett pozicionálnom a sugaras mintavételezés középpontját (3.3. ábra), módosítottam a vízszintes relevanciatartomány határszögeinek értékeit és a függőleges relevanciatartományt, szükség szerint szimmetrikusan felcseréltem az élkiemelő mátrixok elemeinek pozícióit (3.6. ábra), változtattam az élkiemelő és átlagoló mátrixok méretét, a radiális mintavételezés során beállított sugárirányú hatótávolságot, valamint a görbék simításához szükséges DCT-együtthatók számát.

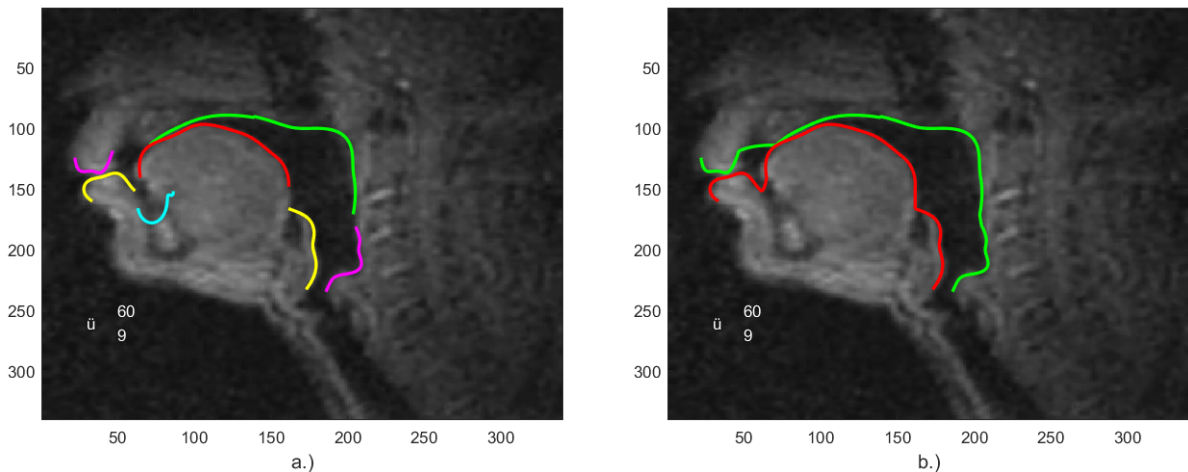


7.4. ábra. A vokális traktus különböző tartományainak kontúrjai szeparált (a.) és interpolációval egyesített (b.) formában az *o* hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)

A kontúrvonalak kiterjesztését a 7.4. ábra illusztrálja az *o* hang példáján keresztül. A 7.4.a ábrán az eredetileg is felvett piros nyelvkontúr és zöld szájpadkontúr mellett a vokális traktus további szakaszainak határvonalai is ki vannak emelve különböző színekkel.

A gége felőli oldalon sárga és lila görbék jelzik a gégefedő és a hátsó garatfal kontúrjait, a szájnyílás felőli oldalon szintén sárga és lila görbékkel vannak megrajzolva az alsó és felső ajkak kontúrjai, a kék görbe pedig a szublingvális kontúrt adja meg. A kapott határvonalakat interpolációval kapcsoltam össze úgy, hogy a gégefedőtől az alsó ajakig, illetve a hátsó garatfaltól a felső ajakig egybefüggő, folytonos kontúrok jöjjenek létre. Az interpolációt szakaszonként köbös Hermite-polinomokra alapozott technikával valósítottam meg (PCHIP – Piecewise Cubic Hermite Interpolating Polynomial). Az interpolációval kialakított folytonos kontúrokat a 7.4.b ábra demonstrálja.

A 7.4.a ábrán látható kontúrszakaszok közül a szublingvális görbe kezelése igényelte a legnagyobb körültekintést, ugyanis ez a régió csak akkor válik relevánssá, ha a nyelv hátsó pozícióba húzódik, ahogyan ez a fentebbi példában az *o* hang esetén is nyilvánvaló. Ha azonban a nyelv elülső állást vesz fel, akkor a szublingvális tartomány a kontúrkeresés szempontjából semmiféle szerepet nem játszik. Éppen ezért speciális feltételt róttam ki a szublingvális kontúr figyelembevétele kapcsán. Ennek megfelelően hátsó nyelvállásban a nyelvkontúr a szublingvális görbe közbeiktatásával csatlakozik az alsó ajakkontúrhoz, elülső nyelvállásban pedig a szublingvális görbe ignorálva van, és a nyelvkontúr közvetlenül kapcsolódik az alsó ajakkontúrhoz. Ez utóbbi művelet eredménye a 7.5. ábrán figyelhető meg az *ü* hang esetében.



7.5. ábra. A vokális traktus különböző tartományainak kontúrjai szeparált (a.) és interpolációval egyesített (b.) formában az *ü* hang esetében. (Az ábra tengelyeinek mértékegysége pixel.)

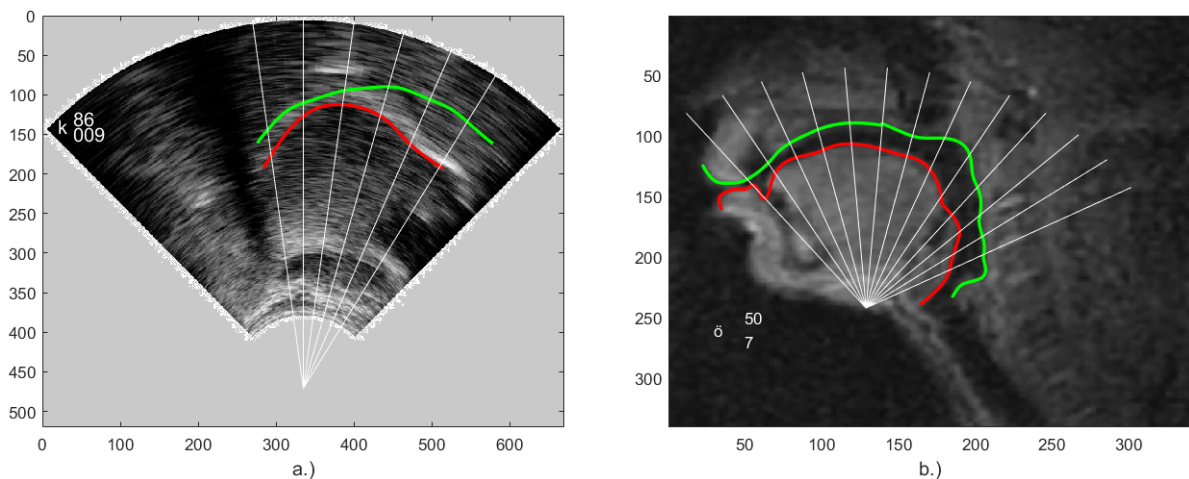
Miután az MRI-felvételeken nem látszanak a beszélő fogai, így ebben a tartományban a kontúrkövetés során csak az alsó és felső fogmeder határvonalai állapíthatók meg.

7.4. Dinamikus távolságmérés az UH- és MRI-felvételeken

A beszédszintézishez szükséges vizuális geometriai adatokat az UH- és MRI-felvételek nyelv- és szájpadkontúrjai között mérhető szagittális radiális távolságok formájában állítottam elő. A távolságméréshez a 3. és 4. fejezetek UH-görbéit, valamint a 7.3. alfejezet kiterjesztett MRI-görbéit használtam fel, és két különböző elveken alapuló mérési mechanizmust dolgoztam ki az adatok dinamikus kinyerésére, melyeket a következő alfejezetekben taglalok.

7.4.1. Távolságmérés 1

A módszer kidolgozását radiális geometriában valósítottam meg. Ehhez igazodva, elsőként egy rögzített középpontból indulva, sugárirányú metszeteket képeztem az UH- és MRI-felvételek releváns körcíkkéi által definiált tartományokban, amint azt a 7.6. ábra fehér vonalai is érzékeltetik. A 7.6.a ábrán látható UH-keret C pontjának koordinátáit továbbra is az (5.1)-ben megadott értékek határozzák meg, a 7.6.b ábra MRI-keretének C' pontját pedig az 5.5.4. alfejezetben ismertetett optimalizációs procedúra eredményeinek figyelembevételével az 5.8. táblázatban feltüntetett koordináták szerint jelöltem ki. A 7.6.a ábra egy k hanghoz tartozó nyelvkontúrt jelenít meg, a 7.6.b ábra pedig egy $\bar{o}$ hanghoz rendelt nyelvkontúrt prezentál. A radiális metszetek a 0° -nál elhelyezkedő vertikális egyenes adott középpont körüli elforgatásával hozhatók létre a választott szögtartományon belül. A 7.6. ábrán megrajzolt radiális metszetek az UH-képek esetében a $[-8^\circ, 32^\circ]$, az MRI-képek esetében pedig a $[-36^\circ, 60^\circ]$ intervallumok által megszabott szögtartományt fedik le úgy, hogy az egyes metszetek 8° -onként követik egymást. A releváns szögtartományokat mindkét forrásnál úgy állítottam be, hogy az összes tanulmányozott bemondás képkockáira megfelelő legyen, azaz a nyelvkontúr a nyelv mozgása során ne lépjen ki a felvett szögintervallumból.

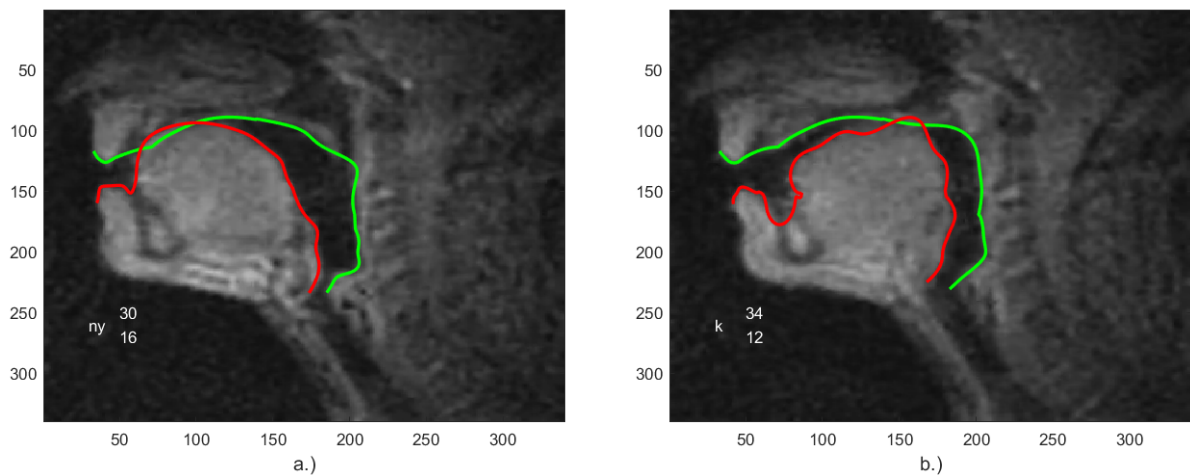


7.6. ábra. A szagittális radiális távolságok méréséhez felvett sugárirányú metszetek egy k hanghoz tartozó UH-kereten (a.), illetve egy $\bar{o}$ hanghoz rendelt MRI-kereten (b.). (Az ábra tengelyeinek mértékegysége pixel.)

Az algoritmus második lépésében az 'intersect' függvény segítségével megkerestem a radiális metszetek és a nyelv-, illetve a szájpadkontúr által képzett metszéspontok koordinátáit, ami a 7.6. ábrával egyetértésben azt jelenti, hogy a fehér-piros, illetve a fehér-zöld görbepárok közös pontjait kell megtalálni. Ez a geometriailag egyszerű művelet azonban korántsem tekinthető triviálisnak, hiszen a radiális metszetek, valamint a nyelv- és szájpadkontúrok pontjait jellemző koordináták nem egész számok, hanem lebegőpontos valós változók. Ez azt vonja maga után, hogy az adott görbepárnak szigorú értelemben véve nincsenek közös pontjai, mivel két lebegőpontos számérték csak a legritkább esetekben egyezik meg egymással több tizedesjegynyi pontosságig. Ezért, hogy áthidaljam ezt a problémát, kerekítési módszerek bevetéséhez folyamodtam. Egy adott görbepár esetén mindkét görbe pontjai x és y koordinátákkal írhatók le, tehát összesen négy halmazban

kell kerekítéseket végezni, melynek alapvetően három típusát különböztethetjük meg: a közelebbi egész értékre kerekítés ('round'), a negatív végtelen, azaz a kisebb egész érték irányába történő kerekítés ('floor'), valamint a pozitív végtelen, azaz a nagyobb egész érték irányába történő kerekítés ('ceil'). A görbepárok x és y koordinátáinak szintjén a szóban forgó kerekítési mechanizmusok összes lehetséges kombinációját megalkotva, *if* feltételek sokaságának kijelölésével sikerült leolvasni a görbepárok metszéspontjait. Ez összesen 80 lehetséges kombinációt takar, vagyis 80 egymást követő és teleszkopikusan egymásba ágyazott *if* feltétel láncolatán keresztül történik a két görbe közös pontjainak megtalálása. A görbepár metszéspontjainak birtokában már egyszerűen kiszámolhatók a nyelv- és szájpaddockontúr között mérhető radiális távolságok, hiszen az adott sugárirányú metszet mentén található két metszéspont $(x; y)$ koordinátáinak felhasználásával és az (5.9) típusú pitagoraszsi összefüggés alkalmazásával megadható a sugárirányú távolság.

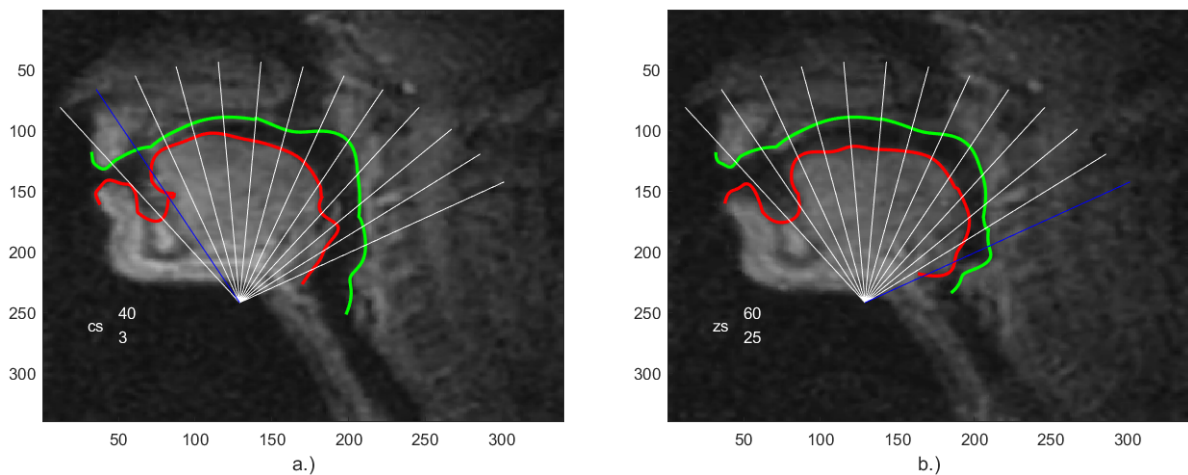
Az anatómiai kontúrvonalak követése során előfordulhatnak olyan anomáliák, melyek nem felelnek meg a valóságban elvárható körülményeknek, és érdemi módon befolyásolják a távolságmérés eredményeit. Ilyen helyzet áll elő például abban az esetben, amikor az artikuláció során a nyelv érintkezésbe kerül a kemény vagy lágy szájpaddal. Ekkor a két szövettartomány képi összeolvadása miatt bizonytalanná válik a kontúrkeresés, ami ahhoz vezethet, hogy a nyelvkontúr túllépi a szájpaddockontúrt. Ezt a hibatípust szemlélteti a 7.7. ábra az *ny* és *k* hangok példái által a kemény és lágy szájpadd tartományaiban. Ezekben az esetekben a távolságmérés nyilvánvalóan fals kimenetelt produkál, ami korrigálásra szorul. A korrekciót oly módon végeztem el, hogy amennyiben a nyelvkontúr túlsordul a szájpaddockontúron, azaz – a 7.7. ábra vertikális skálájával összhangban – az adott radiális metszet mentén a nyelvkontúr pontjának függőleges koordinátája kisebb vagy egyenlő a szájpaddockontúr pontjának függőleges koordinátájával, akkor az (5.9) szerint számolt távolságot a megfelelő negatív értékkel helyettesítettem, majd ezeket az elemeket nullára cseréltem, mivel a nyelv és a szájpadd érintkezésekor a két kontúr közötti távolságnak valóban nullának kell lennie.



