Acta Medicinae et Sociologica (2020)
Vol. 11. No. 31. (55-65)

doi:

UNIVERSITY OF DEBRECEN FACULTY OF HEALTH

NYÍREGYHÁZA

Analysing MRI and ultrasound scans in speech synthesis

MRI- és UH-felvételek geometriai elemzése a beszédszintézisben

Trencsényi Réka

egyetemi adjunktus, Debreceni Egyetem, Villamosmérnöki Tanszék

INFO

Trencsényi Réka
trencsenyi.reka@science.unideb.hu

Keywords
MRI, ultrasound, speech synthesis, mechanical speech

ABSTRACT

Abstract. The articulatory speech synthesis is a new trend in producing machine speech which is based on processing visual information related to voice formation. The profound knowledge of static and dynamic geometrical parameters of speech organs plays a fundamental role in the realization of speech synthesis. To visualize these data MRI and ultrasound scans, which have different geometry, could serve as appropriate sources. The pixels of ultrasound frames can conveniently be managed by setting a polar coordinate system, while for the description of MRI frames a Desceartes coordinate system can serve as a start. Since the ultrasound scans, as opposed to MRI, do not show the back part and the apex of the tongue, only partial information is gained on the movement of the tongue. Consequently, it is important and not trivial at all to concert the geometry of MRI and Ultrasound resources. This writing presents a possible way of geometrical transformation.

Kulcsszavak
MRI, UH, beszédszintézis, gépi beszéd

Absztrakt. A gépi beszéd előállításának egyik új vonulata az artikulációs beszédszintézis, ami a hangképzéshez kapcsolódó vizuális információk feldolgozásán alapszik. A hangképző szervek statikus és dinamikus geometriai paramétereinek pontos ismerete alapvető szerepet játszik a beszédszintézis megvalósításában. Ezen adatok vizuális kinyerésének alkalmas forrásai lehetnek a beszéd közben készült MRI- és UH-felvételek, melyek különböző geometriával jellemezhetők. Az UH-keretek képpontjai egy polárkoordináta-rendszer kijelölésével kezelhetők a legkényelmesebben, míg az MRI-keretek képpontjainak leírásához egy descartes-i koordináta-

rendszer adhat megfelelő kiindulópontot. Mivel az UH-felvételeken nem látható a nyelv hátsó része és a nyelvhegy, így az MRI-hez képest csak részleges információt kaphatunk a nyelv mozgásáról. Ennélfogva fontos és egyben nem triviális feladat az MRI- és UH-források geometriájának összehangolása. A publikációban bemutatom a geometriai transzformációk egy lehetséges módját.

Bevezető

Az emberi beszéd fiziológiai, akusztikai, lingvisztikai, prozódiai, illetve percepciós vonatkozásainak tanulmányozása két alapvető tématerület köré csoportosul, melyeket az artikulációs és akusztikai jellemzők között megvalósítandó transzformáció iránya szerint beszédfelismerésnek vagy beszédszintézisnek nevezünk. Mindkét irányzat az ember-gép kapcsolat eszközeit és módszereit hivatott fejleszteni, de céljukat tekintve élesen elhatárolhatók egymástól.

A beszédfelismerő rendszerek (Averbuch, Bahl 1987; Goffin, Allauzen 2005; Huang, Acero 1995; Paul 1989; Velkei és Vicsi 2004) az emberi hangot a gép által értelmezhető kóddá alakítják át. A legegyszerűbb beszédfelismerők az akusztikai produktumot írott szöveggé transzformálják anélkül, hogy képesek lennének megérteni a hordozott jelentéstartalmat. Emellett azonban erőteljes törekvések mutatkoznak a hagyományos beszédfelismerés komplex beszédészleléssé történő kiterjesztésére, ami gyakorlatilag mindent magába foglal az emberi percepció során megállapítható információkból, mint például a beszélő személyazonossága, a beszélő érzelmi és fizikai állapotát tükröző szupraszegmentális elemek (beszédritmus, hangerő, hangmagasság, hangszín, hanglejtés, hangsúly) (Czap és Pintér 2015), illetve a hangzó szöveg jelentése és kontextusa.

