# Szinkronizált beszéd- és nyelvultrahang-felvételek a SonoSpeech rendszerrel

Csapó Tamás Gábor[1,2], Deme Andrea[1,3], Gráczi Tekla Etelka[1,4],
Markó Alexandra[1,3], Varjasi Gergely[1,3]

[1] MTA-ELTE Lendület Lingvális Artikuláció Kutatócsoport,
[2] Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
csapot@tmit.bme.hu
[3] Eötvös Loránd Tudományegyetem, Fonetikai Tanszék
{marko.alexandra,deme.andrea}@btk.elte.hu,
varjasi.gergely@gmail.com
[4] MTA Nyelvtudományi Intézet, Fonetikai Osztály
graczi.tekla.etelka@nytud.mta.hu

**Kivonat:**
A jelen ismertetés az MTA–ELTE Lingvális Artikuláció Kutatócsoport ultrahangos vizsgálatainak technikai hátterét, az alkalmazott hardver- és szoftverkörnyezetet, illetőleg a folyó és tervezett kutatásokat mutatja be. A magyar és nemzetközi szakirodalmi előzmények tárgyalása után ismerteti az ultrahangnak mint az artikuláció vizsgálatában alkalmazott eszköznek a sajátosságait, összevetve más kísérleti eszközökkel és módszertanokkal. Kitér a kutatási nehézségekre is, mint például az ultrahangkép beszélőfüggő minősége, a nyelvkontúr manuális és automatikus meghatározása, végül bemutatja a kutatócsoport főbb céljait és terveit, mind az alap-, mind pedig az alkalmazott kutatások területén.

**Kulcsszavak:** artikuláció, fonetika, beszédtechnológia

## 1 Bevezetés

Az artikuláció (a beszédképző szervek koordinált mozgása) és az akusztikum (a keletkező beszédjel) kapcsolata az 1700-as évek óta foglalkoztatja a beszédkutatókat [1]. Ahhoz, hogy a beszédképző szervek (pl. hangszalagok, nyelv, ajkak) mozgását vizsgálni tudjuk, speciális eszközökre van szükségünk, mivel a legtöbb ilyen szerv nem látható folyamatosan beszéd közben. Magyar nyelvre eddig kevés olyan artikulációs vizsgálat született, amely dinamikus adatokon (azaz nem csak statikus állóképeken) alapul. Lotz az 1960-as években [2, 3], Szende az 1970-es években [4], majd Bolla az 1980-as években [5, 6] röntgenfilm (ún. röntgenogram/ kinoröntgenografikus vizsgálat) technológiával vizsgálta a magyar beszéd artikulációját. Bolla kutatásaiban az összes magyar magánhangzót és mássalhangzót elemezte: a folyamatos röntgenfelvételekből a vizsgált beszédhangokról öt-öt képet átrajzoltak számítógépre, majd a rajzokat fonetikai szempontból elemezték. A tanulmányokban közlésre adták az összes így keletkezett konfigurációt rajzokon, illetve a toldalékcső méreteit táblázatos formában. Ezek az adatok amellett, hogy segítik a magyar beszédképzés mechaniz-

musainak megismerését és az artikulációs bázis feltárását, akár egy mai modern artikulációs elvű beszédszintetizátorhoz is felhasználhatóak lennének.

Bolla és munkatársai egy későbbi tanulmányban részletesen ismertetik a röntgenogramok készítéséhez használt eszközöket és a felvételek módszertanát [7]. Ebből kiderül, hogy a mikroszámítógépes technikát úgy dolgozták ki, hogy az interlingvális hangtani egybevetésekre is alkalmas legyen. Bolla emellett kísérletezett az ajkak (fotolabiogram) és a szájpadlás (palatogram) vizsgálatával is [8]. Az 1980-as évek röntgenes kísérletei után hosszú ideig nem történtek magyar nyelvű artikulációs kutatások, majd 2008-ban Mády elektromágneses artikulográffal vizsgálta a magyar magánhangzókat normál és gyors beszédben [9]. A magyar magánhangzók vizsgálatára újabb vizsgálat is született, melyben a magánhangzókra a beszédben és az éneklésben az alapfrekvencia függvényében jellemző nyelvkontúrokat, ajakpozíciót és az áll helyzetét (azaz az állkapocs nyitásszögét) elemezték szintén az elektromágneses artikulográfia módszerével [10].