7.7. ábra. A nyelvkontúr túlsordul a szájpaddockontúron az *ny* (a.) és *k* (b.) hangok esetében. (Az ábra tengelyeinek mértékegysége pixel.)

A nyelvkontúrok fentebb bemutatott rendellenessége mellett másféle komplikációk is felléphetnek a távolságok mérésekor, amik egyrészt abból adódhatnak, hogy a radiális metszővonalak és az adott kontúr metszéspontjainak keresésekor alkalmazott kerekítési

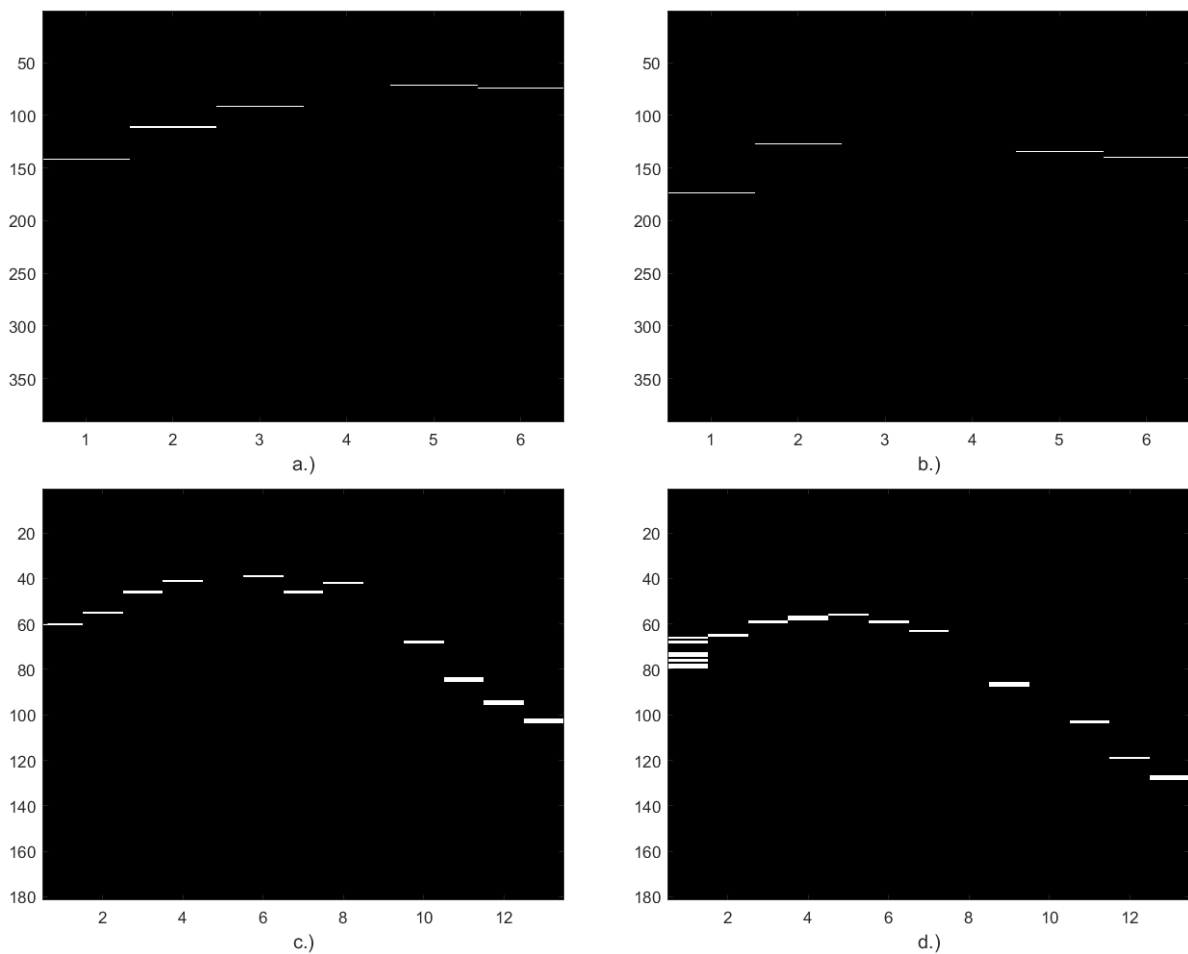
technikák kombinálása többszörös multiplicitású metszéspontokat is eredményezhet, ami abban nyilvánul meg, hogy a közös pont ugyanolyan koordinátákkal többször is előfordulhat vagy több, különböző koordinátákkal jellemzett közös pont jön létre. A metszéspontok sokszorozódásának másik oka az lehet, ha egy kontúr bizonyos tartományokban túlságosan nagy görbülettel rendelkezik, így az adott radiális metszővonal a görbeszakasz több pontját is keresztezi. Az utóbbi helyzetet demonstrálja a 7.8. ábra a *cs* és *zs* hangok esetében. Megfigyelhető, hogy a 7.8.a ábrán a balról számított, kékkel kiemelt második metszet három pontban érinti a nyelvkontúrt, és hasonlóképpen a 7.8.b ábrán a jobb szélső kék metszővonal a nyelvkontúr két pontján is áthalad. Miután az algoritmus nem tudja megkülönböztetni a metszéspontok két eltérő forrásból eredő multiplicitását, ennél fogva úgy jártam el, hogy egynél több közös metszéspont esetén a radiális metszetek pozícióit meghatározó szögértékek, illetve a többszörös multiplicitású metszéspontok valós és képzetes koordinátáinak segítségével kitűzött feltételek érvényesítésével szűrtem ki a helyes metszéspontot. Ennek lényege abban áll, hogy amennyiben a szögérték nulla, azaz a vertikális egyenes metszi a görbét, akkor nyelvkontúrok esetén a legkisebb képzetes koordinátájú pont, szájpadkontúrok esetén pedig a legnagyobb képzetes koordinátájú pont adja meg a helyes metszéspontot. Amennyiben a szögérték pozitív (vagy negatív), azaz a vertikális egyeneshez képest a jobb (vagy bal) oldali képtartományban helyezkedik el a radiális metszet, akkor nyelvkontúrok esetén a legnagyobb (vagy legkisebb) valós koordinátájú pont, szájpadkontúrok esetén pedig a legkisebb (vagy legnagyobb) valós koordinátájú pont jelöli ki a helyes metszéspontot. A fentebbi feltételekben szereplő szélsőértékeket az UH- és MRI-keretek vízszintes és függőleges tengelyskáláinak iránya, valamint a nyelv- és szájpadkontúrok szájüregi helyzete alapján állítottam be úgy, hogy az adott radiális metszet mentén nyelvkontúrok esetén mindig a legfelső, szájpadkontúrok esetén pedig mindig a legalsó pont legyen meghagyva, biztosítva ezzel azt, hogy alulról és felülről is a szájüreg felé közelítsünk a többszörös multiplicitású metszéspontok halmazában.



7.8. ábra. Többszörös metszéspontok létrejötte a kékkel megjelölt radiális metszetek mentén nagy görbületű kontúrszakaszoknál a *cs* (a.) és *zs* (b.) hangok esetében. (Az ábra tengelyeinek mértékegysége pixel.)

7.4.2. Távolságmérés 2

A módszer kifejlesztése a radiális geometriának négyzetes elrendezésbe történő transzformációján alapult. Az eljárás fő lépése ugyanis az volt, hogy az UH- és MRI-felvételek 7.6. ábrán rögzített középpontjaiból kiindulva és a 7.4.1. alfejezetben kijelölt szögtartományokat és szögbeosztást megőrizve, a létrejövő radiális irányok mentén mintavételeztem a nyelv- és szájpgadkontúrokat, majd a kapott minták sorozatát a 3. fejezetben a 3.5. ábra segítségével elmagyarázott transzformációs technika alkalmazásával mátrixos struktúrába rendeztem, ami a minták láncolatának négyzetes síkba való kifizetését jelenti.



7.9. ábra. UH- (a. és b.) és MRI-keretek (c. és d.) radiális geometriában mintavételezett és négyzetes geometriába transzformált mintasorozatai szájpgad- (a. és c.) és nyelvkontúrok (b. és d.) esetén. (Az ábra tengelyeinek mértékegysége pixel.)

A művelet eredményeként létrejövő, mátrixos geometriába alakított mintasorozatok láthatók a 7.9. ábrán. A 7.9.a és 7.9.b, illetve a 7.9.c és 7.9.d blokkok a 7.6.a, illetve a 7.6.b ábra nyelv- és szájpgadkontúrjaiból származnak. Ahogyan azt a 7.9.a-b, illetve 7.9.c-d ábrák vízszintes tengelyei is mutatják, az UH-képek esetében 6, míg az MRI-képek esetében 13 oszlopot tartalmaznak a mátrixos struktúrák, ami azzal van összefüggésben, hogy 8° -onként beosztva a $[-8^\circ, 32^\circ]$, illetve a $[-36^\circ, 60^\circ]$ intervallumokat, éppen 6, illetve 13 különböző szögérték áll elő. A 7.9. ábrán az is szembetűnő, hogy a képmátrixnak lehetnek olyan oszlopai,

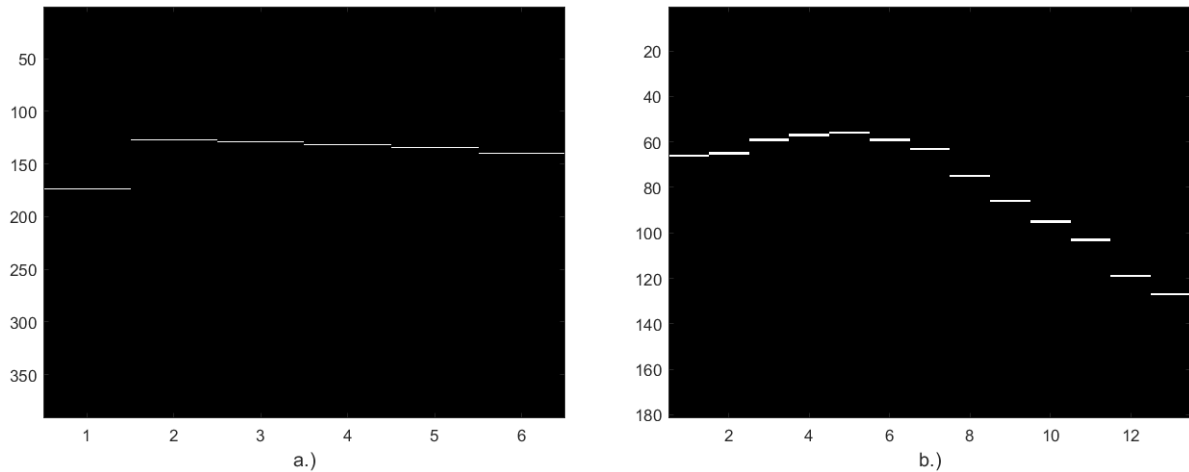
melyekben nem jelenik meg vízszintes fehér vonal, vagyis ezek az oszlopok nem hordoznak mintákat az adott kontúrból. Ennek az az oka, hogy a nyelv- és szájpaddockontúrokat végső soron diszkrét pontok alkotják, ami a 3. fejezetben kifejtett kontúrkereső algoritmus diszkrét radiális képi mintavételezésének a következménye. Ennélfogva megtörténhet az, hogy a kész kontúrok távolságméréshez szükséges mintavételezésekor egy olyan sugárirányú metszet mentén haladunk, ahol nincs kontúrponthoz. Ilyen hézagokat figyelhetünk meg például a 7.9.b ábra 3. és 4., valamint a 7.9.c ábra 5. és 9. oszlopaiban. Ezeken túlmenően az is világosan látszik, hogy a 7.9. ábra mátrixainak egyes oszlopaiban akár egynél több vízszintes fehér vonal is felsorakozhat, ami vagy a speciális radiális mintavételezés vagy a nagy görbülettel rendelkező kontúrszakaszok hatásaként bukkan fel. Ilyen többszörös minták rajzolódhatnak ki például a 7.9.d ábrán, ahol a mátrix 4., 9. és 13. oszlopai kétszeres mintákat foglalnak magukba, és a mátrix 1. oszlopában hét minta tornyosul egymás fölé. A kétszeres minták a radiális mintavételezés sajátosságainak tudhatók be, a hétszeres minta pedig a 7.6.b ábra szájnnyílás felőli első metszete mentén keletkezik, miután a kiterjesztett nyelvkontúr gyakorlatilag párhuzamos a metszővonallal az érintett szakaszon. Vegyük észre, hogy ez utóbbi többszörös minták megfeleltethetők a 7.4.1. alfejezetben ismertetett módszer által produkált többszörös metszéspontoknak, melyek a nagy görbületű kontúrszakaszokon jönnek létre. Összefoglalva tehát megállapítható, hogy a nyelv- és szájpaddockontúrokból eredő mintasorozatok négyzetes kifestése rendkívül szemléletesen tükrözi a hézagokat és a többszörös mintákat is, ami nagymértékben megkönnyítette ezen anomáliák egyértelmű azonosítását és kiküszöbölését.

Elsőként a többszörös minták által generált bizonytalanságokat oldottam fel úgy, hogy minden képmátrixban megkerestem a többszörös mintákat tartalmazó oszlopokat, és a nyelvkontúrok esetén csak a legfelső, míg a szájpaddockontúrok esetén csak a legalsó elemet tartottam meg az adott oszlopbeli mintatömbből, a többi elemet pedig elimináltam. A többszörös minták kiiktatásának ezen stratégiáját az indokolja, hogy a képmátrixok függőleges tengelyét tekintve, a nyelvkontúrok esetén a kisebb koordináták irányába (függőlegesen felfelé), a szájpaddockontúrok esetén pedig a nagyobb koordináták irányába (függőlegesen lefelé) haladva közelítünk a szájüreg felé. Ez a szemlélet egybevág azzal, ahogyan a 7.4.1. alfejezetben vázolt módszer redukálja a többszörös mintákat. Maradva a 7.9. ábra példájánál, ez azt jelenti, hogy a szájüreg a 7.9.a és 7.9.c blokkok mintasorozatai alatt, illetve a 7.9.b és 7.9.d blokkok mintasorozatai fölött helyezkedik el, így természetes, hogy az említett mintatömbökben a szájpaddockontúrok esetében a legalsó, a nyelvkontúrok esetében pedig a legfelső vízszintes fehér vonal jellemzi reálisan a szájpaddock vagy a nyelvfelszín adott szakaszát.

A következő feladat a képmátrixok üres oszlopainak kitöltése volt, amivel megszüntethetők a mintasorozatok hézagjai. A mintasorozatok üres elemei rendkívül változatosan helyezkedhetnek el a vízszintes fehér vonalak által alkotott lánc mentén. Mindenekelőtt azt mértem fel, hogy a hézagok potenciálisan milyen pozíciókban és milyen multiplicitással fordulhatnak elő a mintasorozatok sokaságában. A vizsgálataim során arra jutottam, hogy a hézagok megjelenésének típusa alapvetően kétféle csoportba sorolható. Ennek megfelelően elkülöníthetők a mintasorozat belsejében lévő egyszeres vagy többszörös üres elemek, illetve a mintasorozat elején vagy végén lévő egyszeres vagy többszörös üres elemek. Miután a fentebbi két hibaosztály eltérő tulajdonságú halmazokat határoz meg, így a hibákat is más-más módon kellett kijavítanom, hogy végül hézagmentes mintasorozatok álljanak a rendelkezésemre. A belső hézagokat lineáris interpolációval, a szélső hézagokat pedig

lineáris extrapolációval töltöttem fel.