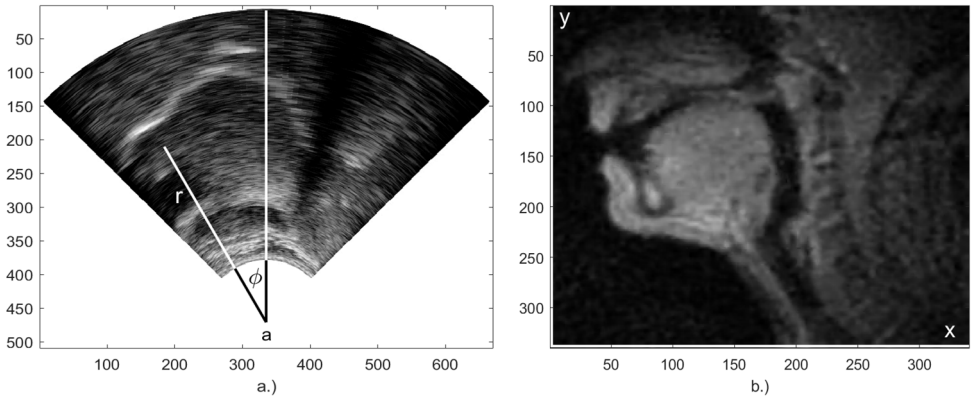
A beszédszintetizátorok (Besacier, Barnard 2014; Németh, Olaszy, Fék 2006; Olaszy 1999; Olaszy, Németh 2000; Schröder és Trouvain 2003; Sproat 1997) az artikulációs-akusztikai konverziót fordított irányban valósítják meg, azaz a gépi kód emberi beszédet utánzó hangsorozat formájában jön létre. A szintetizátorok legjellemzőbb változatait a szövegfelolvasó rendszerek alkotják, melyek emberi hangminták felhasználásával írott szövegeket szólaltatnak meg eltekintve mindenféle prozódiai tényezőtől. Az újítások – a beszédfelismeréshez hasonlóan – többek között arra irányulnak, hogy a gépi beszédben élethű módon megelevenedjenek a fentebb említett szupraszegmentális elemek is (Sheikhan 2013; Tachibana, Yamagishi 2006). A szövegfelolvasók által képviselt klasszikus koncepciók mellett újabb tendenciák is kezdenek életre kelni, melyek merőben új alapokra helyezik a gépi beszéd előállítását. Ebben a kategóriában kap helyet például az artikulációs beszédszintézis, ami az akusztikai produktum utánzását emberi hangminták helyett az emberi hangképzés és artikuláció gépi leképezése révén próbálja megvalósítani. Ebben a megközelítésben az artikulációs-akusztikai konverzió végrehajtása a beszédhez kapcsolódó vizuális információkra épül (Czap és Mátyás 2005a, 2005b), melyekhez különböző képalkotó eljárások segítségével (például mágnesesrezonancia-képalkotás (MRI), komputertomo-

gráfia (CT), ultrahang (UH) juthatunk hozzá. A vizsgálatok legalkalmasabb kiindulópontjai a beszéd közben készült UH- vagy MRI-felvételek lehetnek, melyek előnye a jó térbeli és időbeli felbontás, a kép- és hanganyag szinkronizálhatósága, illetve a beszélő alany sugárterheléstől való mentesítése.

A beszédszintézishez vezető úton rendkívül változatos módszerek és modellek alkalmazására és kidolgozására nyílik lehetőségünk, melynek során fel kell fedeznünk a bonyolultságban rejtőző egyszerűséget, miközben törekednünk kell a hatékony megoldások kifejlesztésére. Ennek fényében észszerű törekvés a hangképzés háromdimenziós folyamatát elsőként kétdimenziós alakzatokból kiindulva tanulmányozni. A beszéd közben készült kétdimenziós metszetek ugyanis technikailag könnyebben elkészíthetők és hozzáférhetők, mint a háromdimenziós felvételek. Hiteles eredmények felmutatásához azonban elengedhetetlen a síkbeli és térbeli források összekapcsolása (Douros, Tsukanova 2019; Ventura, Freitas, Tavares 2008). A vizsgálatok további szempontja lehet a különböző típusú képalkotó eljárásokkal megalkotott felvételek összehangolása (Cleland, Wrench 2011; Lulich 2018). Ez meglehetősen komplex feladat, ami hasznos lehet például az UH- és MRI-képek párhuzamos tanulmányozásában. Ebben a vonatkozásban nagy hangsúlyt kaphat az UH, hiszen míg az MRI klinikai körülmények között elérhető berendezéseket igényel, addig az UH akár egy hordozható készülék segítségével is biztosítható.