1. táblázat: A nyelv mozgásának vizsgálatára használható technológiák összehasonlítása. EMA = elektromágneses artikulográf. MRI = mágnesrezonancia-képalkotás. PMA = permanens mágneses artikulográf.

| Technológia | Előnyök | Hátrányok |
|---|---|---|
| Röntgen | kiváló térbeli felbontás | káros az egészségre<br>nyelvkontúr követése szükséges |
| Ultrahang | jó időbeli és térbeli felbontás<br>elérhető ár | nyelvkontúr követése szükséges<br>csak az ultrahangfejre merőleges nyelvállás látszik jól |
| EMA | kiváló időbeli felbontás<br>pontonként alacsony mérési hiba | csak pontszintű mérés<br>kábelek befolyásolják a beszélőszervek mozgását |
| MRI | jó a tér- vagy időbeli felbontás (trade-off) | trade-off a tér- és időbeli felbontás között<br>fekvő helyzet, zajos körülmények<br>nyelvkontúr követése szükséges |
| PMA | jó időbeli és térbeli felbontás | nem adja meg a nyelv pontos pozícióját, csak a becsült helyzetét |

A nemzetközi szakirodalomban is számos példát találhatunk a beszéd közbeni artikuláció vizsgálatára, melyek közül a nyelv mozgásának elemzésére a következő technológiák alkalmasak: röntgen [11], ultrahang [12, 13], elektromágneses artikulográf (EMA) [14], mágnesesrezonancia-képalkotás (MRI) [15, 16] és permanens mágneses artikulográf (PMA) [17, 18]. Az egyes technikák előnyeit és hátrányait az 1. táblázatban hasonlítjuk össze. A hat technológia közül az ultrahang pozitívuma, hogy egyszerűen használható, elérhető árú, valamint nagy felbontású (akár 800 × 600 pixel), és nagy sebességű (akár 100 képkocka / másodperc) felvétel készíthető vele. A jó térbeli felbontás azért fontos, hogy a nyelv alakjáról minél pontosabb képet kapjunk, míg a jó időbeli felbontás ahhoz szükséges, hogy a beszédhangok képzésének gyors változását (pl. zárfelpattanás, koartikuláció) is vizsgálni tudjuk. Az ultrahang hátránya viszont az, hogy a hagyományos beszédkutatási kísérletekhez a rögzített képsorozatból ki kell nyerni a nyelv körvonalát ahhoz, hogy az adatokon további vizsgálatokat lehessen

végezni. Ez elvégezhető manuálisan, ami rendkívül időigényes, vagy automatikus módszerekkel, amelyek viszont ma még nem elég megbízhatóak [19, 20]. Ugyanakkor az ultrahang az egyik legelterjedtebb technológia az artikulációs kutatással foglalkozó beszédkutató laboratóriumokban [21].

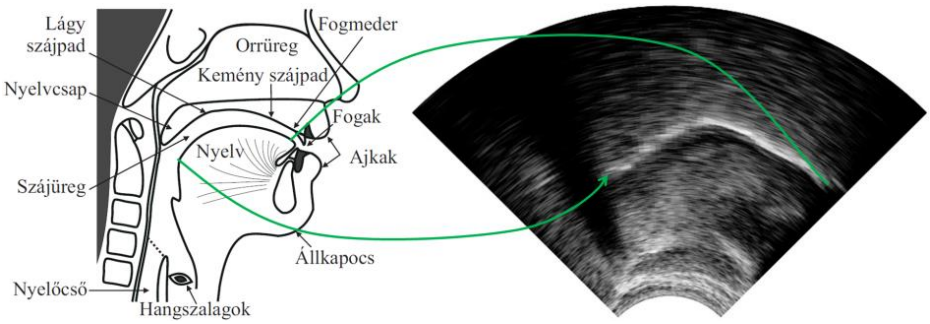
A jelen cikk célja, hogy bemutassuk az MTA–ELTE Lendület Lingvális Artikuláció Kutatócsoport magyar beszéden történő nyelvultrahangos vizsgálatainak technikai hátterét, továbbá azt, hogy a rögzített adatok milyen módon használhatóak fel beszédkutatáshoz.

## 2 A szinkronizált beszéd és nyelvultrahang felvételének módszertana

Az első kísérleti felvételek az ELTE Fonetikai Tanszékének egyik csendes szobájában készültek, a szakirodalomban javasolt helyzetben és beállításokkal [13], az 1. ábrán látható módon.