A fentebb bemutatott korrekciós eljárások alkalmazásával kijavított pixelláncokat szemléltetnek a 7.10.a és 7.10.b ábrák, melyek a 7.9.b és 7.9.c ábrákon látható mintasorozatokból származnak.



7.10. ábra. A többszörös metszéspontok és a hézagok megszüntetésével előállított, korrigált mintasorozatok egy UH- (a.) és egy MRI-nyelvkontúr (b.) esetében. (Az ábra tengelyeinek mértékegysége pixel.)

Itt jegyzem meg, hogy a 7.4.1. alfejezetben tárgyalt módszer esetében gyakorlatilag a kerekítési mechanizmusok kombinált alkalmazása oldja fel a mintasorozatok belsejében vagy széleinél kialakuló egyszeres és többszörös hézagokat, hiszen a kerekítések révén olyan pontokban is létrejöhetnek a radiális metszővonalak és a görbék metszéspontjai, ahol egyébként nincs kontúrpont.

A többszörös metszéspontoktól és hézagoktól megszabadított mintasorozatok már kényelmes alapot biztosítottak a nyelv- és szájpadkontúrok közötti távolságméréshez, amit a 7.9. ábrán is példázott négyszöges geometriában úgy hajtottam végre, hogy az összetartozó nyelv- és szájpadkontúrt leképező két mátrix minden oszlopában kiszámítottam a görbepár ugyanazon radiális metszet mentén lokalizálható pontjait megjelenítő két vízszintes fehér vonal függőleges koordinátáinak különbségét. Ezzel a módszerrel tehát a nyelv- és szájpadkontúrok között mérhető radiális távolságokat a transzformált négyszöges struktúrákból olvastam ki, feltételezve, hogy a két görbe relatív távolsága ugyanaz a radiális és a négyszöges geometriában. Ez a feltételezés logikusnak tűnik, mivel mindkét kontúr ugyanabban a radiális környezetben van felvéve, és a két görbe transzformációja ugyanolyan módon, ugyanolyan paraméterhalmaz beállításával zajlik. Másképpen megfogalmazva, a transzformáció megtartja a relatív távolságokat. A módszer a 7.4.1. alfejezetbeli eljárással ellentétben képes közvetlenül előjeles távolságokat mérni, amivel sokkal egyszerűbbé válik a hibás távolságadatok kiszűrése. A negatív távolságértékek ugyanis azonnal jelzik azokat az anomáliákat, amikor a nyelvkontúr bizonyos szakasza túllép a szájpadkontúron, így nem szükséges extra feltételt kikötni a hibás adatok korrigálásához, hanem elegendő a negatív értékeket nullára cserélni, ami megfelel a nyelv és a szájpád érintkezésének.

7.4.3. A távolságmérő módszerek eredményeinek összehasonlítása

A radiális távolságmérés 7.4.1. és 7.4.2. alfejezetekben áttekintett módszerei által adott eredményeket egy tetszőlegesen kiválasztott UH- és MRI-kereten megrajzolt, *á* hanghoz tartozó nyelv- és szájpaddocktúrok példáin keresztül hasonlítottam össze. A 7.1., illetve 7.2. táblázatok a jelzett UH-, illetve MRI-keret görbepárjai közötti távolságadatokat foglalják össze a *Távolságmérés 1* és *Távolságmérés 2* metódusok által produkált számértékek összevetésével, ahol a távolságok pixel egységben vannak kifejezve. A táblázatok első soraiban figyelemmel kísérhető a $[-8^\circ, 32^\circ]$, illetve a $[-36^\circ, 60^\circ]$ intervallumok radiális metszeteit meghatározó előjeles szögértékek. A 7.1. táblázatban a legkisebb negatív, míg a 7.2. táblázatban a legnagyobb pozitív értéktől indulva haladunk balról jobbra az adott szögtartomány pontjai mentén. Ez az irány mindkét táblázat esetében megfelel a gégétől a szájnylás felé történő mozgásnak, ami a két táblázatban azért fordított sorrendű, mert az UH- és MRI-felvételek anatómiai szempontból tükrözött módon ábrázolják a hangképző szerveket, tehát az UH-képeken a negatív, az MRI-képeken pedig a pozitív szögértékekkel jellemzett radiális metszetek vannak közelebb a gégéhez. A 7.1. és 7.2. táblázatok utolsó sorai a *Távolságmérés 1* és *Távolságmérés 2* módszerekkel mért távolságok metszetenkénti különbségeit tárják fel. Látható, hogy igen jó egyezés mutatkozik a két eljárással kapott kimenetek között, hiszen a két adatsor elemei között legfeljebb néhány pixelnyi eltérés tapasztalható. Ez a tendencia a felhasznált UH- és MRI-keretek túlnyomó részében helytálló.

7.1. táblázat. Egy kiválasztott UH-keret nyelv- és szájpaddocktúrai között mérhető radiális távolságok a *Távolságmérés 1* és *Távolságmérés 2* metódusok alkalmazásával.

UH						
Metszet	-8°	0°	8°	16°	24°	32°
Távolságmérés 1	72	78	65	68	76	74
Távolságmérés 2	71	78	65	71	76	74
Különbség	1	0	0	-3	0	0

7.2. táblázat. Egy kiválasztott MRI-keret nyelv- és szájpaddocktúrai között mérhető radiális távolságok a *Távolságmérés 1* és *Távolságmérés 2* metódusok alkalmazásával.

MRI													
Metszet	60°	52°	44°	36°	28°	20°	12°	4°	-4°	-12°	-20°	-28°	-36°
Távolságmérés 1	22	19	13	22	28	27	28	29	34	32	24	23	36
Távolságmérés 2	22	19	13	22	28	27	27	29	32	31	24	23	35
Különbség	0	0	0	0	0	0	1	0	2	1	0	0	1

7.5. Beszédszintézis

A beszédszintézisre irányuló törekvéseim első állomásán önálló gépi beszédhangokat próbáltam meg előállítani. Ehhez olyan magánhangzókat választottam, melyek a rendelkezésemre álló UH- és MRI-felvételeken elhangzó bemondásokban közös halmazt alkotnak, tehát egyidejűleg kinyerhetők mindkét forrásból. Ezzel az indítatással továbbra is azt az irányvonalat szerettem volna követni, hogy összekapcsoljam az UH- és MRI-forrásból származó adatokat, amire már az 5. és 6. fejezetekben is kísérletet tettem a nyelv- és szájjpadkontúrok optimalizációja, illetve gépi betanítása révén. Ezúttal szintén a gépi tanulás eszköztárához folyamodtam, mivel a szintézishez szükséges artikulációs paramétereket geometriai adatokra alapozva, neurális hálózatok beiktatásával konstruáltam meg. A közös hangkészletet az $\{ɔ, a:, ε, e:, i, o, ø, u, y\}$ IPA szimbólumokkal ellátott $\{a, á, e, é, i, o, ö, u, ü\}$ magánhangzók csoportja alkotta. Hasonló metodikát követve, az önálló magánhangzókon túl folyamatos beszéd szintézisét is előirányoztam, amihez az MRI-felvételeken elhangzó szavakat és az UH-felvételeken rögzített mondatokat használtam fel. A szavak és mondatok listája az 1. és 2. függelékben olvasható. Összesítve tehát 9 magánhangzó, 47 szó és 40 mondat kapott szerepet a vizsgálataimban. A következő alfejezetekben ismertetem a neurális hálózathoz kapcsolódó megfontolásokat, a szintézis módszereit, illetve az eredmények kvalitatív és kvantitatív értékelését.

7.5.1. A neurális hálózatok felépítése

Neurális hálózatok alkalmazásával különböző artikulációs paraméterek betanítását valósítottam meg geometriai adatok segítségével. A magánhangzók esetében lehetőség nyílt az UH- és MRI-források összekapcsolására, így a rendszer bemeneti paramétereit az UH-keretektől, kimeneti paramétereit pedig az MRI-keretektől nyertem ki. A szavak és mondatok esetében azonban a hálózat bemeneti és kimeneti adathalmazát is ugyanabból a forrásból származtattam, tehát a szavak előállításához csak az MRI-felvételeket, a mondatok szintetizálásához pedig csak az UH-felvételeket használtam fel.

A rendszer bemeneti paramétereiként szagittális radiális távolságokat vagy DCT-együtthatókat értelmeztem. A radiális távolságok számítását a 7.4. alfejezetben részletezett módszerek alkalmazásával hajtottam végre, ami a $[-8^\circ, 32^\circ]$ UH-szögtartományban keretenként 6, a $[-36^\circ, 60^\circ]$ MRI-szögtartományban pedig keretenként 13 távolságadatot definiál. A DCT-együtthatókat a nyelvkontúrokból vontam ki a 6.2.2. alfejezetben bevezetett ismereteknek megfelelően, és figyelembe véve a 6.4.1.2., 6.4.3., illetve a 6.4.4. alfejezetekben kikristályosodott eredményeket, a DCT-együtthatók számát minden esetben 9-re állítottam be. A magánhangzók tekintetében szeretném megjegyezni, hogy miután az említett közös halmazban szereplő hangok több szóban is megjelennek a 40 különböző mondatot tartalmazó UH-csomagban, így egy adott magánhangzóhoz akár több száz képkocka, illetve nyelvkontúr tartozhat, melyek alaki szempontból igen nagy változatosságot mutathatnak. Ezért, hogy egységes és konzisztens legyen a távolságadatok kinyerése, minden magánhangzó esetében a 6.4.2. alfejezet lépései szerint produkált és a 6.7. táblázat által összefoglalt redukált UH-nyelvkontúrok átlagaként meghatározott görbével dolgoztam.

A rendszer kimeneti paramétereit a 7.2. alfejezetben tárgyalt lineáris predikció elvén keresztül a 7.1. alfejezetben bemutatott akusztikus csőmodellhez kapcsolódó mennyiségek formájában vettem fel úgy, hogy az adott bemondás akusztikus jeléből az 'lpc' függvény felhasználásával kiszámítottam az LPC-együtthatókat, melyekből az 'lpcar2rf', majd az

'lpcrf2aa' függvények alkalmazásával meghatároztam a reflexiós tényezők, illetve a vokális traktus keresztmetszeteinek halmazát. Megjegyzem, hogy az 'lpc' függvény közvetlenül a (7.3) kifejezésben foglalt β_i együtthatókat adja meg, ahol $\beta_0 = 1$, és az $\alpha_i = -\beta_i$ ($i = 1, 2, \dots, p$) transzformációval juthatunk el az α_i koefficiensekhez. A lineáris predikció és az akusztikus csőmodell szoros kapcsolatban áll egymással, hiszen a predikció p fokszáma megegyezik az akusztikus csőmodell keretében a vokális traktus felosztására kreált elemi csőszakaszok N számával. Ennélfogva az $N + 1$ darab β_i együttható N számú α_i faktort, azaz N LPC-együtthatót és ezzel együtt N csőszakaszt eredményez, ami $N + 1$ reflexiós tényezőt és $N + 2$ keresztmetszetet deklarál. Az akusztikus csőmodell alkalmazásakor kulcsfontosságú tényező a csőszakaszok N számának helyes megválasztása. Ennek a fizikai jellemzőnek az értéke ugyanis nagymértékben függ a beszédjel f_s mintavételi frekvenciájától, ami a folytonos akusztikus jelből vett diszkrét minták időegységenkénti számát definiálja. Ebből kiindulva a

$$T_s = \frac{1}{f_s} \quad (7.5)$$

reláció által meghatározható a mintavételi periódusidő, valamint a csőszakaszok maximális

$$l = v T_s \quad (7.6)$$

hossza, ahol $v = 343,14$ m/s a levegőben mért hangsebesség értéke. Mindezek ismeretében a csőszakaszok N számát a vokális traktus teljes L hosszának figyelembevételével az

$$N = \frac{L}{l} \quad (7.7)$$

összefüggés alapján kaphatjuk meg. L átlagos értéke férfiak esetében 17 cm, nők esetében pedig 14 cm. A (7.7) formula a modellt felépítő csőszakaszok számára vonatkozóan egy alsó korlátot szab meg, tehát akkor járunk el korrekt módon, ha a csőszakaszok számát legalább olyan értékre rögzítjük, amit (7.7) diktál. A rendelkezésemre álló UH- és MRI-felvételek mintavételi frekvenciája $f_s = 22050$ Hz, ami a (7.5)-(7.7) kifejezések értelmében a csőszakaszok minimális számára az UH-keretekre az $N = 8,996$, az MRI-keretekre pedig az $N = 10,924$ értéket írja elő, vagyis legalább 11 csőszakasz felvétele indokolt a vokális traktus mentén. Erre hivatkozva, a felbontás finomításának érdekében 17 csőszakaszt értelmeztem, és ezzel összhangban az LPC-együtthatók számát is 17-re állítottam be. Ez tehát keretenként 17 α_i és 18 β_i koefficiens definiál, melyekhez 18 reflexiós tényező és 19 keresztmetszet tartozik.

A neurális hálózat $m \times n$ -es mátrixok formájában kezelt bemeneti és kimeneti blokkjainak dimenzióit az összes rendszerbeállítás esetében a 7.3. táblázat foglalja össze. Az m sorindex a vizsgálatban részt vevő magánhangzók, szavak vagy mondatok számát adja meg, míg az n oszlopindex a felvett radiális távolságok, DCT-együtthatók, reflexiós tényezők vagy keresztmetszetek számát fejezi ki. Ahogyan azt a 7.3. táblázat is tükrözi, a magánhangzók esetében nem elemeztem a DCT-együtthatók által kínált opciót. Ennek az az oka, hogy a jelenleg használt algoritmusok a DCT-együtthatókat csak az egyes keretekhez illesztett, egyedi nyelvkontúrok esetében képesek előállítani, ennélfogva a fentebb említett átlagolt görbe szintjén technikailag nem elérhetők a DCT-hez kapcsolódó számítások.

7.3. táblázat. A neurális hálózat bemeneti és kimeneti adathalmazainak dimenziói az összes rendszerkonfiguráció esetében.