Célkitűzések

Aktuális kutatásaim a beszéd közben készült UH- és MRI-felvételek szimultán elemzésére fókuszálnak, ami elősegítheti az emberi artikulációt jellemző statikus és dinamikus paraméterek vizuális módon támogatott komplexebb kinyerését. Az MRI-felvételeket a Dél-kaliforniai Egyetem honlapján szabadon hozzáférhető multimédiás csomagból válogattam ki, az UH-felvételek pedig az MTA-ELTE Lendület Lingvális Artikuláció Kutatócsoport SonoSpeech rendszerével készült audiovizuális anyagok formájában álltak rendelkezésemre. A felvételek az emberi testet bal és jobb oldali részekre osztó szagittális síkban jelenítik meg a szájüregi régiót, így egy kétdimenziós metszeten láthatóvá válik a nyelv fel-le, illetve előre-hátra irányú mozgása. A dinamikus mozgóképek statikus képkockákra bonthatók, melynek folytán a beszédkeltés egymást követő mozzanatai lépésről lépésre tanulmányozhatók. Az 1. ábra egy UH-, illetve MRI-keretet mutat be, melyek egy női, illetve férfi bemondásból származó *k* hangnak megfelelő nyelvállást vizualizálnak.



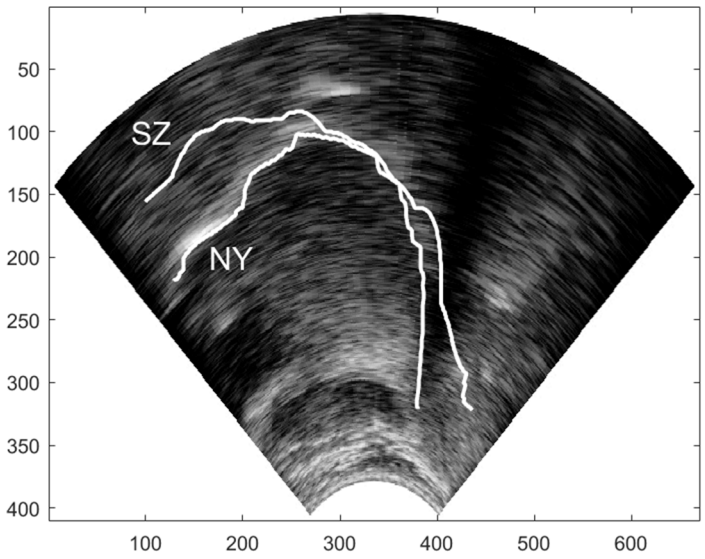
1. ábra. Egy *k* hanghoz tartozó UH- (a.) és MRI-keret (b.)

Az UH-felvételen a nyelvkontúr egy világos sávként tűnik fel, amit a nyelv és a fölötte lévő levegő határán visszaverődött sugárzás hoz létre, és a nyelvhát vonala a világos sáv alsó peremén jelölhető ki. Mivel a nyelvcsont és az állcsont részben leárnyékolja az UH-hullámokat, az UH-fej nem képes a szájüreg tartományát maradéktalanul szondázni. Ez a hiányosság a kép bal és jobb oldalán, a nyelv elülső és hátsó részénél kialakuló sötét sáv formájában mutatkozik meg, ami eltakarja a nyelvgyök és a nyelvhegy mozgását, így – a teljes szájüregi tartományt megjelenítő MRI-felvételekkel szemben – a nyelv alakjáról és mozgásáról csak részleges információt kaphatunk. További különbségként jegyezhető az a körülmény, hogy az UH-kereteken nem azonosítható a szájpad kontúrja, míg az MRI-kereteken kielégítő pontossággal behatárolható a kemény szájpad körvonala, és detektálható a lágy szájpad mozgása is. Az 1. ábrán az is megfigyelhető, hogy az UH-keretek egy körcikk által lefedett zónában vannak kifeszítve, így az egyes képpontok pozíciójának jellemzésére kényelmesen alkalmazhatók a síkbeli polárkoordináták, melyek a kör **a** középpontjától mért **r** sugár, illetve a kép függőleges szimmetriatengelyéhez viszonyított **φ** szög által definiálhatók. Ezáltal a képkocka síkjában tetszőlegesen felvett pixel helyzetét az (**r**,**φ**) koordinátapár egyértelműen meghatározza. A **φ** szög értéke minden UH-keret esetén -45° és 45° között változhat. Az MRI-keretek kezeléséhez szükséges legkomfortosabb vonatkoztatási rendszert viszont egy síkbeli descartes-i koordináta-rendszer adhatja, melyben a képkocka kijelölt pontjának pozícióját az (**x**,**y**) koordinátapár rögzíti. A kutatómunka célkitűzése az UH-felvételek radiális, illetve az MRI-felvételek négyszöges elrendezésének összehangolása a megfelelő geometriai transzformációk megállapításával.