A beszélők jelentés nélküli VCVCV szerkezetű hangsorokat és mondatokat olvastak fel. Az ultrahangkép tipikus orientációjára a 2. ábra mutat egy példát.

**1. ábra.** Beszéd és ultrahang felvétele rögzítő sisakkal az ELTE Fonetikai Tanszéken.



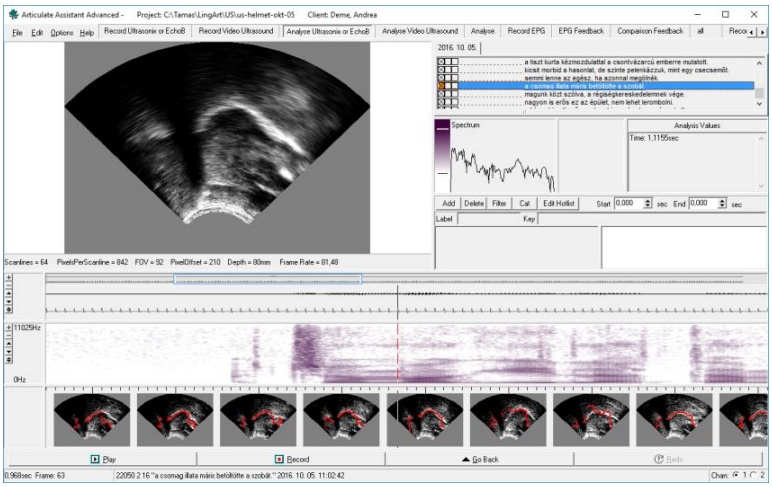

**2. ábra.** Az ultrahangos kép orientációja. A bal oldali ábra forrása: [22].

## 2.1 Hardveres környezet

A nyelv középvonalának mozgását a SonoSpeech rendszerrel rögzítettük (Articulate Instruments Ltd.) egy 2–4 MHz frekvenciájú, 64 elemű, 20 mm sugarú konvex ultrahang-vizsgálófejjel, 80–100 fps sebességgel. A felvételek során ultrahangrögzítő sisakot is alkalmaztunk (Articulate Instruments Ltd.), melyet az 1. ábra mutat. A rögzítő sisak használata azt biztosítja, hogy a felvétel során az ultrahang-vizsgálófej ne mozduljon el (pl. az orientációja ne változzon). A beszédet az első kísérletekben Monacor ECM 100 kondenzátormikrofonnal rögzítettük, melyet a kísérleti alany a kezében tartott (az 1. ábrán látható módon). A későbbiekben a beszédet Audio-Technica - ATR 3350 omnidirekcionális kondenzátormikrofonnal rögzítettük, amely a sisakra volt csíptetve, a szájtól kb. 20 cm-re. A hangot 22050 vagy 44100 Hz mintavételi frekvenciával digitalizáltuk M-Audio – MTRACK PLUS hangkártyával. Az ultrahang és a beszéd szinkronizációja a SonoSpech rendszer 'Frame sync' kimenetét használva történt: minden elkészült ultrahangkép után ezen a kimeneten megjelenik egy néhány nanoszekundum nagyságrendű impulzus, amit egy 'Pulse stretch' egység szélesebb négyszögugrássá alakít, hogy digitalizálható legyen (l. 2.3 fejezet). Ez utóbbi jelet szintén a hangkártya rögzítette.

## 2.2 Szoftveres környezet

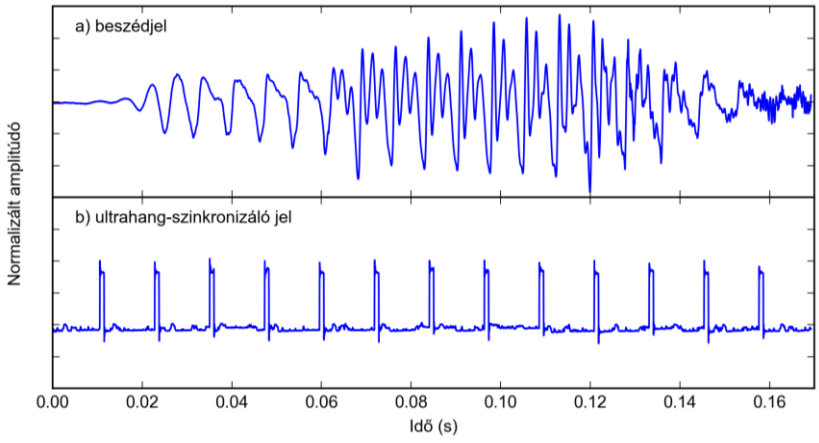
A felolvasandó mondatokat az Articulate Assistant Advanced (Articulate Instruments Ltd.) szoftver segítségével jelenítettük meg a képernyőn. Az adatokat ugyanezzel a szoftverrel rögzítettük. Az AAA szoftver az adatok elemzésére is használható: egyszerre látszik az ultrahangkép, a beszéd hullámformája, FFT-spektruma és spektrogramja (3. ábra). Emellett az ábra alján látható módon automatikus nyelvkontúrkövetésre is alkalmas, és az ultrahangképeket a beszéddel szinkronizáltan jeleníti meg.