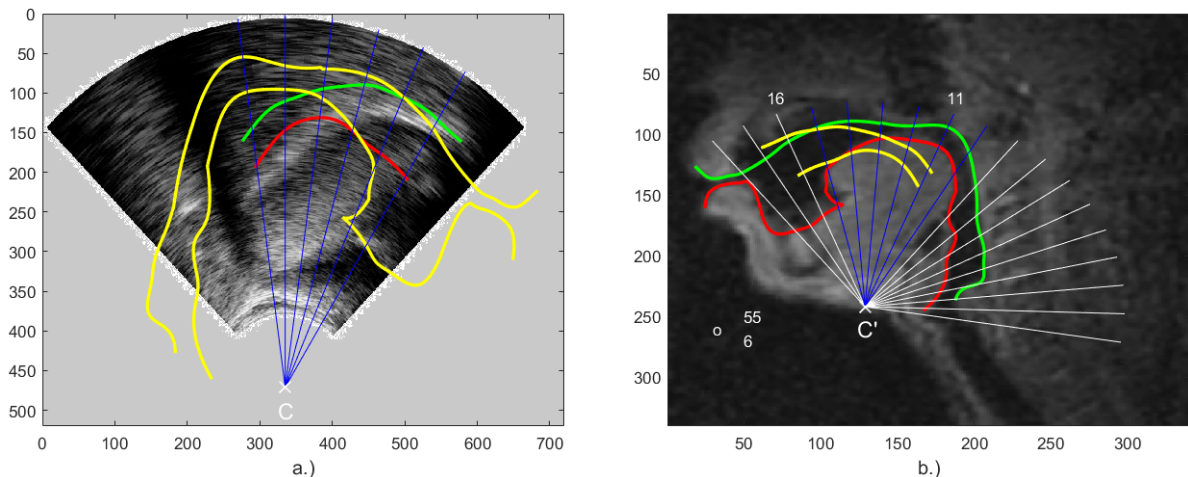
	A bemenet mérete		A kimenet mérete	
	Radiális távolság	DCT-koefficiens	Reflexiós tényező	Keresztmetszet
Magánhangzók	9×6	–	9×18	9×19
Szavak	47×13	47×9	47×18	47×19
Mondatok	40×6	40×9	40×18	40×19

A 7.3. táblázatban feltüntetett paramétereket a magánhangzók, illetve a szavak és mondatok esetében eltérő technikával vontam ki a beszédjelből. A magánhangzók halmozásában úgy jártam el, hogy miután a kitartott beszédhangok képzésekor lényegében statikus a hangképző szervek helyzete, így a rendszer bemeneti oldalán a radiális távolságok kiolvasásakor elegendő volt hangonként egyetlen keretet felhasználni, amelyen az adott hanghoz rendelt, korábban említett átlagolt nyelvkontúr és a szájpadkontúr szerepelt. A kimeneti adatok kinyeréséhez pedig az adott magánhangzóhoz tartozó hangminták középső elemének szimmetrikus környezetében kivágtam egy olyan hosszúságú részletet, amiben hallás után már felismerhető a hang, és az így létrejövő magánhangzókernelből származtattam a reflexiós tényezőket és keresztmetszeteket. A szavak és mondatok esetében komplikáltabb volt a paraméterek korrekt megállapítása, mivel, folyamatos beszédről lévén szó, az egymáshoz kapcsolódó beszédhangok láncolata mentén haladva dinamikusan változik a vokális traktus alakja, és ezt a dinamikát a paraméterhalmaznak is tükröznie kell. Ezért a bemeneti adatok számításakor az adott bemondás összes keretét figyelembe kellett vennem úgy, hogy minden egyes kerethez hozzárendeltem az aktuális nyelv- és szájpadkontúr alkalmazásával kapott radiális távolságokat és DCT-együtthatókat. A kimeneti adatok kivonásakor pedig az adott bemondás teljes mintaszámát elosztva az összes keret számával, meghatároztam a keretenkénti minták számát, így – a bemeneti oldalon érvényes módszerhez hasonlóan – ez esetben is minden egyes kerethez társítottam az aktuális hangminták csoportjából eredeztetett reflexiós tényezőket és keresztmetszeteket. Ennek az lett az eredménye, hogy a hálózat bementi és kimeneti tömbjeinek első dimenziója a szavak esetében 47-ről 4 354-re, a mondatok esetében pedig 40-ről 6 278-ra duzzadt.

A tanulóalgoritmus konstrukciójakor ezúttal is a 6.3. alfejezetben kifejtett technikát, azaz az SCG módszert alkalmaztam a kimeneti paraméterek betanítására. A rendszerben egy rejtett réteget helyeztem el, melybe a magánhangzók esetében 10, a szavak és mondatok esetében pedig 100 neuront ültettem be. A hálózat kimeneti rétegének aktivációs függvényét a kimeneti paraméter típusához igazodva választottam meg. Mivel a reflexiós tényezők értéke lehet pozitív és negatív is, esetükben megtartottam a standard lineáris átvitelt, a keresztmetszetek tanításakor azonban a hiperbolikus tangens szigmoid függvény mellett döntöttem, kizárva ezzel a negatív keresztmetszetek kialakulásának lehetőségét.

7.5.2. A neurális hálózat bemeneti és kimeneti paramétereinek kapcsolata UH-MRI viszonylatban

Úgy ítéltém meg, hogy a magánhangzók betanításakor kombinált UH- és MRI-környezetek összehasonlítása céljából érdemes tanulmányozni a megszerkesztett neurális hálózat bemeneti és kimeneti adatai között fennálló kapcsolatot, melynek megértését a 7.11. ábra segíti az *o* hangot reprezentáló kontúrvonalak feltüntetésével.



7.11. ábra. Az UH- és MRI-környezet eredeti (zöld és piros) és transzformált (sárga) szájpap- és nyelvkontúrjainak relatív elhelyezkedése a megfelelő radiális metszővonalak által kijelölt tartományokban.
(Az ábra tengelyeinek mértékegysége pixel.)

A 7.11.a ábra nyugalmi nyelválláshoz tartozó UH-keretén zölddel van megrajzolva a szájpap kontúrja, a redukált nyelvkontúrok átlagaként kapott görbe pedig piros színben jelenik meg. Az (5.1) által rögzített koordinátákkal rendelkező C pontból indított, kékkel megjelölt sugárirányú metszővonalak a 7.6.a ábrához hasonlóan továbbra is azokat a radiális irányokat definiálják, melyek mentén a görbepár közötti 6 távolságvértéket megmérjük. Elemzéseim során arra a kérdésre kerestem a választ, hogy ez a 6 radiális metszet geometriailag hogyan viszonyul a 7.11.b ábra által vizualizált MRI-környezethez, ahol a kiterjesztett szájpap-, illetve nyelvkontúrt a zöld, illetve piros görbék érzékeltetik. Ennek érdekében, igazodva az alkalmazott radiális geometriához, a gégtől a szájnylásig terjedő tartományban sugárirányú metszővonalak mint osztóvonalak segítségével vettem fel a korábban említett 17 csőszakaszt. A fehérrel kiemelt radiális metszetek a $[-36^\circ, 100^\circ]$ intervallummal behatárolt szögtartományt fedik le, és – biztosítva a vizsgálatok koherenciáját – ugyanabból a C' pontból erednek, mint amit az 5.5.4. alfejezetben taglalt optimalizációs eljárás végeredményeként kaptam az 5.8. táblázat $(C'_x; C'_y)$ koordinátái szerint. Ezek után az 5.8. táblázat R nagyításának és Φ_0 szögelfordulásának beállításával végrehajtottam a 7.11.a ábra nyelv- és szájpapkontúrjainak az (5.6) műveletekre alapozott, MRI-környezetbe történő transzformációját, amit a 7.11.b ábra sárga görbepárja realizál. Megfigyelhető, hogy a görbepár az MRI-geometriába való beágyazódása révén a 11. és 16. radiális metszetek által kihasított köröket foglalja el, amint azt a kékkel megkülönböztetett metszővonalak csoportja is szemlélteti. Az UH-kereteken felvett nyelv- és szájpapkontúroknek a fentebbi technikával megvalósított transzformációja a vizsgálatban részt vevő többi magánhangzó esetében is ugyanerre az elrendezésre vezetett, azaz a

transzformált görbepár minden esetben a 11. és 16. radiális metszetek között kiterjedő zónába kerül. Ez a körcikk a 17 csőszakaszból 5 csőszakaszt fed le, ami az 5:17 arány értelmében azt jelenti, hogy az elérhető részleges UH-tartomány a teljes vokális traktust felölelő MRI-tartománynak megközelítőleg 29%-át teszi ki. A 7.11.b ábra nyelv- és szájpaddockontúrjainak az 5.8. táblázat paramétereire, valamint az (5.7) operációkra épülő, UH-környezetbe történő transzformációját is megvalósítottam, melynek eredményét a 7.11.a ábra sárga görbepárja demonstrálja.

7.5.3. A beszédszintézis módszerei

A gépi beszéd szintetizálását kétféle megközelítésben vittem véghez. Elsőként a 7.1. alfejezetben vázolt akusztikus csőmodell, majd pedig a 7.2. alfejezetben prezentált lineáris predikció felhasználásával kíséreltem meg a gépi hangok előállítását. A két módszer lényegi aspektusait a következő paragrafusokban foglalom össze.

7.5.3.1. Az akusztikus csőmodellen alapuló szintézis [69] ♦ Az akusztikus csőmodell keretében a vokális traktusban terjedő hanghullámot speciális fizikai mennyiségekkel jellemezhetjük, melyek közül programozástechnikai szempontból az $u(x, t)$ térfogatáram a legalkalmasabb a folyamatok modellezésére. A kétváltozós térfogatáram a vokális traktus tengelye mentén mért x pozíciótól és a t időtől függ, és megmutatja az artikulációs csatorna adott keresztmetszetén egységnyi idő alatt átáramló levegő térfogatának mértékét. A térfogatáramot vektorfizikai mennyiségként kezeljük, mivel a nagyságán túl az iránya is mérhető. A pozitív, illetve negatív x irányba terjedő térfogatáramot az $u^+(x, t)$, illetve az $u^-(x, t)$ szimbólumokkal láthatjuk el. A modell hipotéziseinek betartásával levezethető egy olyan komplett egyenletrendszer, mely egyértelműen összekapcsolja a szomszédos csőszakaszok határánál létrejövő térfogatáramokat, méghozzá oly módon, hogy a vokális traktus három sajátos tartományát külön-külön kezelve, eltérő szerkezetű egyenletcsoport alakul ki a gégénél található gerjesztési oldalon, a közbenső csőszakaszok régiójában, valamint a szájnylásnál lévő sugárzási oldalon. A keletkező egyenletrendszer az

$$u_1^+(t) = \frac{1+r_g}{2} u_g(t) + r_g u_1^-(t) \quad (a)$$

$$\begin{aligned} u_{k+1}^+(t) &= (1+r_k) u_k^+(t - \tau_k) + r_k u_{k+1}^-(t) \\ u_k^-(t + \tau_k) &= -r_k u_k^+(t - \tau_k) + (1-r_k) u_{k+1}^-(t) \end{aligned} \quad (b)$$

$$\begin{aligned} u_N^-(t + \tau_N) &= -r_s u_N^+(t - \tau_N) \\ u_s(t) &= (1+r_s) u_N^+(t - \tau_N) \end{aligned} \quad (c) \quad (7.8)$$

alakot ölti, ahol az (a), (b), (c) blokkok rendre a gerjesztési oldal, a közbenső csőszakaszok, illetve a sugárzási oldal térfogatáram-viszonyait jellemzik. A (7.8) egyenletrendszer tömörített jelölésformái szerint a k -adik csőszakasz elején pozitív vagy negatív irányba haladó térfogatáram az

$$u^\pm(x_k, t) = u_k^\pm(t), \quad (7.9)$$

míg a k -adik csőszakasz végénél pozitív vagy negatív irányba haladó térfogatáram az

$$u^\pm(x_k + l_k, t) = u_k^\pm(t \mp \tau_k) \quad (7.10)$$

szimbólumoknak megfelelően egyszerűsödik, $k = 1, 2, \dots, N$ indexelés mellett. A (7.9) és (7.10) egyenlőségek bal oldali argumentumában szereplő helykoordinátát tehát a jobb oldali alsó index váltja fel, a (7.10)-ben megjelenő τ_k paraméter pedig azt az időeltolódást határozza meg, ami ahhoz szükséges, hogy a $v = 343,14$ m/s sebességgel terjedő hanghullám áthaladjon az l_k hosszúságú csőszakaszon. A (7.8.a) és (7.8.c) egyenletekben szereplő $u_g(t)$ és $u_s(t)$ függvények speciálisan a gége által indukált gerjesztőjelet, illetve a szájnýílán keresztül a szabad térbe sugárzott beszédjelet testesítik meg, az r_g, r_k, r_s faktorok pedig a 7.1. ábra szerint bevezetett reflexiós tényezőkkel azonosíthatók. A gépi beszéd produkciója a (7.8) egyenletrendszer programszintű ciklikus megvalósításával lehetséges, melynek során döntő szerepe van az $u_g(t)$ gerjesztőjelnek, hiszen a jel alakja alapvetően befolyásolja a szintézis eredményét. A gerjesztőjel jellegét természetesen az előállítandó hang típusa határozza meg attól függően, hogy zöngés vagy zöngétlen módon képződik. Zöngés gerjesztés esetén a gerjesztőjel periodikusan ismétlődő impulzusok formájában jön létre, zöngétlen gerjesztés esetén pedig véletlenszerűen változik. Az egymást követő beszédhangok típusának véletlenszerűségéhez igazodó zöngés-zöngétlen fluktuáció legegyszerűbben úgy indukálható, hogy gerjesztőjelként a lineáris predikció keretében értelmezett (7.3) hibajelet alkalmazzuk. A csőmodell megvalósításához szükséges reflexiós tényezőket részben közvetlenül a neurális hálózat kimeneti csatornájából nyertem eleve betanított paraméterek formájában, másrészt pedig a betanított keresztmetszeteket átalakítottam reflexiós tényezőkké az 'lpcaa2rf' függvény segítségével.

7.5.3.2. A lineáris predikció elvén alapuló szintézis [69] ♦ A predikciós elvű beszéd szintézis végrehajtása a 7.2. és 7.3. ábrákon vázolt analízis- és szintézismodellek programszintű megalkotása révén lehetséges. Ehhez mindkét modell esetében ismernünk kell a teljes rendszer átviteli függvényét. Az analízismodell szerint megszerkesztett rendszer átviteli függvényének értelmezéséhez elsőként képeznünk kell (7.3) első egyenlőségének diszkrét idejű Laplace-transzformáltját, ami a

$$H(z) = S(z) - \bar{S}(z) \quad (7.11)$$

kifejezésre vezet. Felismerve, hogy a (7.4) által bevezetett átviteli függvény a 7.2. ábra szaggatott vonallal határolt belső blokkjához csatlakozó kimeneti és bemeneti $\bar{s}(n)$ és $s(n)$ jelek diszkrét idejű Laplace-transzformáltjainak hányadosaként definiálható

$$F(z) = \frac{\bar{S}(z)}{S(z)} \quad (7.12)$$

szerint, a (7.11) formula átírható a

$$H(z) = S(z)[1 - F(z)] \quad (7.13)$$

alakra. A (7.13) egyenlet

$$A(z) = 1 - F(z) = \frac{H(z)}{S(z)} \quad (7.14)$$

típusú átrendezése arra a felismerésre vezet, hogy a $h(n)$ predikciós hiba egy $A(z)$ átviteli függvénnyel rendelkező, $s(n)$ jellel gerjesztett rendszer válaszaként kezelhető, ami pontosan megfelel a 7.2. ábra által teremtett körülményeknek. A szintézismodell keretében

felépített rendszer átviteli függvényét a (7.14) egyenlet

$$B(z) = \frac{1}{A(z)} = \frac{1}{1 - F(z)} = \frac{S(z)}{H(z)} \quad (7.15)$$

szerinti transzformációjával nyerhetjük, ami azt mutatja, hogy az $s(n)$ jel egy $B(z)$ átviteli függvénnyel leírható rendszer válaszként detektálható, amennyiben a $h(n)$ hibafüggvény gerjesztőjelként szolgál, egyetértésben a 7.3. ábra szerkezetével. A (7.4) összefüggés figyelembevételével tehát az analízismodellben az

$$A(z) = 1 - \sum_{i=1}^p \alpha_i z^{-i} = \sum_{i=0}^p \beta_i z^{-i}, \quad (7.16)$$

a szintézismodellben pedig a

$$B(z) = \frac{1}{1 - \sum_{i=1}^p \alpha_i z^{-i}} = \frac{1}{\sum_{i=0}^p \beta_i z^{-i}} \quad (7.17)$$

alakú átviteli függvény releváns. A (7.16)-(7.17) átviteli függvényeket alapul véve, az analízis- és szintézismodell kimeneti jelei a 'filter' függvény segítségével generálhatók 'filter(a, b, x)' struktúrával, ahol az x -szel jelölt bemeneti adatokat a szűrésnek alávetett szintetizálendő hangminták vagy a hibajel alkotják, az a és b paraméterek pedig a vokális traktus átviteli függvényének számlálójában és nevezőjében megjelenő koefficienseket adják meg. Ennek megfelelően, a (7.16)-(7.17) kifejezésekhez igazodva, az analízismodell hibajelét az $a = \beta_i$ és $b = 1$, a szintézismodellrel produkált beszédjelet pedig az $a = 1$ és $b = \beta_i$ együtthatók beállításával biztosíthatjuk. A hibajel előállításához szükséges β_i LPC-együtthatókat az 'lpc' függvény alkalmazásával vontam ki az eredeti beszédjelből. A hangminták szintetizálásában részt vevő β_i LPC-együtthatókat pedig a betanított reflexiós tényezőkből, illetve keresztmetszetekből származtattam az 'lpcrf2ar', illetve az 'lpcaa2rf' függvények felhasználásával.

