

Kriminalisztikai alapú beszélőprofil-alkotás

Beke András

Spicy Analytics Kft.
1026 Budapest, Harangvirág utca 5.
{beke.andras}@gmail.com

Kivonat: A beszélő hangja alapján az ismert személyek felismerése mellett képesek vagyunk az ismeretlen személyekről profilt készíteni, vagyis olyan általános információkat becsülni, mint például a nem, az életkor, a testalkat vagy a beszélő hangulata. Korábbi kutatások igazolták, hogy erős összefüggés van a toldalékcso hossza és a beszélő személy fizikai állapota, mint az életkor, a nem, a testmagasság stb. között. Ezen összefüggés alapján feltételezzük, hogy az emberi beszéd akusztikai jellemzői kódolják az adott beszélő testi fizikai felépítésére utaló jegyeket. A jelen kutatásban ezen összefüggés érvényességét vizsgáljuk tanuló algoritmusok segítségével. A kutatásban elemezzük, hogy a beszédből milyen eredményességgel lehet automatikusan becsülni a beszélő nemét, életkorát, testsúlyát, illetve testtömegét. A fizikai tulajdonságok becsüléséhez a beszédből kinyert akusztikai jellemzőket használunk: prozódiai alapú, beszédminőség-alapú, spektrális alapú. Az eredmények azt mutatják, hogy a nem, a testtömeg és a testsúly becslése nagy pontosságú, míg az életkor becslése kevésbé.

1 Bevezetés

Az elmúlt években láthatóan megnövekedett az érdeklődés a hangalapú biometria iránt, ami ma már szintén kiegészítő eleme a személyazonosításnak, beléptető, illetve felügyeleti eszközöknek stb. Ugyanakkor fontos megjegyezni, hogy a hangalapú biometria korántsem rendelkezik olyan megbízhatósági szinttel, mint például a DNS, az ujjlenyomat vagy íriszminta. Ezzel szemben a hangalapú biometria alkalmazásának költsége jóval alacsonyabb és gyorsabb, mint a fent említetteké, így mind a gyakorlat, mind a tudományos szakma egyre nagyobb figyelmet fordít ennek fejlesztésére. A kurrens hangalapú biometriai eszközök közé tartozik a beszélőazonosítás, a beszélőhitelesítés. A beszélőazonosítás során egy n -elemű halmazból (általában a gyanúsítottak köréből) kell kiválasztani a legvalószínűbb beszélőt (vagyis a kérdéses mintát hozzárendelni a gyanúsítottak valamelyikéhez, vagy kizárni azt a csoportból). A beszélőhitelesítéskor azt a hipotézist kell ellenőrizni, hogy az adott beszélő az a személy-e, akinek állítja magát. A közös mindkettőben az, hogy az adott/vizsgált beszélői modell ismert személytől származik. Ugyanakkor a kriminalisztikai gyakorlatban sokszor olyan esetek is felmerülnek, amikor a beszélő személy kiléte nem ismert, illetve a hatóság nem rendelkezik gyanúsítottal körrel. A hangalapú biometriai eljárások közül a beszélőprofil-alkotás foglalkozik ennek a kérdések a

vizsgálatával. A beszélőprofil-alkotás a kriminalisztika egy olyan ága, amely deduktív érveléseket használ, hogy logikus következtetéseket vonjon le az elkövetőről annak beszédbeli és/vagy nyelvi ismérvei alapján. Ez a tudomány arra a megfigyelésre épül, hogy a beszélő hangja alapján az ismert személyek felismerése mellett képesek vagyunk az ismeretlen személyekről profilt készíteni, vagyis olyan általános információkat becsülni, mint a nem [14], az életkor [7,20], a testalkat [2,8,24] vagy a beszélő hangulata [22] stb. Korábbi kutatások igazolták, hogy erős összefüggés van a toldalékcso hossza és a beszélő személy fizikai állapota, mint az életkor, a nem, a testmagasság stb. között [4,5]. Ezen tényre alapulva számos olyan kutatás született, amely a beszédből automatikusan kívánta becsülni a beszélő fizikai tulajdonságait annak beszédbeli jellemzői alapján.

A nemzetközi szakirodalomban igen sok kutatás történt a beszélő nemének és életkorának automatikus becslésére annak beszéde alapján. Mindemellert van néhány olyan munka, amely a beszélő testsúlyának, illetve testmagasságának becsélét vizsgálja a beszélő hangja alapján. A hazai beszélői profilalkotás előzményeinek tekinthető: [7,8]. Ezen kutatások azonban a jelenség percepció oldalát vizsgálták, vagyis hogy a humán hallgató milyen eredményességgel képes a beszélő fizikai paramétereinek becslésére. A következőkben bemutatjuk az egyes beszélői fizikai paraméterek becslésére történt kísérleteket.