**3. ábra.** Az Articulate Assistant Advanced szoftver használata.

## 2.3 Beszéd és ultrahang szinkronizálása

Ahhoz, hogy az ultrahangot és a beszédet később együttesen lehessen kezelni (azaz például meg tudjunk nézni egy zárfelpattanáshoz kapcsolódó ultrahangképet), nem elég a két jel párhuzamos felvétele, hanem szinkronizálni is kell azokat. A SonoSpeech ultrahang 'Frame sync' kimenetét (illetve ennek digitalizálható változatát, l. 2.1 fejezet) a kétcsatornás hangkártyára kötve, a mikrofonból származó beszédjelet és az ultrahang-szinkronizáló jelet párhuzamosan fel tudjuk venni (4. ábra). A szinkronizálójelben a négyszögek felfutó élét egy jelfeldolgozási algoritmussal megkeresve meg tudjuk határozni az egyes ultrahangképek pontos helyét a beszédhez képest.



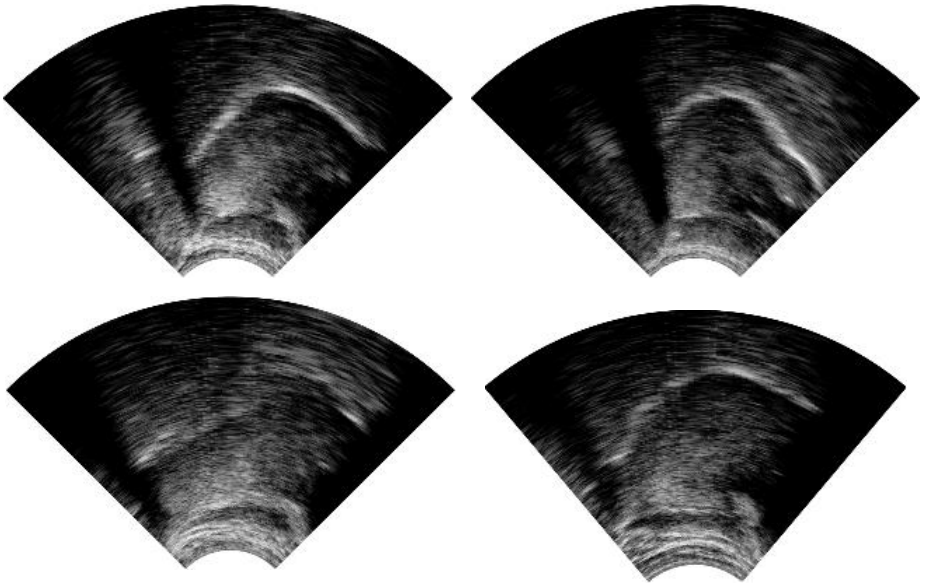

**4. ábra.** Beszédjel és ultrahang-szinkronizáló jel. Az alsó jelben lévő tüskék az egyes ultrahangképek elkészültét jelölik.

# 3 Beszéd- és nyelvultrahang-felvételek

## 3.1 Az ultrahang beszélőfüggősége

Az ultrahangfelvételek képi minősége eltérő lehet az egyes beszélők között. Stone leírja, hogy a fiatal, női, sovány adatközlők artikulációjának ultrahangos rögzítése adja a legjobb képminőséget [13]. Ezt természetesen az artikulációs szervek szöveteinek állapota (pl. hidratáltság) is befolyásolja. Az 5. ábrán négy különböző beszélő azonos beállításokkal készített felvételeiből rögzített képeket láthatunk. A bal oldali sötétebb rész a nyelvcsont helyére, míg a jobb oldali sötétebb rész az állkapocscsont helyére utal (mivel az ultrahang-hullám a csontokon nem tud áthatolni). Látható, hogy a négy különböző beszélő nyelvének felszíne nem egyformán jól látszik. Ennek az is lehet az oka, hogy a rögzítősisak különböző fejméretek esetén máshogy (más orientációban) tartja az ultrahang-vizsgálófejet.