7.5.4. A beszéd szintézis eredményei

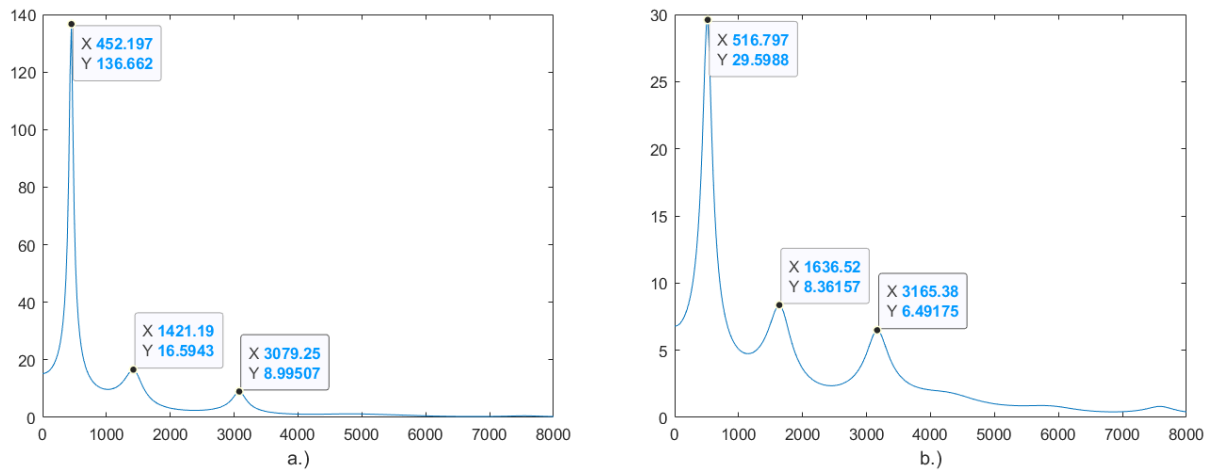
A 7.5.3.1. és a 7.5.3.2. paragrafusokban ismertetett stratégiákat követve és a 7.5.1. alfejezet szerint megvalósított gépi tanítás eredményeire támaszkodva végrehajtottam az 1. és 2. függelékben csatolt magánhangzók, szavak és mondatok szintézisét. Az eredeti és a szintetizált beszédjelek elérhetők és meghallgathatók az alábbi linken:

https://drive.google.com/drive/folders/1qbTEuNdl5T_pn2FYY-b5vyFMDTLjITyX?usp=sharing

A 'HANGOK', 'SZAVAK' és 'MONDATOK' főmappákban található almappák elnevezése minden esetben 'X_Y_Z' szerkezetű, ahol 'X' a szintézis során alkalmazott módszert rejti, 'Y' és 'Z' pedig a neurális hálózat bemeneti és kimeneti paramétereit egyértelműsíti, tehát az 'X_Y_Z' címke 'Y' adatokkal táplált és 'Z' adatokat betanuló neurális hálózat kimeneti eredményeit felhasználó, 'X' módszerrel realizált szintézist takar. Ezzel összhangban az 'X' helyére 'cso' vagy 'lpc' kerül, ami az akusztikus csőmodellt vagy a lineáris predikció elvét jelzi. Az 'Y' pozícióban 'tav' vagy 'det' feliratok lehetségesek, melyek a radiális távolságokat vagy a DCT-együtthatókat jelölik. A 'Z' elem 'ref' vagy 'ker' alakú, utalva a reflexiós tényezőkre vagy a keresztmetszetekre. Ezek mellett mindhárom főmappa tartalmazza az eredeti bemondások audiofelvételeit is.

Az eredményeim elemzése során arra a következtetésre jutottam, hogy a beszédszintézis minden esetben sikerrel zárult, mivel kvalitatív és kvantitatív szinten is relatíve jó egyezést tapasztaltam az eredeti és a szintetizált bemondások között.

A kvalitatív értékelés során azt állapítottam meg, hogy a szintetizált bemondások mindegyike elég jól érthető és a beszéd felismerhető az eredeti hanginformáció nélkül is, bár megjegyzem, hogy a szintetizált jelekben zajkomponensek is tapasztalhatók, ami valamelyest rontja az akusztikai élményt, de természetesen a kutatómunkám későbbi fázisaiban szeretném majd a torzításokat a lehető legjobb mértékben kiküszöbölni és javítani a beszéd minőségén.



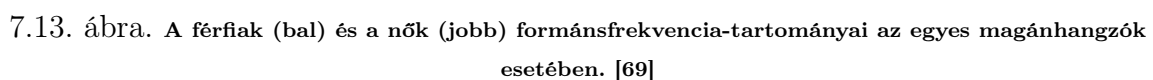
7.12. ábra. Az eredeti (a.) és a szintetizált (b.) beszédhang esetében előálló átviteli karakterisztikák az $\bar{o}$ hang példájában a lokális maximumok helyein azonosítható formánsfrekvenciák értékeivel. (Az ábra vízszintes tengelyeinek mértékegysége Hz, függőleges tengelyeinek nincs mértékegysége.)

Kvantitatív analízis céljából a magánhangzók esetében összehasonlítottam az eredeti és a szintetizált hangokat jellemző formánsfrekvenciákat. A formánsok a zöngés beszédhangok artikulációja során képződnek, amikor is a hangszalagok kváziperiodikus rezgésének eredményeképpen létrejövő zöngé rezgései gerjesztik a vokális traktust. A gerjesztés folytán többféle frekvenciakomponens jelenik meg a hangban, kialakítva az alapharmonikust és a felharmonikusokat. Bizonyos felharmonikusok meghatározott rezonanciafrekvenciák környékén felerősödnek, és létrehozzák a formánsokat, amik egyedi módon jellemzik az adott hangot. A vizsgálathoz a 'freqz(a,b)' függvény segítségével meghatároztam a vokális traktus mint szűrő frekvenciaválasztát, ami megegyezik a rendszer átviteli karakterisztikájával. A (7.17) átviteli függvény struktúrájához visszatérve, az $a = 1$ és $b = \beta_i$ beállítás a számláló és nevező koeficienseit rögzíti. A b paraméter helyére először az eredeti, majd pedig a szintetizált beszédjelhez tartozó β_i LPC-együtthatókat helyettesítve, eljutottam az átviteli karakterisztikákhoz. Egy szűrő frekvenciaválasza információt ad a rendszer frekvenciatartománybeli áteresztőképességéről, megmutatva a szűrő erősítését különböző frekvenciákon, így a kapott görbe lokális maximumainál egyértelműen megtaláljuk a formánsfrekvenciák értékeit. Az eredeti és a szintetizált beszédhang esetében előálló átviteli karakterisztikákat szemléltet a 7.12. ábra az $\bar{o}$ hang példáján keresztül. Az ábrák függőleges tengelyein az erősítést, a vízszintes tengelyein pedig a frekvenciát mérjük, ennél fogva a görbéken megjelölt pontok X koordinátájából olvashatjuk ki az első három formánsfrekvenciát. A 7.12. ábrán vázolt grafikus megközelítést alkalmazva, a frekvenciaválasz-görbék

7.4. táblázat. Az eredeti és a reflexiós tényezők alapján szintetizált magánhangzók formánsfrekvenciái.

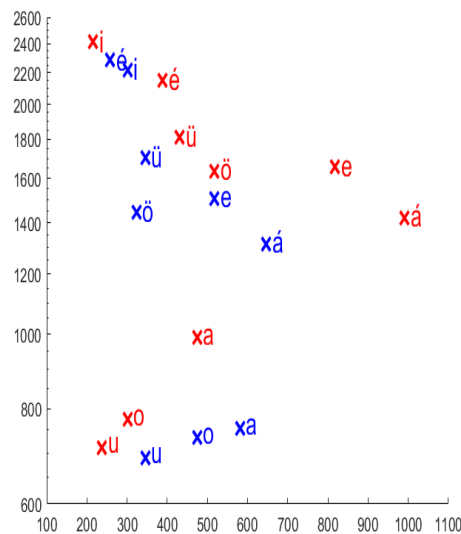
7.5. táblázat. Az eredeti és a keresztmetszetek alapján szintetizált magánhangzók formánsfrekvenciái.

A táblázatok adatai alapján megállapítható, hogy az eredeti és a szintetizált hangok formánsfrekvenciái között létrejött különbségek a 0-302 Hz intervallumba esnek. Ezen eltérések mértékének megítélése céljából eredményeimet összehasonlítottam a 7.13. ábrán látható formánstérképpel is, ami kétdimenziós vetület formájában ábrázolja az egyes magánhangzókat általánosan jellemző formánsfrekvencia-tartományokat.



90

érvényes értékhalmozokat tünteti fel. Az MRI-felvételek férfi bemondójához igazodva, vizsgálataimban a bal oldali rajz volt a mérvadó, és arra a következtetésre jutottam, hogy a 7.4. és 7.5. táblázatok frekvenciái javarészt beilleszthetők a 7.13. ábra által vizualizált intervallumokba. Ez alól csak néhány adat képez kivételt, amit pirossal kiemeltem a táblázatokban. Emellett a 7.4. és 7.5. táblázatok szintetizált hangjainak formánsfrekvenciáit grafikusán is ábrázoltam, ami a 7.14. ábrán követhető nyomon. A vízszintes tengelyen az F1, a függőleges tengelyen pedig az F2 formánsfrekvenciát vettem fel Hz-ben. A piros értékhalmoz a reflexiós tényezők, a kék értékhalmoz pedig a keresztmetszetek alapján produkált beszédhangokhoz tartozik. A 7.13. és 7.14. ábrák összevetése azt mutatja, hogy a szintetizált magánhangzók formánsfrekvenciáinak relatív elrendeződése többnyire megtartja az elvárt formánstérkép struktúráját, ami alátámasztja a szintézis sikerességét.



7.14. ábra. A reflexiós tényezők (piros), illetve a keresztmetszetek (kék) alapján produkált beszédhangok formánstérképei. (Az ábra tengelyeinek mértékegysége Hz.)

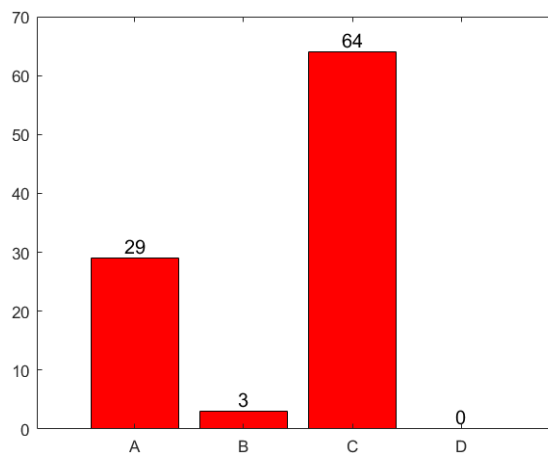
A szavak és mondatok esetében kapott eredményeim értékeléséhez néhány szintetizált bemondást szubjektív audioteszt formájában véleményezésre bocsátottam egy erre a feladatra felkért, sok résztvevős célcsoport körében, akik független minősítőként semmilyen módon nem kapcsolódtak a kutatáshoz. A felmérés célja az volt, hogy a szubjektív audioteszt kimenetelének ismeretében egyértelműen állást lehessen foglalni, hogy a 7.3. táblázatban rendszerezett konfigurációk közül melyik bizonyul a leghitelesebbnek a beszéd-szintézisben, azaz melyik neurális hálózat beállításával lehet a legjobb minőségű beszédet előállítani. Ezzel a vizsgálattal nem a modellszintű hatékonyságot szerettem volna tesztelni, ezért az akusztikus csőmodell és a lineáris predikció révén kapott hangminták összehasonlítására irányuló kísérlet nem történt. Számomra inkább az volt a legfőbb eldöntendő kérdés, hogy a 7.3. táblázat mely paraméterkombinációjának alkalmazásával érhető el a legtisztábban érthető gépi beszéd. A szubjektív minősítés megvalósításához kiválasztottam egy lineáris predikció szerint szintetizált mondatot a 2. függelékben felsorolt bemondások közül, és azt négy különböző változatban tártam a célcsoport elé. A tesztelendő bemondás típusát alapvetően két szempont alapján választottam ki. Egyrészt a magánhangzók, szavak és mondatok csoportjait tekintve a mondatok adják a leghosszabb idejű bemondásokat, így talán ezeket könnyebb hallás után megítélni, mint a magánhangzókat és szavakat megszó-

laltató, rövidebb időtartamú audiofelvételeket. Másrészt pedig miután a célcsoport minden tagja magyar anyanyelvű, így a hallás utáni szövegértés megbízhatóbb eredményeket ígérhet magyar nyelvű bemondások próbájakor. A kiválasztott mondat A, B, C, D címkékkel ellátott négy verzióját a 7.6. táblázat sűríti, ahol azonosítható a szintézis alapjául szolgáló neurális hálózat bemeneti és kimeneti paramétertípusainak összekapcsolása. A minősítésben részt vevő személyeket nem szerettem volna semmilyen módon befolyásolni, így nem hoztam a tudomásukra, hogy az A, B, C, D felvételekhez milyen paraméterek vannak hozzárendelve, mindössze annyit közöltem velük, hogy a négy mondatot négy különböző módszer szerint generáltam, és pusztán az auditív percepció alapján kellett eldönteniük, hogy melyik esetben érthető legjobban a beszédjel.

7.6. táblázat. A szubjektív tesztelésre bocsátott szintetizált audiofelvétel négy különböző változata a neurális hálózat paramétereinek függvényében.

	Bemeneti paraméter	Kimeneti paraméter
A	Radiális távolság	Reflexiós tényező
B	Radiális távolság	Keresztmetszet
C	DCT-koefficiens	Reflexiós tényező
D	DCT-koefficiens	Keresztmetszet

A szubjektív értékeléshez összesen 96 alany csatlakozott. Véleményük megoszlását a 7.15. ábra szemlélteti, ahol megfigyelhető, hogy 3 résztvevő kivételével senki nem voksolt a B és D variánsokra, a szavazatok lényegében az A és C verziók szintjén differenciálódnak többségében C szerinti állásfoglalással a 29:64 arálynak megfelelően, vagyis az A-hoz viszonyítva kb. kétszer annyian jelölték be a C-t. Ez az eredmény tehát arra enged következtetni, hogy a keresztmetszetekkel szemben a reflexiós tényezők betanításával jobb minőségű gépi beszéd állítható elő, és a tanítóalakzatok szintjén a radiális távolságokkal szemben előnyt élveznek a DCT-koefficiensek.



7.15. ábra. Az auditív percepción alapuló szubjektív tesztben részt vevők szavazatainak megoszlása négy különböző módon szintetizált mondat (A, B, C, D) próbájakor.