Eredmények

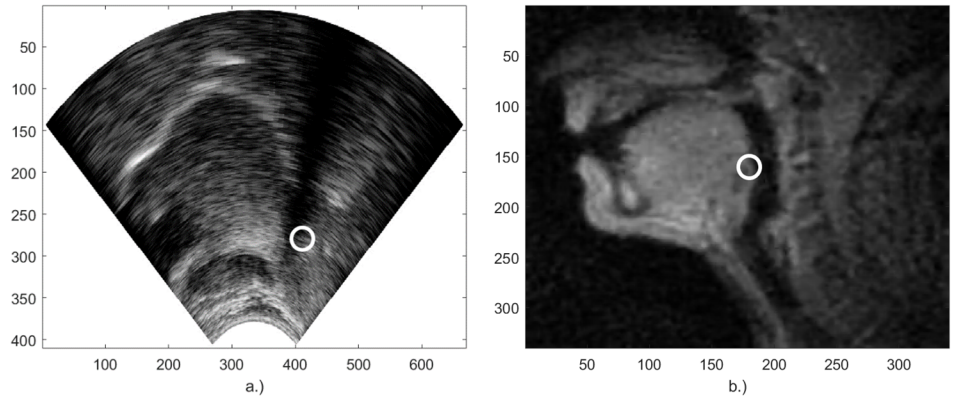
A nyelvet csak részlegesen ábrázoló UH-keretekből az MRI-keretekhez képest szűkebb információhalmazt nyerhetünk a nyelv elhelyezkedését illetően, ezért az említett geometriai transzformációk kiindulópontjául az UH-felvételek jellemző kontúrvonalait jelöltem ki. Mivel az artikuláció dinamikai leírásában a nyelvfelszín és a szájpad

relatív helyzetének változása játszik mérvadó szerepet a szájüregi régióban, a vizsgálatok feltétlenül szükségessé teszik a nyelv- és szájpadkontúr lehető legpontosabb ismeretét. A nyelvhát kontúrját dinamikus programozáson alapuló automatikus nyelvkontúrkövető algoritmusok segítségével határoztam meg. A szájpad pozíciója azonban csak becslés útján adható meg azáltal, hogy az artikuláció során a nyelvfelszín által érintett, legmagasabb helyzetben lévő pontok kiválasztásával valószínűsítjük a nyelv és a szájpad határvonalát. Ehhez természetesen olyan mássalhangzók vizsgálatára kell szorítkoznom, melyek képzése során a nyelv biztosan érintkezik a szájpad palatális (kemény szájpad) vagy veláris (lágy szájpad) zónájával. Ez a feltétel a rendelkezésemre álló, különböző bemondásokat tartalmazó UH-csomag esetében automatikusan teljesül, hiszen a rögzített mondatokban szereplő mássalhangzók képzésekor a nyelv más-más helyeken kerül kontaktusba a szájpad ívével. A szájpad kontúrjának kirajzolását lényegében egy szélsőérték-keresési probléma megoldásaként valósítottam meg, melynek eredményét a 2. ábra SZ görbéje prezentálja, a képkockához tartozó nyelvkontúrt pedig az NY görbe demonstrálja.



2. ábra. Az UH-kereteken valószínűsített szájpadkontúr (SZ) és az illesztett nyelvkontúr (NY)

A 2. ábra görbéinek transzformációjához egy olyan viszonyítási pontot kerestem, ami az UH- és MRI-kereteken is meggyőző biztonsággal azonosítható. Ezt a pontot a gégefedő csúcspontjánál definiáltam, melynek helyzetét a 3. ábra képein megrajzolt fehér körvonal lokalizálja.

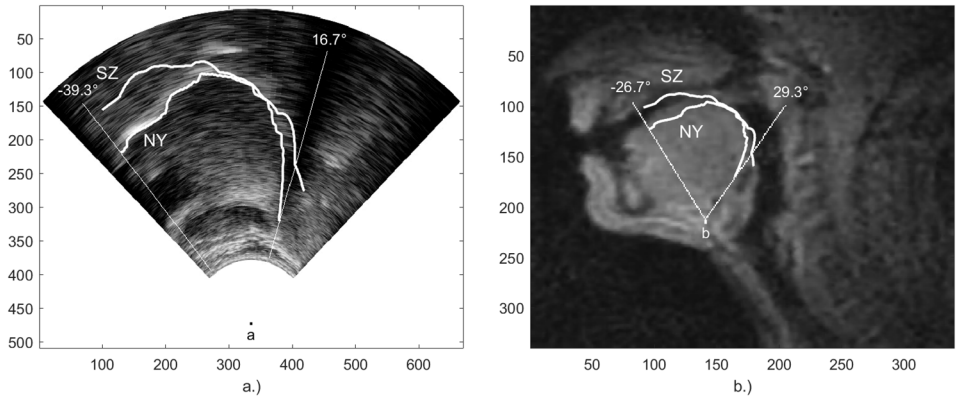


3. ábra. A gégefedő csúcspontja az UH- (a.) és MRI-kereten (b.)