Természetesen a szoftver lehetőséget ad az ultrahangos hardver paramétereinek (pl. vizsgálófej frekvenciája, látómező, mélység, dinamikatartomány, vonalsűrűség stb.) állítására, ez azonban nem minden beszélő esetében kínál elégséges megoldást.



**5. ábra.** Beszélőnkénti eltérések a nyelvultrahang képeken.
Bal felső: 42 éves nő, jobb felső: 29 éves nő, bal alsó: 31 éves férfi, jobb alsó: 33 éves nő.

A képek minősége befolyásolja a nyelvkontúrkövetéshez használt szoftver teljesítményét is. Automatikus követésre maga az AAA szoftver is kínál lehetőséget, emellett számos más program is rendelkezésre áll. Ilyen például az EdgeTrak [23], a ToungeTrack [24], és az AutoTrace [25] szoftver; illetve a legújabb nyelvkontúrkövető módszerek is használhatóak [26]. Ezen szoftverek között eltér

például, hogy igényelnek-e előzetes manuális betanító adatbázist, avagy képi hasonlóságon alapulnak (összehasonlítás: [19]).

Mindebből következik, hogy a kutatások megtervezésének egyik elengedhetetlen lépése a megfelelő beállítások és a precíz nyelvkontúrkövető módszerek feltárása.

## 4 Laptopos bemutató

Demonstrációnkban 5 magyar anyanyelvű beszélővel készült ultrahangvideók alapján mutatjuk be az ultrahangos nyelvkontúrkövetés módszereit. A bemutatóban szerepel az ultrahangos mérések képpé alakításának folyamata, specifikációi, a felvételi körülmények ismertetése. Ezután az AAA szoftver működésére térünk rá, különös tekintettel a beállítási lehetőségekre. Emellett a nyelvkontúr követésének korábban említett lehetőségeit ismertetjük, azok előnyeinek, hátrányainak és korlátainak bemutatásával.

## 5 Kutatási tervek

Megkezdett és jövőbeni kutatásaink során egyrészt a koartikulációnak a nyelvmozgásban detektálható mintázatait elemezzük, másrészt az ultrahangos artikulációkövetésnek a képfeldolgozási és beszédtechnológiai alkalmazásban rejlő lehetőségeit kívánjuk feltárni.

Megindultak kutatásaink az artikuláció ultrahangos képi feldolgozási lehetőségei [27] és az artikuláció alapján történő akusztikumbecslés terén [28], illetve megkezdtük az artikulációs tempó magánhangzóejtésre és ennek kontextusfüggő jellemzőire gyakorolt hatásának feltárását.

## Bibliográfia


1. Kempelen, F.: Az emberi beszéd mechanizmusa, valamint a szerző beszélőgépének leírása. Szépirodalmi Könyvkiadó, Budapest. (1989). [Eredeti cím: Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine. (1791)].
2. Lotz, J.: Egy magyar röntgen-hangosfilm és néhány fonológiai kérdés. Magyar Nyelv. 62, 257–266 (1966).
3. Lotz, J.: Hangos röntgenfilm-vetítés a magyar nyelv hangképzéséről. Nyelvtudományi Értekezések. 58, 255–258 (1967).
4. Szende, T.: A magyar hangrendszer néhány összefüggése röntgenográfiai vizsgálatok tükrében. Magyar Nyelv. 70, 68–77 (1974).
5. Bolla, K.: A magyar magánhangzók és rövid mássalhangzók képzési sajátságainak dinamikus kinoröntgenográfiai elemzése. Magyar Fonetikai Füzetek. 8, 5–62 (1981).
6. Bolla, K.: A magyar hosszú mássalhangzók képzése. (Kinoröntgenografikus vizsgálat számítógépppel). Magyar Fonetikai Füzetek. 8, 7–55 (1981).
7. Bolla, K., Földi, É., Kincses, G.: A toldalékcső artikulációs folyamatainak számítógépes vizsgálata. Magyar Fonetikai Füzetek. 15, 155–165 (1985).
8. Bolla, K.: Magyar fonetikai atlasz. A szegmentális hangszerkezet elemei. Nemzeti Tankönyvkiadó, Budapest (1995).