7.6. T3 tézis

A kétdimenziós UH- és MRI-felvételekből kiindulva artikulációs beszéd-szintézis valósítható meg mesterséges intelligencia felhasználásával. Szagittális radiális távolságokra, valamint DCT-együtthatókra alapozva neurális hálózatok segítségével betaníthatók a beszédjel artikulációs paramétereiként kezelhető reflexiós tényezők, illetve a vokális traktusbeli keresztmetszetek, melyekből az akusztikus csőmodell és a lineáris predikciós kódolás szerint is rekonstruálható az eredeti beszédjel. A szintézis során a keresztmetszetekkel szemben a reflexiós tényezők betanításával jobb minőségű gépi beszéd állítható elő, és a tanítóalakzatok szintjén a radiális távolságokkal szemben előnyt élveznek a DCT-koefficiensek. (Kapcsolódó saját publikáció: [S12])

8. A tézisek korlátai, érvényessége és újdonságai

Korlátok: A vizsgálataim során beszélőfüggő eredmények születtek, hiszen az UH- és MRI-források esetében is egy-egy adatközlő felvételeivel dolgoztam, akik nemüket és anyanyelvüket tekintve is különbözőek voltak, ezért a következtetések több beszélő vagy ugyanazon személy UH- és MRI-felvételeinek esetében változhatnak.

Érvényesség: A következtetések érvényességét abban az esetben lehetne általánosan kiterjeszteni, ha több beszélő vagy ugyanazon személy UH- és MRI-felvételeit is bevonnánk az elemzésekbe.

Újdonság: Legjobb tudomásom szerint a szakirodalom nem tartalmaz olyan közleményt, melyek az UH-MRI fúziót anatómiai kontúrvonalak optimalizációs algoritmusokra vagy gépi tanulásra támaszkodó kombinálásával valósítanák meg.

9. Összefoglaló

Doktori értekezésemben a beszédképzés különböző aspektusait vizsgáltam MATLAB-környezetben. Elemzéseim során olyan ultrahangos (UH), illetve mágneses rezonancias képalkotási (MRI) technikával készült audiovizuális forrásokra támaszkodtam, melyek a kétdimenziós szagittális síkban vizualizálják a vokális traktus hangképző szerveinek relatív helyzetét és mozgását, miközben rögzítik a beszélő által kibocsátott audiojelet. Így a kép- és hangtartalom a szinkronizációnak köszönhetően egyértelmű módon összekapcsolódik.

Az UH- és MRI-felvételek jó alapot biztosítottak ahhoz, hogy geometriai és akusztikus paramétereket nyerhessek ki a kép- és hangforrásokból. A geometriai adatokhoz való hozzáférést nagyban megkönnyítette az automatikus kontúrkövető algoritmusok alkalmazása, melyekkel elvégeztem az UH-keretek nyelvkontúrjainak, valamint az MRI-keretek nyelv- és szájjpadkontúrjainak dinamikus letapogatását. Az így kapott görbecsoportokat kiegészítettem az UH-felvételekre rajzolt szájjpadkontúrral, amit egy általam kidolgozott eljárással konstruáltam meg. A nyelvkontúrok ismeretében a görbék mentén kijelöltem véges számú pontthalmazokat (tanítópontok), és származtattam a görbék diszkrét koszinusz-transzformációjában részt vevő együtthatókat (DCT-együtthatók). Ezek a paraméterek

a geometriai leírásban relevánsak, mivel kvantitatív adalékkal szolgálnak a nyelv szájszármagjának és alakjának jellemzésében. Emellett a nyelv- és szájszármagkontúrokból kiindulva kifejlesztettem két különböző módszert a nyelv és szájszármag anatómiai felszínei között mérhető szagittális radiális távolságok dinamikus meghatározására. Az akusztikus paramétereket a beszédjelek időfüggvényeiből eredeztettem a lineáris predikció elvéhez kapcsolódó LPC-együtthatók formájában, melyeket az akusztikus csőmodellben értelmezett reflexiós tényezőkké, illetve a vokális traktus keresztmetszeti adataivá konvertáltam.

Doktori munkám egyik legfontosabb célkitűzése az UH- és MRI-források geometriáinak összehangolása volt, amit többféle megközelítésben is megvalósítottam. Egy lehetséges irányvonalként olyan geometriai transzformációkat alkottam meg, melyek segítségével kölcsönösen egyértelműen egymásra vetíthetők az UH- és MRI-felvételek nyelv- és szájszármagkontúrjai. Az egymásnak megfeleltetett kontúrok legnagyobb mértékű átfedését a transzformációk paramétereinek optimalizációjával biztosítottam. Ezenkívül olyan gépi tanulóalgoritmusokat is létrehoztam, melyek az MRI-nyelvkontúrok betanítását végzik az UH-nyelvkontúrokra alapozva. A neurális hálózat tekintetében számos rendszerkonfigurációt alakítottam ki, változtatva a hálózat rejtett rétegeinek és neuronjainak számát, valamint bemeneti és kimeneti adattömbjeinek típusát és dimenzióját. A rendszer bemenetén és kimenetén a nyelvkontúrokból kivont tanítópontokat, illetve DCT-együtthatókat állítottam be különböző párosításokban.

A kutatómunkám másik lényegi irányvonalát az artikulációs beszéd-szintézis kivitelezése képezte, melynek során beszédhangokat, illetve folyamatos beszédet állítottam elő. A szintézis előkészítéseként olyan neurális hálózatokat szerkesztettem, amik bemeneti adatként fogadják a vokális traktus szagittális radiális távolságaival és a nyelvkontúrokból kivont DCT-együtthatókkal felépített mátrixokat, a kimeneten pedig a beszédjelből származtatott reflexiós tényezők, valamint keresztmetszetek által alkotott struktúrákat produkálnak. A betanított paraméterek felhasználásával beprogramoztam az akusztikus csőmodellt, illetve a lineáris predikció analízis- és szintézismodelljeit, melyek segítségével véghez vittem a gépi beszédprodukción.

10. Summary

In my doctoral thesis, I investigated the different aspects of speech production in MATLAB environment. During my analyses, I relied on such audiovisual sources made by ultrasound (US) and magnetic resonance imaging (MRI) technique that visualise the relative positions and movement of the vocal organs of the vocal tract in the two-dimensional sagittal plane, while they record the audio signal emitted by the speaker. So, the image and sound contents are connected due to the synchronisation in an unambiguous manner.

The US and MRI records provided a good basis for extracting geometric and acoustic parameters from the image and sound sources. The application of the automatic contour tracking algorithms facilitated the access to the geometric data, with which I performed the dynamic scanning of the tongue contours of the US frames, as well as the tongue and palate contours of the MRI frames. I supplemented the obtained groups of curves with the palate contour drawn in the US images, which I constructed by a procedure elaborated by me. Knowing the tongue contours, I designated a finite number of sets of points along the curves (feature points) and derived the coefficients taking part in the discrete cosine transform of the curves (DCT coefficients). These parameters are relevant in the

geometric description, as they serve as quantitative components in the characterisation of the oral position and shape of the tongue. Besides this, starting from the tongue and palate contours, I developed two different methods for the dynamic determination of the sagittal radial distances measured between the anatomic surfaces of the tongue and palate. I deduced the acoustic parameters from the time functions of the speech signals in the form of LPC coefficients connected to the principle of linear prediction, which I converted to reflection coefficients interpreted in the acoustic tube model and to cross-sectional data of the vocal tract.

One of the most important goal of my doctoral work was the harmonisation of the geometries of the US and MRI sources, which I realised in various approaches. As a possible line, I created such geometric transformations by dint of which the tongue and palate contours of the US and MRI images can be projected onto each other mutually unambiguously. I ensured the greatest extent of overlap between the corresponding contours by the optimisation of the parameters of the transformations. In addition, I generated such machine learning algorithms, as well, which make the training of MRI tongue contours based on US tongue contours. Regarding the neural network, I formed several system configurations, modifying the number of hidden layers and neurons, as well as the type and dimension of the input and output array of data of the network. At the input and output of the system, I set feature points and DCT coefficients gained from the tongue contours in different pairings.

The other essential line of my research work was given by the implementation of articulatory speech synthesis, during which I generated speech sounds and continuous speech. As a preparation of the synthesis, I designed such neural networks that receive matrices built up by sagittal radial distances of the vocal tract and DCT coefficients extracted from the tongue contours as input data and produce structures created by reflection coefficients arising from the speech signal and cross-sections at the output. Using the trained parameters, I programmed the acoustic tube model and the analysis and synthesis models of linear prediction, by dint of which I accomplished machine speech production.

Köszönetnyilvánítás

Őszinte hálámat fejezem ki a Témavezetőmnek, Prof. Dr. Czap László Tanár Úrnak, hogy értékes és inspiráló szakmai tanácsaival, tiszteletreméltó emberi magatartással, példamutató önzetlenséggel és végtelen türelemmel segítette a kutatómunkám minden fázisát és a doktori disszertációm elkészítését.

Köszönetet mondok az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoportjának is, hogy rendelkezésemre bocsátották a Micro rendszerrel készült ultrahangfelvételeket.

Hivatkozások

- [1] N. Kaur, P. Singh, *Speech waveform reconstruction from speech parameters for an effective text to speech synthesis system using minimum phase harmonic sinusoidal model for Punjabi*, Multimedia Tools and Applications, 81(18), 26101-26120, 2022.
- [2] S. P. Panda, A. K. Nayak, *A waveform concatenation technique for text-to-speech synthesis*, International Journal of Speech Technology, 20, 959-976, 2017.
- [3] R. Carlson, B. Granström, *Rule-based speech synthesis*, Springer Handbook of Speech Processing, 429-436, 2008.
- [4] R. Mahum, A. Irtaza, A. Javed, *Text to speech synthesis using deep learning*, In Intelligent Multimedia Signal Processing for Smart Ecosystems, Cham: Springer International Publishing, 289-305, 2023.
- [5] G. Olaszy, G. Gordos, *Automatic text-to-speech system applied in a reading machine*, In Proceedings of European Conference on Speech Technology (EUROSPEECH '87), Edinburgh, UK, 25-29, 1987.
- [6] G. Kiss, A. Arató, J. Lukács, J. Sulyán, T. Vaspöri, *BraiLab, a full Hungarian text-to-speech microcomputer for the blind*, In Proceedings of World Conference on Phonetics, 1-4, 1987.
- [7] J. Király, *A PC TALKER beszédszintetizátor és digitális hangrögzítő-visszajátszó rendszer*, Magyar Elektronika, VI. évf., 12. szám, 7. o., 1989.
- [8] G. Olaszy, *Multivox – A flexible text-to-speech system for Hungarian, Finnish, German, Esperanto, Italian, and other languages for IBM-PC*, First European Conference on Speech Communication and Technology (EUROSPEECH '89), Paris, France, 2525-2528, 1989.
- [9] G. Olaszy, G. Németh, P. Olaszi, G. Kiss, C. Zainkó, G. Gordos, *Profivox — A Hungarian text-to-speech system for telecommunications applications*, International Journal of Speech Technology, 3, 201-215, 2000.
- [10] G. Balogh, E. Dobler, T. Grobler, B. Smodies, C. Szepesvári, *FlexVoice: A parametric approach to high-quality speech synthesis*, IEE Seminar on State of the Art in Speech Synthesis, 189-194, 2000.
- [11] R. Sproat, *Multilingual text-to-speech synthesis: The Bell Labs approach*, Springer, 1997.
- [12] M. Schröder, J. Trouvain, *The German text-to-speech synthesis system MARY: A tool for research, development and teaching*, International Journal of Speech Technology, 6, 365-377, 2003.
- [13] P. Taylor, *Text-to-speech synthesis*, Cambridge University Press, 2009.
- [14] H. U. Mullah, *A comparative study of different text-to-speech synthesis techniques*, International Journal of Scientific Engineering and Research, 6(6), 287-292, 2015.

- [15] S. Ö. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, A. Ng, J. Raiman, S. Sengupta, M. Shoeybi, *Deep voice: Real-time neural text-to-speech*, In Proceedings of the 34th International Conference on Machine Learning, PMLR, 70, 195-204, 2017.
- [16] Y. Shiga, J. Ni, K. Tachibana, T. Okamoto, *Text-to-Speech Synthesis*, In Y. Kidawara, E. Sumita, H. Kawai (eds) Speech-to-Speech Translation. Springer Briefs in Computer Science. Springer, Singapore, 2020.
- [17] X. Zhou, M. Zhang, Y. Zhou, Z. Wu, H. Li, *Accented text-to-speech synthesis with limited data*, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 32, 1699-1711, 2024.
- [18] J. James, B. T. Balamurali, C. I. Watson, B. MacDonald, *Empathetic speech synthesis and testing for healthcare robots*, International Journal of Social Robotics, 13(8), 2119-2137, 2021.
- [19] K. Ashok, M. Ashraf, J. Thimmia Raja, M. Z. Hussain, D. K. Singh, A. Haldorai, *Collaborative analysis of audio-visual speech synthesis with sensor measurements for regulating human-robot interaction*, International Journal of System Assurance Engineering and Management, 1-8, 2022.
- [20] B. Pang, J. Teng, Q. Xu, Y. Song, X. Yuan, Y. Li, *Chinese personalised text-to-speech synthesis for robot human-machine interaction*, IET Cyber-Systems and Robotics, 5(3), e12098, 2023.
- [21] B. H. Story, I. R. Titze, E. A. Hoffman, *Vocal tract area functions for an adult female speaker based on volumetric imaging*, Journal of the Acoustical Society of America, 104(1), 471-487, 1998.
- [22] P. Birkholz, D. Jackel, B. J. Kröger, *Construction and control of a three-dimensional vocal tract model*, In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, 1, I-I, 2006.
- [23] Y. C. Kim, J. Kim, M. Proctor, A. Toutios, K. Nayak, S. Lee, S. S. Narayanan, *Toward automatic vocal tract area function estimation from accelerated three-dimensional magnetic resonance imaging*, In Speech Production in Automatic Speech Recognition, Lyon, France, 40-43, 2013.
- [24] Z. I. Skordilis, A. Toutios, J. Töger, S. Narayanan, *Estimation of vocal tract area function from volumetric Magnetic Resonance Imaging*, In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 924-928, 2017.
- [25] J. Zhang, K. Honda, J. Wei, *Tooth visualization in vowel production MR images for three-dimensional vocal tract modeling*, Speech Communication, 96, 37-48, 2018.
- [26] S. Dabbaghchian, M. Arnela, O. Engwall, O. Guasch, *Reconstruction of vocal tract geometries from biomechanical simulations*, International Journal for Numerical Methods in Biomedical Engineering, 35, e3159, 2019.

- [27] P. Birkholz, S. Kürbis, S. Stone, P. Häsner, R. Blandin, M. Fleischer, *Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties*, Scientific Data, 7(1), 255, 2020.
- [28] M. Leppävuori, E. Lammentausta, A. Peuna, M. K. Bode, J. Jokelainen, J. Ojala, M. T. Nieminen, *Characterizing vocal tract dimensions in the vocal modes using magnetic resonance imaging*, Journal of Voice, 35(5), 804-e27, 2021.
- [29] J. Beskow, "Trainable articulatory control models for visual speech synthesis," International Journal of Speech Technology, 7(4), 335-349, 2004.
- [30] T. Hueber, G. Aversano, G. Cholle, B. Denby, G. Dreyfus, Y. Oussar, P. Roussel, M. Stone, *Eigentongue feature extraction for an ultrasound-based silent speech interface*, In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, 1, I-1245, 2007.
- [31] B. Denby, T. Schultz, K. Honda, T. Hueber, J.M. Gilbert, J.S. Brumberg, *Silent speech interfaces*, Speech Communication, 52(4), 270-287, 2010.
- [32] W. Mattheyses, W. Verhelst, *Audiovisual speech synthesis: An overview of the state-of-the-art*, Speech Communication, 66, 182-217, 2015.
- [33] K. Richmond, Z. Ling, J. Yamagishi, *The use of articulatory movement data in speech synthesis applications: An overview—application of articulatory movements using machine learning algorithms—*, Acoustical Science and Technology, 36(6), 467-477, 2015.
- [34] J. A. Gonzalez-Lopez, A. Gomez-Alanis, J. M. M. Donas, J. L. Pérez-Córdoba, A. M. Gomez, *Silent speech interfaces for speech restoration: A review*, IEEE access, 8, 177995-178021, 2020.
- [35] Y. Jin, Y. Gao, X. Xu, S. Choi, J. Li, F. Liu, Z. Li, Z. Jin, *EarCommand: "Hearing" your silent speech commands in ear*, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6(2), 1-28, 2022.
- [36] B. Cao, S. Ravi, N. Sebkhi, A. Bhavsar, O. T. Inan, W. Xu, J. Wang, *Magtrack: A wearable tongue motion tracking system for silent speech interfaces*, Journal of Speech, Language, and Hearing Research, 66(8S), 3206-3221, 2023.
- [37] A. Toutios, S. S. Narayanan, *Articulatory synthesis of French connected speech from EMA data*, In Interspeech, 2738-2742, 2013.
- [38] Y. Otani, S. Sawada, H. Ohmura, K. Katsurada, *Speech synthesis from articulatory movements recorded by real-time MRI*, In Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech, 127-131, 2023.
- [39] B. Denby, M. Stone, *Speech synthesis from real time ultrasound images of the tongue*, In 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1, I-685, 2024.