A dinamikus UH-felvételek vizuális tanulmányozása során arra a következtetésre jutottam, hogy a nyelv bizonyos hangok képzésekor annyira hátrahúzódik, hogy érintkezik a gégefedővel. Ennélfogva a szájpadkontúr kezdőpontjaként a gégefedő csúcspontját jelöltem ki, és meghatároztuk a kiválasztott k hanghoz tartozó nyelvkontúr által lefedett szögtartományt, amit a -39.3° és a 16.7° értékek határolnak. A nyelv- és szájpadkontúr görbéinek transzformációját a polárkoordináta-rendszerben végeztem el a görbék pontjait jellemző (**r**,**φ**) értékpárok által adott sugár- és szögtartomány skálázásával, illetve a szögtartomány $\boldsymbol{\varphi_0}$ kezdőszögének eltolásával az

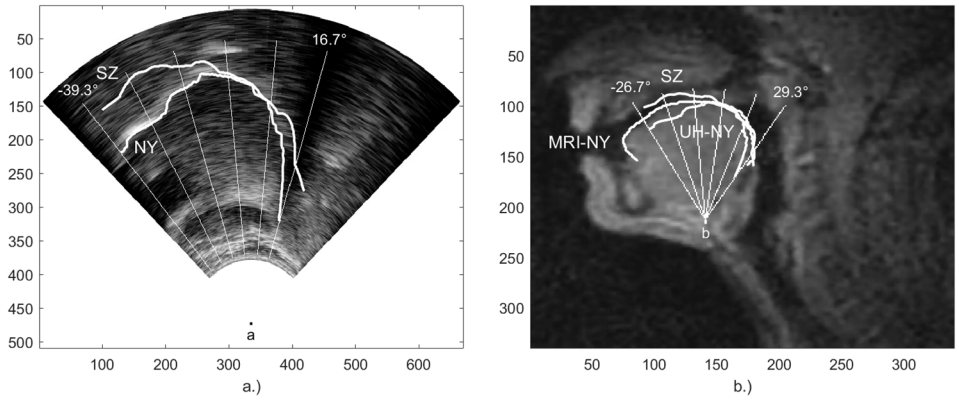
$$r' = R \cdot r,$$
$$\varphi' = FI \cdot \varphi,$$
$$\varphi_0{}' = \varphi_0 + FIKORR \qquad (1)$$

formulák szerint. Az (1) összefüggések R és FI skálafaktorai a sugár- és szögtartomány normálását teszik lehetővé, a FIKORR tag pedig a szögtartomány elforgatásáért felel. Az R=0.31, FI=1, FIKORR=12.6° értékek rögzítésével sikerült megvalósítanom a nyelv- és szájpadkontúr MRI-keretre történő átültetését. A 4. ábra tanúsága szerint a nyelv- és szájpadkontúr elfogadható módon illeszkedik az MRI-keretre, ahol a nyelvkontúr szögtartománya a -26.7° és a 29.3° értékek közé tehető. Végeredményben tehát a polárkoordináták rendszerében végrehajtott (1) transzformációkkal az UH-keretek radiális geometriáját beágyaztam az MRI-keretek négyszöges geometriájába.



4. ábra. A nyelv- és szájpadkontúr relatív helyzete az UH- (a.) és MRI-kereten (b.)

A transzformáció révén lehetővé válik az UH-, illetve MRI-keretekre illesztett nyelvkontúrok pontjainak kölcsönösen egyértelmű megfeleltetése, ami az 5. ábra segítségével követhető nyomon. A FI=1 faktornak köszönhetően a transzformáció szögtartó, ezért az UH-kereten kiválasztott négy belső radiális metszet által kijelölt négy kontúrpont az MRI-kereten megrajzolt ugyanazon négy radiális metszet mentén leképezhető az MRI-NY görbével illusztrált MRI-nyelvkontúrra, így egy adott metszet mentén haladva az UH-NY és MRI-NY görbéken megtaláljuk azt a két pontot, ami egyértelműen párba rendelhető.