9. Mády, K.: Magyar magánhangzók vizsgálata elektromágneses artikulográffal normál és gyors beszédben. Beszédkutatás 2008. 52–66 (2008).
10. Deme, A., Greisbach, R., Markó, A., Meier, M., Bartók, M., Jankovics, J., Weidl, Z.: Tongue and jaw movements in high-pitched soprano singing: A case study. Beszédkutatás 2016 [Speech Research 2016]. 24, 121–138 (2016).
11. Öhman, S., Stevens, K.: Cineradiographic studies of speech: procedures and objectives. J. Acoust. Soc. Am. 35, 1889 (1963).
12. Stone, M., Sonies, B., Shawker, T., Weiss, G., Nadel, L.: Analysis of real-time ultrasound images of tongue configuration using a grid-digitizing system. J. Phon. 11, 207–218 (1983).
13. Stone, M.: A guide to analysing tongue motion from ultrasound images. Clin. Linguist. Phon. 19, 455–501 (2005).
14. Schönle, P.W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., Conrad, B.: Electromagnetic articulography: use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. Brain Lang. 31, 26–35 (1987).
15. Baer, T., Gore, J., Gracco, L., Nye, P.: Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels. J. Acoust. Soc. Am. 90, 799–828 (1991).
16. Woo, J., Murano, E.Z., Stone, M., Prince, J.L.: Reconstruction of high-resolution tongue volumes from MRI. IEEE Trans. Biomed. Eng. 59, 3511–3524 (2012).
17. Cheah, L.A., Bai, J., Gonzalez, J.A., Ell, S.R., Gilbert, J.M., Moore, R.K., Green, P.D.: A user-centric design of permanent magnetic articulography based assistive speech technology. In: Proc. BioSignals. pp. 109–116 (2015).
18. Gonzalez, J.A., Moore, R.K., Gilbert, J.M., Cheah, L.A., Ell, S., Bai, J.: A silent speech system based on permanent magnet articulography and direct synthesis. Comput. Speech Lang. 39, 67–87 (2016).
19. Csapó, T.G., Lulich, S.M.: Error analysis of extracted tongue contours from 2D ultrasound images. In: Proc. Interspeech. pp. 2157–2161. , Dresden, Germany (2015).
20. Csapó, T.G., Csopor, D.: Ultrahangos nyelvkontúr követés automatikusan: a mély neuronhálókon alapuló AutoTrace eljárás vizsgálata. Beszédkutatás 2015. 177–187 (2015).
21. Wrench, A.: Ultrasound speech analysis: State of the art. In: Ultrafest VI. , Edinburgh, UK (2013). http://materials.articulateinstruments.com/Technical/State_of_Art.ppt
22. Németh, G., Olaszy, G. eds: A MAGYAR BESZÉD; Beszédkutatás, beszédtechnológia, beszédinformációs rendszerek. Akadémiai Kiadó, Budapest (2010).
23. Li, M., Kambhamettu, C., Stone, M.: Automatic contour tracking in ultrasound images. Clin. Linguist. Phon. 19, 545–554 (2005).
24. Tang, L., Bressmann, T., Hamarneh, G.: Tongue contour tracking in dynamic ultrasound via higher-order MRFs and efficient fusion moves. Med. Image Anal. 16, 1503–1520 (2012).
25. Hahn-powell, G. V., Archangeli, D., Berry, J., Fasel, I.: AutoTrace: An automatic system for tracing tongue contours. J. Acoust. Soc. Am. 136, 2104 (2014).
26. Xu, K., Gábor Csapó, T., Roussel, P., Denby, B.: A comparative study on the contour tracking algorithms in ultrasound tongue images with automatic re-initialization. J. Acoust. Soc. Am. 139, EL154-EL160 (2016).
27. Xu, K., Roussel, P., Csapó, T.G., Denby, B.: Convolutional neural network-based automatic classification of midsagittal tongue gestures using B-mode ultrasound images. submitted to J. Acoust. Soc. Am. Express Lett., (2016).
28. Csapó, T.G., Grósz, T., Tóth, L., Markó, A.: Beszédszintézis ultrahangos artikulációs felvételekből mély neuronhálók segítségével. In: Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017), Szeged, Magyarország, (2017).