- [40] T. Kaburagi, *Determining the length and cross-sectional area of the vocal tract jointly from formants using acoustic sensitivity function*, Acoustical Science and Technology, 35(6), 290-299, 2014.
- [41] T. Kaburagi, *A method for estimating vocal-tract shape from a target speech spectrum*, Acoustical Science and Technology, 36(5), 428-437, 2015.
- [42] M. S. Ribeiro, J. Cleland, A. Eshky, K. Richmond, S. Renals, *Exploiting ultrasound tongue imaging for the automatic detection of speech articulation errors*, Speech Communication, 128, 24-34, 2021.
- [43] E. Sugden, J. Cleland, *Using ultrasound tongue imaging to support the phonetic transcription of childhood speech sound disorders*, Clinical Linguistics and Phonetics, 36(12), 1047-1066, 2022.
- [44] K. Al-Hammuri, F. Gebali, I. Thirumarai Chelvan, A. Kanan, *Tongue contour tracking and segmentation in lingual ultrasound for speech recognition: A review*, Diagnostics, 12(11), 2811, 2022.
- [45] P. Saha, Y. Liu, B. Gick, S. Fels, *Ultra2speech-a deep learning framework for formant frequency estimation and tracking from ultrasound tongue images*, In Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Proceedings, Part III 23, 473-482, 2020.
- [46] J. E. Allen, J. Cleland, M. Smith, *An initial framework for use of ultrasound by speech and language therapists in the UK: Scope of practice, education and governance*, Ultrasound, 31(2), 92-103, 2023.
- [47] Y. Yu, A. H. Shandiz, L. Tóth, *Reconstructing speech from real-time articulatory MRI using neural vocoders*, In 2021 29th European Signal Processing Conference (EUSIPCO), 945-949, 2021.
- [48] C. E. Wiltshire, M. Chiew, J. Chesters, M. P. Healy, K. E. Watkins, *Speech movement variability in people who stutter: a vocal tract magnetic resonance imaging study*, Journal of Speech, Language, and Hearing Research, 64(7), 2438-2452, 2021.
- [49] A. D. Gomez, M. L. Stone, J. Woo, F. Xing, J. L. Prince, *Analysis of fiber strain in the human tongue during speech*, Computer Methods in Biomechanics and Biomedical Engineering, 23(8), 312-322, 2020.
- [50] R. Mannem, N. Gaddam, P. K. Ghosh, *Air-tissue boundary segmentation in real time magnetic resonance imaging video using 3-D convolutional neural network*, In Interspeech, 1396-1400, 2020.
- [51] J. Kim, A. Toutios, S. Lee, S. S. Narayanan, *Vocal tract shaping of emotional speech*, Computer Speech and Language, 64, 101100, 2020.
- [52] M. J. Ball, V. Gracco, M. Stone, *A comparison of imaging techniques for the investigation of normal and disordered speech production*, Advances in Speech Language Pathology, 3(1), 13-24, 2001.

- [53] M. Aron, M.-O. Berger, E. Kerrien, *Multimodal fusion of electromagnetic, ultrasound, and MRI data for building an articulatory model*, In Proceedings of the 8th International Seminar On Speech Production - ISSP'08, 349–352, 2008.
- [54] K. Kansy, J. Hoffmann, N. Mistele, V. Shavlokhova, M. Bendszus, S. Heiland, J. Krisam, A. Geschwinder, J. Gradl, *Visualization and quantification of tongue movement during articulation: Is ultrasound a valid alternative to magnetic resonance imaging?*, Journal of Cranio-Maxillofacial Surgery, 46(11), 1924-1933, 2018.
- [55] <https://www.alamy.com/stock-photo/midsagittal-plane.html?sortBy=relevant>
- [56] <http://www.articulateinstruments.com/Micro%20Research%20Brochure.pdf>,
<https://magneticmag.com/wp-content/uploads/2021/08/Siemens-He-MRI-1-1024x723.jpg>
- [57] <https://sail.usc.edu/span/index.html>
- [58] G. V. Hahn-Powell, D. Archangeli, *AutoTrace: An automatic system for tracing tongue contours*, Journal of the Acoustical Society of America, 136(4_Supplement), 2104-2104, 2014.
- [59] K. Xu, T. G. Csapó, P. Roussel, B. Denby, *A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization*, Journal of the Acoustical Society of America, 139(5), EL154-EL160, 2016.
- [60] K. Al-Hammuri, F. Gebali, I. Thirumarai Chelvan, A. Kanan, *Tongue contour tracking and segmentation in lingual ultrasound for speech recognition: A review*, Diagnostics, 12(11), 2811, 2022.
- [61] L. Zhao, L. Czap, *A nyelvkontúr automatikus követése ultrahangos felvételeken*, Beszédkutatás, 27(1), 331-343, 2019.
- [62] A. J. Viterbi, *Error bounds for convolutional codes and an asymptotically optimum decoding algorithm*, IEEE Transactions on Information Theory, 13(2), 260–269, 1967.
- [63] T. G. Csapó, A. Deme, T. E. Grácsi, A. Markó, *Comparison of distance measures in tongue contour traces of ultrasound images*, Ultrafest VIII, 65-66, 2017.
- [64] <https://insightsimaging.springeropen.com/articles/10.1186/s13244-019-0694-x>
- [65] K. R. Rao, P. Yip, *Discrete cosine transform: algorithms, advantages, applications*, Academic Press, 2014.
- [66] L. Czap, *Impact of preprocessing features on the performance of ultrasound tongue contour tracking, via dynamic programming*, Acta Polytechnica Hungarica, 18(2), 159-176, 2021.
- [67] M. F. Moller, *A scaled conjugate gradient algorithm for fast supervised learning*, Neural networks, 6, 525-533, 1993.

- [68] H. Akima, *A new method of interpolation and smooth curve fitting based on local procedures*, Journal of the Association for Computing Machinery, 17(4), 589-602, 1970.
- [69] G. Gordos, Gy. Takács, *Digitális beszédfeldolgozás*, Műszaki Könyvkiadó, 1983.

A doktori disszertáció témájához kapcsolódó közlemények:

- S1: R. Trencsényi, *MRI- és UH-felvételek geometriai elemzése a beszédszintézisben*, Acta Medicinae et Sociologica 11(31), 55-65, 2020.
- S2: R. Trencsényi, L. Czap, *UH-és MRI-nyelvkontúrok optimalizációja*, In Speech Research Conference, Hungarian Research Institute for Linguistics, Budapest, Hungary, 14-15th December 2020, 86-88, 2020.
- S3: R. Trencsényi, *A nyelvkontúrkövető algoritmusok és a gépi tanulás összekapcsolhatóságának vizsgálata*, In XVI. Magyar Számítógépes Nyelvészeti Konferencia, MSZNY 2020, Szeged, Magyarország, 2020. január 23–24., 233-244, 2020.
- S4: R. Trencsényi, L. Czap, *Possible methods for combining tongue contours of dynamic MRI and ultrasound records*, Acta Polytechnica Hungarica, 18(4), 143-160, 2021.
- S5: R. Trencsényi, L. Czap, *A possible optimisation procedure for US and MRI tongue contours*, In Proceedings of the 1st Conference on Information Technology and Data Science, CITDS 2020, CEUR Workshop Proceedings, Debrecen, Hungary, 6-8th November 2020, 259-269, 2021.
- S6: R. Trencsényi, L. Czap, *Machine learning applied in speech science*, In 23rd International Carpathian Control Conference, ICCC 2022, Piscataway (NJ), Amerikai Egyesült Államok: IEEE, Sinaia, Romania, 29th May - 1st June 2022, 309-314, 2022.
- S7: R. Trencsényi, L. Czap, *Articulatory data of audiovisual records of speech connected by machine learning*, In 2nd Conference on Information Technology and Data Science, CITDS 2021, Proceedings Piscataway (NJ), Amerikai Egyesült Államok: IEEE, Debrecen, Hungary, 16-18th May 2022, 297-301, 2022.
- S8: R. Trencsényi, L. Czap, *A neural network based approach for combining ultrasound and MRI data of 2-D dynamic records of human speech*, In 13th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2022, Piscataway (NJ), Amerikai Egyesült Államok: IEEE, Budapest, Hungary, 22nd-23rd September, 47-52, 2022.
- S9: R. Trencsényi, L. Czap, *Optimisation techniques in speech processing*, In Doktoranduszok Fóruma 2021, Miskolc-Egyetemváros, Magyarország, 98-104, 2022.
- S10: R. Trencsényi, *A gépi tanulóalgoritmusok hatékonyságának vizsgálata kétdimenziós ultrahang- és MRI-felvételek adatainak összekapcsolásában*, In Doktoranduszok Fóruma 2022, Miskolc-Egyetemváros, Magyarország, 84-89, 2023.
- S11: R. Trencsényi, L. Czap, *Association of relevant anatomic contours of ultrasound and MRI images by optimisation via different geometric parameters*, In 25th International Carpathian Control Conference, ICCC 2024, Krynica Zdrój, Poland, 22nd-24th May 2024, pp. 1-6.

S12: R. Trencsényi, L. Czap, *Ultrasound- and MRI-based speech synthesis applying neural networks*, In 25th International Carpathian Control Conference, ICCCC 2024, Krynica Zdrój, Poland, 22nd-24th May 2024, pp. 1-6.

Egyéb közlemény:

S13: R. Trencsényi, L. Czap, *Artikulációs fonetikai jellemzők verifikálása kvantitatív adatokkal*, Beszédtudomány/Speech Science, 2(1), 243-260, 2021.

1. függelék

1. bad
2. bait
3. bat
4. bawed
5. bayed
6. bead
7. bed
8. beet
9. Bert
10. bet
11. bid
12. bide
13. bird
14. bit
15. bite
16. boat
17. bod
18. bode
19. booed
20. boot
21. boud
22. bought
23. bout
24. bowed
25. Boyt
26. bud
27. but
28. butte
29. byued
30. had
31. hawed
32. hayed
33. head
34. heard
35. heed
36. hid
37. hide
38. hod
39. hoed
40. hood
41. how'd
42. hoy'd
43. hud
44. hues
45. pot
46. put
47. who'd

2. függelék

1. Most a túta meg a túta volt.
2. Most a kéka meg a kéka volt.
3. Most a teta meg a teta volt.
4. Most a pepa meg a pepa volt.
5. Most a kika meg a kika volt.
6. Most a púpa meg a púpa volt.
7. Most a küka meg a küka volt.
8. Most a tóta meg a tóta volt.
9. Most a kaka meg a kaka volt.
10. Most a kóka meg a kóka volt.
11. Most a tita meg a tita volt.
12. Most a téta meg a téta volt.
13. Most a koka meg a koka volt.
14. Most a tota meg a tota volt.
15. Most a tata meg a tata volt.
16. Most a köka meg a köka volt.
17. Most a kúka meg a kúka volt.
18. Most a popa meg a popa volt.
19. Most a tuta meg a tuta volt.
20. Most a pápa meg a pápa volt.
21. Most a kőka meg a kőka volt.
22. Most a papa meg a papa volt.
23. Most a tüta meg a tüta volt.
24. Most a tūta meg a tūta volt.
25. Most a tōta meg a tōta volt.
26. Most a pőpa meg a pőpa volt.
27. Most a tőta meg a tőta volt.
28. Most a kika meg a kika volt.
29. Most a pípa meg a pípa volt.
30. Most a püpa meg a püpa volt.
31. Most a káka meg a káka volt.
32. Most a pépa meg a pépa volt.
33. Most a pipa meg a pipa volt.
34. Most a keka meg a keka volt.
35. Most a títa meg a títa volt.
36. Most a púpa meg a púpa volt.
37. Most a pupa meg a pupa volt.
38. Most a pöpa meg a pöpa volt.
39. Most a kuka meg a kuka volt.
40. Most a kúka meg a kúka volt.

3. függelék

Ábrajegyzék

- 2.1. ábra. Az emberi test középszagittális síkja.
- 2.2. ábra. A vokális traktus elemeinek elhelyezkedése egy statikus UH- (a.) és MRI-kereten (b.).
- 2.3. ábra. Dinamikus audiovizuális beszédfelvételek készítésére alkalmas UH-készülék (balra) és MRI-berendezés (jobbra).
- 2.4. ábra. Az UH-felvételek radiális geometriáját meghatározó $(r; \phi)$ polárkoordináták (a.), illetve az MRI-felvételek négyszöges geometriáját jellemző $(x; y)$ koordináták (b.).
- 3.1. ábra. Az 1° -onként berajzolt sugárirányú metszetek egy o hanghoz tartozó UH-kereten.
- 3.2. ábra. Egy o (a.), illetve egy t (b.) hanghoz tartozó UH-kereten feltüntetett radiális metszetek a vízszintes relevanciatartomány kezdő- és végszögének megjelölésével.
- 3.3. ábra. A négyszöges geometriájú MRI-kereteken definiált $(r; \phi)$ polárkoordináták (a.), és a 10° -onként berajzolt sugárirányú metszetek egy o hanghoz tartozó MRI-kereten (b.).
- 3.4. ábra. Egy o (a.), illetve egy t (b.) hanghoz tartozó MRI-kereten feltüntetett radiális metszetek a vízszintes relevanciatartomány kezdő- és végszögének megjelölésével.
- 3.5. ábra. A radiális metszetek mátrixos szerkezetbe történő transzformációja egy o hanghoz tartozó UH- (a.) és MRI-kereten (b.).
- 3.6. ábra. Élkimelő mátrixok grafikus megjelenítése három különböző konstrukcióban.
- 3.7. ábra. Élkimelő és átlagoló operációk alkalmazása egy o hanghoz tartozó, négyszöges geometriájú UH- (a. és b.) és MRI-kereten (c. és d.).
- 3.8. ábra. A függőleges relevanciatartomány kiszabása egy o hanghoz tartozó UH- (a.) és MRI-kereten (b.).
- 3.9. ábra. A Viterbi-algoritmus segítségével illesztett nyelvkontúrok egy o hanghoz tartozó UH- (a.) és MRI-kereten (b.).
- 3.10. ábra. A radiális geometriában megjelenő nyelvkontúrok simítatlan görbéi egy o hanghoz tartozó UH- (a.) és MRI-kereten (b.).
- 3.11. ábra. A radiális geometriában megjelenő nyelvkontúrok simított görbéi egy o hanghoz tartozó UH- (a.) és MRI-kereten (b.).
- 4.1. ábra. A 40 UH-bemondás összes keretéhez tartozó 11 111 darab nyelvkontúr halmaza (a.), illetve az összes nyelvkontúr által alkotott görbeseregnek megfelelő valószínűségi sűrűségterkép (b.).
- 4.2. ábra. Természetes (a.) és funkcionális (b.) hibákból eredő fals UH-nyelvkontúrok.
- 4.3. ábra. Az első közelítésben kapott burkológörbe az összes nyelvkontúr halmazán (a.) fehérrel, illetve a sűrűségterképen (b.) pirossal megrajzolva.
- 4.4. ábra. A szájpadkontúr hátsó szakaszának másod- (kék), harmad- (sárga) és negyedrendű (rózsaszín) közelítése az összes nyelvkontúr halmazán (a.), illetve a sűrűségterképen (b.).
- 4.5. ábra. A szájpadkontúr első- (felső görbe) és negyedrendű (alsó görbe) közelítése az összes nyelvkontúr halmazán.
- 4.6. ábra. A pirossal kiemelt etalongörbe az összes nyelvkontúr halmazán.
- 4.7. ábra. A szájpadkontúr hátsó szakaszának ötödrendű (kék) közelítése az összes nyelvkontúr halmazán (a.), illetve a sűrűségterképen (b.).
- 4.8. ábra. A szájpadkontúr negyed- és ötödrendű közelítéséből származó görbék egyesítése (a.), illetve az eredeti görbe (piros), valamint annak egyszeres (zöld) és kétszeres (kék) gördülő átlagai (b.).
- 4.9. ábra. A szájpadkontúr kézzel megrajzolt kétszeres gördülő átlaga (a.), illetve a görbe felcsapódó far részének lineáris transzformációja nyomán létrejövő kontúr (b.).