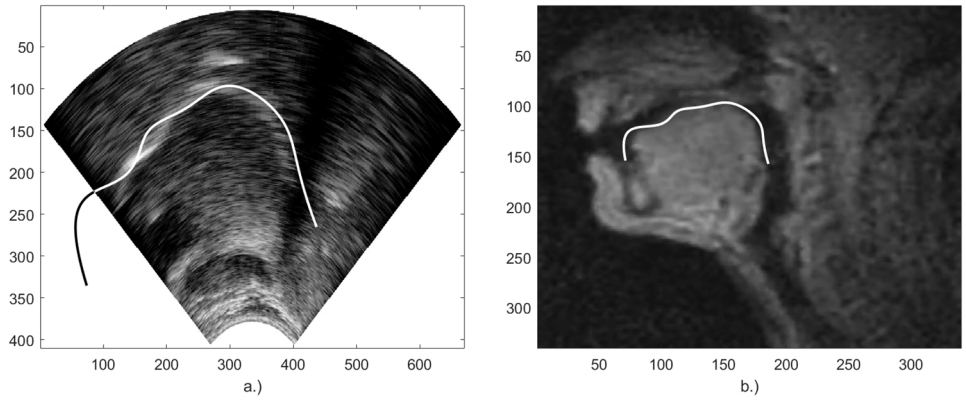
5. ábra. Az UH- és MRI-nyelvkontúr kölcsönösen egyértelmű megfeleltetése

Az UH- és MRI-nyelvkontúr pontjainak összerendelése kulcsfontosságú lehet az UH-felvételekből származó részleges információk kiegészítésében, melynek hatékony eszköze lehet például a gépi tanulóalgoritmusok alkalmazása. A gépi tanulás során a gép bizonyos bemeneti paraméterek halmazából jól definiált belső algoritmusok, ún. mintázatok segítségével előállítja a kimeneti paraméterek halmazát. A gépi tanulóalgoritmus lényegében az emberi agy működését próbálja imitálni a valódi neurális hálózatok gépi leképezésével. Jelen esetben az algoritmus a nyelvkontúr gépi betanítását valósíthatja meg, melynek be- és kimeneti paramétereit a nyelvkontúr közvetlen vagy

származtatott adatai alkothatják. Ennek megfelelően olyan neurális hálózatot szerkesztettem, melynek bemeneti paramétereit az UH-nyelvkontúr 5. ábrán kiválasztott négy pontjának **y** koordinátája, kimeneti paramétereit pedig az MRI-nyelvkontúr diszkrét koszinusz transzformáltjának első húsz együtthatója definiálja. Ezáltal egy olyan tanítási mechanizmus hozható létre, melynek során inverz koszinusz transzformáció segítségével teljes nyelvkontúrt konstruálhatunk részleges UH-adatok felhasználásával. A gépi tanítás eredményeit a 6. ábra összegzi, ahol a kapott MRI-nyelvkontúrt az (1) összefüggések által elrendelt transzformációk

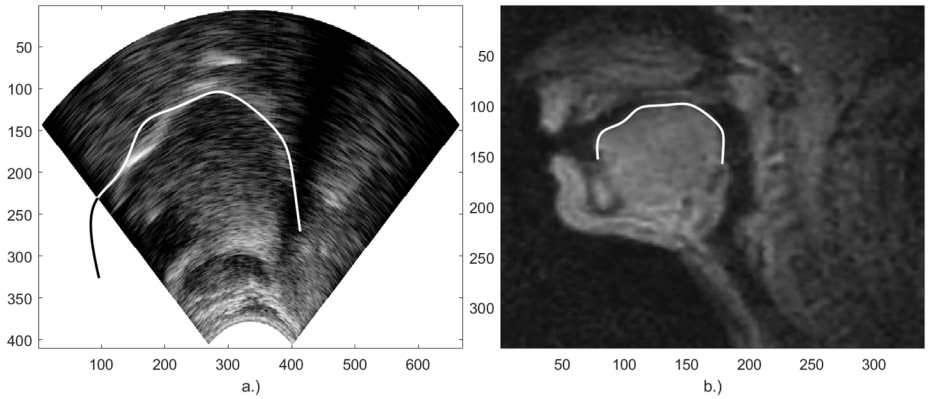
$$r = \frac{r'}{R'},$$
$$\varphi = \frac{\varphi'}{FI'},$$
$$\varphi_0 = \varphi_0{}' - FIKORR \tag{2}$$

alakú inverzének alkalmazásával rávetítettem az UH-keretre. A transzformált nyelvkontúr jól szemlélteti a nyelvhegy azon szakaszát, amely az állcsont árnyékoló hatása miatt nem jelenik meg az UH-felvételen.