- 4.10. ábra. A szájpaddockontúr fehérrel megrajzolt háromszoros gördülő átlaga (a.), illetve annak lineáris extrapolációval meghosszabbított változata (b.).
- 4.11. ábra. A diszkrét koszinusztranszformációval simított szájpaddockontúr végleges alakja az összes nyelvkontúr halmazán (a.), illetve egy nyugalmi nyelválláshoz tartozó UH-kereten (b.).
- 4.12. ábra. Az összes nyelvkontúr által alkotott görbeseregnek megfelelő bináris sűrűségtérkép.
- 4.13. ábra. A szélsőérték-probléma keretében meghatározott (zöld), illetve a kontúrkereső algoritmussal detektált (piros) szájpaddockontúr a bináris sűrűségtérképen.
- 4.14. ábra. A 28 MRI-bemondás összes keretéhez tartozó 3 395 darab keményszájpaddockontúr halmaza.
- 4.15. ábra. A kemény szájpaddockontúrja az összes szájpaddockontúr halmazán (a.), illetve egy nyugalmi nyelválláshoz tartozó MRI-kereten (b.).
- 4.16. ábra. Automatikusan illesztett lágyszájpaddockontúrok egy o (a.), illetve egy n (b.) hanghoz tartozó MRI-kereten.
- 4.17. ábra. A művi úton beszúrt maximális világosságú tartomány (vízszintes fehér csík) egy o (a.), illetve egy n (b.) hanghoz tartozó MRI-kereten.
- 4.18. ábra. A teljes szájpaddockontúr diszkontinuitást tartalmazó (a. és c.) és folytonos (b. és d.) görbéje egy o (a. és b.), illetve egy n (c. és d.) hanghoz tartozó MRI-kereten.
- 5.1. ábra. A gégefedő zöld körvonallal határolt csúcspontja és a transzformáció piros keresztjellel megjelölt középpontja egy-egy nyugalmi UH- (a.) és MRI-kereten (b.).
- 5.2. ábra. Az UH- és MRI-keretekhez rendelt pitagorasz geometria (bal és jobb) elemeinek grafikus megjelenítése.
- 5.3. ábra. A kontúrtranszformációk grafikus vizualizálása: a kék görbe R skálafaktorral történő kicsinyítése (a.), illetve Φ_0 szöggel történő síkbeli elforgatása (b.) a piros görbéhez viszonyítva.
- 5.4. ábra. A két tetszőleges görbe viszonylatában értelmezett NND távolságmérték kétirányú számításának szemléltetése.
- 5.5. ábra. A fixpontok (a gégefedő G csúcspontja és a transzformáció C középpontja) relatív helyzetének szemléltetése egy nyugalmi UH-kereten a nyelvcsont árnyéksávjához viszonyítva.
- 5.6. ábra. A nyelvcsont elhelyezkedése egy sematikus képen és egy CT-felvételen (a.), illetve egy nyugalmi MRI-kereten (b.).
- 5.7. ábra. A gégefedő G' csúcspontjának és a nyelvcsont árnyéksávjának relatív helyzete egy nyugalmi MRI-kereten.
- 5.8. ábra. Egy G' középpontú, R_2 sugarú kör (piros), a nyelvcsont árnyéksávja (sárga), a C' pont potenciális tartománya (zöld) és az előző három geometriai halmaz közös K metszete a körív mentén (kék) egy nyugalmi MRI-kereten (a.) változtatható R_2 sugarú körök illusztrálásával (b.).
- 5.9. ábra. Az a (a.) és k (b.) hanghoz tartozó összes nyelvkontúron felvett átlaggörbe (fekete), illetve az átlaghoz legközelebb eső nyelvkontúr (piros).
- 5.10. ábra. Az optimalizáció eredményeképpen kapott C' (sárga) és G' (zöld) pontok helyzete a nyelvcsont árnyéksávjához (sárga), illetve a gégefedő csúcspontjának valódi pozíciójához (piros) viszonyítva.
- 5.11. ábra. Az $I_R = [0,3; 0,4]$ intervallumban értelmezett 21 nagyításértékhez tartozó kör egy nyugalmi MRI-kereten egészében (a.) és egy kinagyított részletben (b.) ábrázolva.
- 5.12. ábra. Az $r = 0,3$ tűrési paraméter beállításával kapott kék pontok halmaza C' potenciális pozícióira a 21 körív mentén a zöld tartományon belül (a.), illetve az $R = 0,385$

- nagyításértékhez tartozó körív mentén megtalált pontok sorozatának kinagyítása (b.).
- 5.13. ábra. A nyelvcsont árnyéksávjába eső C' pontok kiválogatásának eredménye.
- 5.14. ábra. Az optimalizáció eredményeképpen kapott C' (sárga) pont helyzete a nyelvcsont árnyéksávjához (sárga), illetve a gégefedő csúcspontjának valódi pozíciójához (piros) viszonyítva.
- 5.15. ábra. A C' pont két szélső helyzete a $\Phi_0 = 0,05$ (a.) és a $\Phi_0 = 0,3$ (b.) szögelfordulási értékek esetén.
- 5.16. ábra. A $\Phi_0 = 0,19$ (a.), illetve a $\Phi_0 = 0,27$ (b.) szögelfordulási értékek beállításakor kapott C' pont pozíciói.
- 5.17. ábra. Az optimalizáció vizuális eredményei az e (a. és b.) és t (c. és d.) hangokhoz tartozó UH- (zöld) és MRI-nyelvkontúrok (piros) esetében.
- 5.18. ábra. Az optimalizáció vizuális eredményei az $ö$ hanghoz tartozó UH- (zöld) és MRI-nyelvkontúrok (piros) esetében.
- 5.19. ábra. Egy tetszőlegesen kiválasztott UH-kereten (a.), illetve az MRI-kereten (b.) felvett nyelvkontúr maximumpontja a k hang esetében.
- 5.20. ábra. A kevésbé szigorú optimalizáció vizuális eredményei az e (a. és b.) és t (c. és d.) hangokhoz tartozó UH- (zöld) és MRI-nyelvkontúrok (piros) esetében.
- 6.1. ábra. Egy adott MRI-nyelvkontúr (piros) mentén speciálisan megválasztott 5 tanítópont (zöld) a vízszintesen és függőlegesen levetített koordináták bejelölésével.
- 6.2. ábra. Az s hanghoz tartozó MRI-nyelvkontúrok az inverz transzformáció első 4 (a.), illetve 8 (b.) DCT-együtthatójával történő simítást követően.
- 6.3. ábra. A megszerkesztett neurális hálózatok alapvető struktúrája.
- 6.4. ábra. A tanítópontok sematikus elrendezése a neurális hálózat bemenetének öt különböző beállításában.
- 6.5. ábra. 1 (a.) és 5 (b.) tanítóponttal végzett tanítás eredményeként kapott görbék (piros) az eredetileg illesztett MRI-nyelvkontúrhoz (zöld) viszonyítva az a és t hangok esetében.
- 6.6. ábra. 5 (a.) és 10 (b.) DCT-együtthatóval végzett tanítás eredményeként kapott görbék (piros) az eredetileg illesztett MRI-nyelvkontúrhoz (zöld) viszonyítva az e és k hangok esetében.
- 6.7. ábra. Az átmeneti keretek kiszűrése előtti (a.) és utáni (b.) UH-nyelvkontúrok görbeserege az e hang esetében.
- 6.8. ábra. Negatív iránytangensű fals UH-nyelvkontúrok az o hang esetében.
- 6.9. ábra. A szájpadkontúr zöld görbéje a 40 UH-bemondás összes nyelvkontúrára rávetítve a fals nyelvkontúrok megjelölésével.
- 6.10. ábra. A szájpadkontúron (zöld) túlcsorduló fals UH-nyelvkontúrok a k hang esetében.
- 6.11. ábra. Konvex szakaszokat tartalmazó fals UH-nyelvkontúrok a t hang esetében.
- 6.12. ábra. Az a hanghoz tartozó görbeseregek a kirívó nyelvkontúrokat tartalmazó (a.) és a kirívó nyelvkontúroktól mentes (b.) változatban.
- 6.13. ábra. A betanított (piros) és az illesztett (zöld) MRI-nyelvkontúrok 1 (a.), illetve 5 (b.) tanítópont alkalmazásával a redukált UH-nyelvkontúrok felhasználásával az o hang esetében.
- 6.14. ábra. A betanított (piros) és az illesztett (zöld) MRI-nyelvkontúrok 5 (a.), illetve 10 (b.) DCT-együttható alkalmazásával a redukált UH-nyelvkontúrok felhasználásával a t hang esetében.
- 6.15. ábra. A betanított (piros) és az illesztett (zöld) MRI-nyelvkontúrok 1 (a. és c.), valamint 5 (b. és d.) tanítópont, illetve 5 (e. és g.), valamint 9 (f. és h.) DCT-együttható

alkalmazásával a *g* és *l* hangok esetében.

- 7.1. ábra. A vokális traktus modellezése egyenként állandó keresztmetszetű, egyforma hosszúságú csőszakaszokkal.
- 7.2. ábra. A lineáris predikció analízismodellje.
- 7.3. ábra. A lineáris predikció szintézismodellje.
- 7.4. ábra. A vokális traktus különböző tartományainak kontúrjai szeparált (a.) és interpolációval egyesített (b.) formában az *o* hang esetében.
- 7.5. ábra. A vokális traktus különböző tartományainak kontúrjai szeparált (a.) és interpolációval egyesített (b.) formában az *ü* hang esetében.
- 7.6. ábra. A szagittális radiális távolságok méréséhez felvett sugárirányú metszetek egy *k* hanghoz tartozó UH-kereten (a.), illetve egy *ö* hanghoz rendelt MRI-kereten (b.).
- 7.7. ábra. A nyelvkontúr túlcsordul a szájpaddockontúron az *ny* (a.) és *k* (b.) hangok esetében.
- 7.8. ábra. Többszörös metszéspontok létrejötte a kézzel megjelölt radiális metszetek mentén nagy görbületű kontúrszakaszoknál a *cs* (a.) és *zs* (b.) hangok esetében.
- 7.9. ábra. UH- (a. és b.) és MRI-keretek (c. és d.) radiális geometriában mintavételezett és négyszöges geometriába transzformált mintasorozatai szájpadd- (a. és c.) és nyelvkontúrok (b. és d.) esetén.
- 7.10. ábra. A többszörös metszéspontok és a hézagok megszüntetésével előállított, korrigált mintasorozatok egy UH- (a.) és egy MRI-nyelvkontúr (b.) esetében.
- 7.11. ábra. Az UH- és MRI-környezet eredeti (zöld és piros) és transzformált (sárga) szájpadd- és nyelvkontúrjainak relatív elhelyezkedése a megfelelő radiális metszővonalak által kijelölt tartományokban.
- 7.12. ábra. Az eredeti (a.) és a szintetizált (b.) beszédhang esetében előálló átviteli karakterisztikák az *ö* hang példájában a lokális maximumok helyein azonosítható formánsfrekvenciák értékeivel.
- 7.13. ábra. A férfiak (bal) és a nők (jobb) formánsfrekvencia-tartományai az egyes magánhangzók esetében.
- 7.14. ábra. A reflexiós tényezők (a.), illetve a lineáris predikció elve (b.) alapján produkált beszédhangok formánstérképei.
- 7.15. ábra. Az auditív percepción alapuló szubjektív tesztben részt vevők szavazatainak megoszlása négy különböző módon szintetizált mondat (A, B, C, D) próbájakor.

4. függelék

Táblázatjegyzék

- 2.1. táblázat. A felhasznált UH- és MRI-felvételek térbeli és időbeli felbontásához kapcsolódó paraméterek.
- 5.1. táblázat. Az *Optimalizáció 1* algoritmus futtatásával kapott eredmények.
- 5.2. táblázat. A C' pontok x és y koordinátái, illetve a hozzájuk tartozó szögelfordulások az $R = 0,385$ nagyításérték esetében.
- 5.3. táblázat. Az *Optimalizáció 2.a* algoritmus futtatásával kapott eredmények.
- 5.4. táblázat. Az *Optimalizáció 2.b* algoritmus futtatásával kapott eredmények.
- 5.5. táblázat. Az *Optimalizáció 3.a* algoritmus futtatásával kapott eredmények.
- 5.6. táblázat. Az *Optimalizáció 3.b* algoritmus futtatásával kapott eredmények.
- 5.7. táblázat. Az *Optimalizáció 4* algoritmus futtatásával kapott eredmények.
- 5.8. táblázat. Az *Optimalizáció 4* algoritmus futtatásával kapott legvégső eredmények.
- 6.1. táblázat. A V_1 verzió alkalmazásával kapott eredmények.
- 6.2. táblázat. A V_2 verzió alkalmazásával kapott eredmények.
- 6.3. táblázat. Az összes és a redukált UH-nyelvkontúrok száma hangonkénti bontásban.
- 6.4. táblázat. A betanított és illesztett MRI-nyelvkontúrok között mért átlagos NND távolságok különböző számú tanítópontok alkalmazásával.
- 6.5. táblázat. A betanított és illesztett MRI-nyelvkontúrok között mért átlagos NND távolságok különböző számú DCT-együtthatók alkalmazásával.
- 6.6. táblázat. A négy különböző rendszerkonfiguráció értelmezése a bemeneti és kimeneti paraméterek tükrében.
- 6.7. táblázat. A redukált UH-nyelvkontúrok száma hangonkénti bontásban.
- 6.8. táblázat. A legjobb tanítási eredmények a négy különböző rendszerkonfiguráció esetében.
- 7.1. táblázat. Egy kiválasztott UH-keret nyelv- és szájpaddocktúrajai között mérhető radiális távolságok a *Távolságmérés 1* és *Távolságmérés 2* metódusok alkalmazásával.
- 7.2. táblázat. Egy kiválasztott MRI-keret nyelv- és szájpaddocktúrajai között mérhető radiális távolságok a *Távolságmérés 1* és *Távolságmérés 2* metódusok alkalmazásával.
- 7.3. táblázat. A neurális hálózat bemeneti és kimeneti adathalmazainak dimenziói az összes rendszerkonfiguráció esetében.
- 7.4. táblázat. Az eredeti és a reflexiók tényezők alapján szintetizált magánhangzók formánsfrekvenciái.
- 7.5. táblázat. Az eredeti és a keresztmetszetek alapján szintetizált magánhangzók formánsfrekvenciái.
- 7.6. táblázat. A szubjektív tesztelésre bocsátott szintetizált audiofelvétel négy különböző változata a neurális hálózat paramétereinek függvényében.