6. ábra. Az UH-adatok alapján betanított nyelvkontúr az UH- (a.) és MRI-kereten (b.)

Eredményeim kontrollálása céljából gépi tanulóalgoritmusomat olyan körülmények között is lefuttattam, amikor a be- és kimeneti paraméterek ugyanazon MRI-forrásból erednek, azaz a bemeneti paraméterek halmazát az MRI-nyelvkontúr 5. ábrán megfeleltetett négy pontjának **y** koordinátája, a kimeneti paraméterek halmazát pedig az előző esethez hasonlóan az MRI-nyelvkontúr diszkrét koszinusz transzformáltjának első húsz együtthatója alkotja. Az MRI-adatok alapján betanított nyelvkontúrt a 7. ábra mutatja be, ami kvalitatíve igen jó egyezést mutat a 6. ábra UH-adatok alapján betanított nyelvkontúrjával. Az eredmények hasonlósága a tanulóalgoritmus korrekt működése mellett a geometriai transzformációk helyességét is igazolja, bár megjegyzem, hogy a jelenlegi fázisban még igen korlátozott számú tanító és tesztelő alakzat áll rendelkezésemre, de a forrásadatok folyamatos bővítés alatt állnak.



7. ábra. Az MRI-adatok alapján betanított nyelvkontúr az UH- (a.) és MRI-kereten (b.)

Összefoglaló

A cikk dinamikus UH- és MRI-felvételek anatómiai kontúrvonalainak összehangolását vizsgálja, aminek egy lehetséges módját a megfelelő geometriai transzformációk feltárása adja. A transzformációk a radiális geometriájú UH-keretek, illetve a négyszöges geometriájú MRI-keretek sugár- és szögtartományának skálázásával, illetve a szögtartomány elforgatásával kölcsönösen egyértelmű kapcsolatot teremtenek a két forrás nyelvkontúrjai között. Az eredményeket a *k* hang példáján keresztül követhetjük végig. Fontos hangsúlyozni, hogy az egzakt lépéseket tartalmazó transzformáció nem alkalmazható egységesen az összes beszédhang esetén, mivel a transzformációt leíró paraméterhalmaz hangonként változhat. Ez a körülmény megnehezíti az UH- és MRI-felvételek kompakt összeegyeztetését, de a probléma feloldható a transzformáció paramétereinek több beszédhangra kiterjedő optimalizálásával vagy a statisztikus módszerek élvonalába tartozó gépi tanulóalgoritmusok bevetésével. A transzformáció optimalizálását és a gépi tanulás kiterjedtebb alkalmazását illető lépések a jövőbeni kutatásaim alapvető feladatait fogják meghatározni, hozzájárulva ezzel az eddigi eredményeim továbbfejlesztéséhez. A jelenlegi és a jövőben tervezett vizsgálataim az artikulációs beszédszintézishez vezető út állomásainak tekinthetők. Az artikulációs beszédszintézishez kapcsolódó kutatási törekvéseknek jelentős szerepe lehet például a klinikai célú beszédterápiában, a nem anyanyelvi nyelvtanulási tréningek kialakításában vagy a néma beszéd megszólaltatásához szükséges szintetizátorok konstrukciójában és fejlesztésében.

Köszönetnyilvánítás

Köszönjük az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoportjának, hogy rendelkezésünkre bocsátották a SonoSpeech rendszerrel készült ultrahangfelvételeket.

Hivatkozások

1. Averbuch, A., Bahl, L., Bakis, R., Brown, P., Daggett, G., Das, S., Davies, K., De Gennaro, S., de Souza, P., Epstein, E., Fraleigh, D., Jelinek, F., Lewis, B., Mercer, R., Moorhead, J., Nadas, A., Nahamoo, D., Picheny, M., Shichman, G., Spinelli, P., Van Compernolle, D., Wilkens, H. (1987): *Experiments with the Tangora 20,000 word speech recognizer*, ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing, 701–704
DOI: https://doi.org/10.1109/icassp.1987.1169870
2. Besacier, L., Barnard, E., Karpov, A., Schultz, T. (2014): *Automatic speech recognition for under-resourced languages: A survey*, Speech. Comm., 56, 85–100
DOI: https://doi.org/10.1016/j.specom.2013.07.008
3. Cleland, J., Wrench, A.A., Scobbie, J.M., Semple, S. (2011): *Comparing Articulatory Images: An MRI/Ultrasound Tongue Image Database*, In Proceedings of the 9th International Seminar on Speech Production, 163–170
4. Czap László, Mátyás János (2005a): *Virtual announcer*, Infocommunications J., 60, 2–5
5. Czap László, Mátyás János (2005b): *Virtual speaker*, In: Ádám, Tihamér; Vásárhelyi, József; Varga, Attila (szerk.) Proceedings of 6th International Carpathian Control Conference ICCC 2005 Miskolc, Magyarország: Miskolci Egyetem, 351-358
6. Czap László, Pintér Judit (2015): *Intensity feature for speech stress detection*, In: Ivo, Petras; Igor, Podlubny; Jan, Kacur; Vásárhelyi, József (szerk.) Proceedings of the 16th International Carpathian Control Conference Miskolc, Magyarország: IEEE IAS/IES/PELS, 91–94
DOI: https://doi.org/10.1109/carpathiancc.2015.7145052
7. Douros, I.K., Tsukanova, A., Isaieva, K., Vuissoz, P.A., Laprie, Y. (2019): *Towards a method of dynamic vocal tract shapes generation by combining static 3D and dynamic 2D MRI speech data*, Proc. Interspeech 2019, 879-883
DOI: https://doi.org/10.21437/interspeech.2019-2880
8. Goffin, V., Allauzen, C., Bocchieri, E., Hakkani-Tur, D., Ljolje, A., Parthasarathy, S., Rahim, M., Riccardi, G., Saraclar, M. (2005): *The AT&T WATSON speech recognizer*, Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, I/1033-I/1036
DOI: https://doi.org/10.1109/icassp.2005.1415293
9. Huang, X., Acero, A., Alleva, F., Hwang, M.Y., Jiang, L., Mahajan, M. (1995): *Microsoft Windows highly intelligent speech recognizer: Whisper*, International Conference on Acoustics, Speech, and Signal Processing, 93–96
DOI: https://doi.org/10.1109/icassp.1995.479281
10. Lulich, S.M. (2018): *Registration and fusion of 3D head-neck MRI and 3D/4D tongue ultrasound*, J. Acoust. Soc. Am., 144(3), 1904–1904
DOI: https://doi.org/10.1121/1.5068345
11. Németh Géza, Olaszy Gábor, Fék Márk (2006): *Új rendszerű, korpusz alapú gépi szövegfelolvasó fejlesztése és kísérleti eredményei*, Beszédkutatás, 183-196

12. Olaszy Gábor (1999): *Beszédadatbázisok készítése gépi beszédelőállításhoz*}, Beszédkutatás99, 68–89
13. Olaszy Gábor, Németh Géza, Olaszi Péter, Kiss Géza (2000): *Profivox: a legkorszerűbb hazai beszédszintetizátor*, Beszédkutatás 2000, 167–179
14. Paul, D.B. (1989): *The Lincoln robust continuous speech recognizer*, International Conference on Acoustics, Speech, and Signal Processing, 449–452 DOI: https://doi.org/10.1109/icassp.1989.266460
15. Schröder, M., Trouvain, J. (2003): *The German text-to-speech synthesis system MARY: A tool for research, development and teaching*, Int. J. Speech Tech., 6, 365–377
16. Sheikhan, M. (2013): *Synthesizing suprasegmental speech information using hybrid of GA-ACO and dynamic neural network*, In The 5th Conference on Information and Knowledge Technology, IEEE, 175–180 DOI: https://doi.org/10.1109/ikt.2013.6620060
17. Sproat, R. W. (1997): *Multilingual text-to-speech synthesis*, Boston, KLUWER Academic Publishers Tachibana, M., Yamagishi, J., Masuko, T., Kobayashi, T. (2006): *A style adaptation technique for speech synthesis using HSMM and suprasegmental features*, IEICE transactions on information and systems, 89(3), 1092–1099 DOI: https://doi.org/10.1093/ietisy/e89-d.3.1092
18. Velkei Szabolcs, Vicsi Klára (2004): *Beszédfelismerő modellépítési kísérletek akusztikai, fonetikai szinten, kórházi leletező beszédfelismerő kifejlesztése céljából*, II. Magyar Számítógépes Nyelvészeti Konferencia, 307–315
19. Ventura, S.M.R., Freitas, D.R., Tavares, J.M. (2008): *Three-dimensional modeling of tongue during speech using MRI data*, In CMBBE 2008-8th International Symposium on Computer Methods in Biomechanics and Biomedical Engineering, 1-6