1.1 A beszélő nemének automatikus osztályozása

A beszéd elsődleges információt hordozhat a beszélő nemére vonatkozóan. Ennek háttérben az áll, hogy a beszéd létrehozása biológiailag meghatározott. A nők és a férfiak fiziológiai jellemzőiket – többek között a beszédszervek méretét – tekintve eltérőek. A hangszalagok, valamint az artikulációs csatorna hossza kisebb a nők esetében (a hangszalagok átlagos hossza: 1,5 és 2 cm közötti, az artikulációs csatorna átlagos hossza: 14 cm; térfogata 130 cm³), mint a férfiaké (a hangszalag átlagos hossza: 1,9 és 2,9 cm közötti, az artikulációs csatorna átlagos hossza: 17 cm, térfogata 170 cm³). Az első és legdominánsabb akusztikai következménye ennek, amely hallás alapján is jól feldolgozható, hogy a nők hangmagassága magasabb, mint a férfiaké. Jóllehet a nemek azonosítása beszédhang alapján (akár percepciósan, akár gépi úton) egyszerű feladatnak tűnhet, mégis sok esetben téves eredményt adhat. Nehezíti a feladatot az is, hogy az életkor előrehaladásával a nők hangja mélyül, a férfiaké pedig magasabbá válik. A nemek felismerését nehezítik továbbá az alkoholfogyasztás, a dohányzás, illetve a betegségek is. A fiziológiai különbözőségeken kívül a nők és a férfiak nyelvhasználata is eltérő lehet, amely a szóhasználatban, a létrehozott szöveg nyelvtani összetettségében, a pragmatikai tényezőkben, avagy a nonverbális kommunikációban mutatkozhat meg [12]. A nők például gyakrabban használják a visszakérdézést (*ugye?*, *igaz?*, *hát nem?* stb.) a szolidaritás és az udvariasság eszközeként, mint a férfiak.

1.2 A beszélő életkorának becslése

Az életkor előrehaladtával a beszédszervek változáson mennek keresztül (például a tüdőkapacitás, a gégenagyság, a fogazat és a nyelv izomzata is változik), amelynek következtében jelentősen módosulhat a beszéd hangzása. Ez a változás igen gyakran a beszélő hangmagasságát, artikulációs és beszédtempóját, hangerejét stb. érinti. A nemhez hasonlóan, az életkor is befolyásoló tényező lehet a szóhasználat, a létrehozott szöveg nyelvtani összetettségét, pragmatikai tényezőit, avagy a nonverbális kommunikációt illetően. A beszédkutatók három fő életkori szakaszra irányulnak: gyermekkor (18 éves korig), felnőttkor (18–65 éves kor között), időskor (65 éves kortól). A kutatási eredmények szerint azonban a hallgató az életkor vonatkozásában ennél jóval pontosabb becslést is képes adni, közel 70%-os pontossággal tudja a beszélő életkorát megbecsülni [24]. Gyermekkorban a beszéd a felnőttkori beszédetől az alábbi tényezőkben tér el: az alaphangmagasság magasabb (gyermekkorban a legmagasabb: ekkor a fiúk és lányok között nincs, vagy alig van különbség), a beszéd- és artikulációs tempó lassabb, a beszédhangok kiejtése még nem kiforrott, a beszéd tematikája és grammatikai kidolgozottsága eltérő. Időskorban a beszéd a felnőtt beszédétől az alábbi tényezőkben tér el: a beszéd és artikulációs tempó lassabb, a beszédtervezési nehézségek jellemzőek, a szünettartás gyakoribb, az artikuláció pontatlanabb, az egyes szavak kiejtése hosszabb időtartamú, a hiba típusú megakadások gyakoribbak [3].

1.3 A beszélő testmagasságának becslése

Az elmúlt években szintén több vizsgálatot végeztek a testmagasság becslésére a beszéd alapján. Ezek többsége a humán percepciót tesztelte, vagyis azt, hogy a kutatásban részt vevő hallgatók milyen eredményességgel tudják megítélni a beszélő testmagasságát a hallott beszédminta alapján. Ezek eredményei azonban nem mutatnak egy irányba. Lass [13,14] munkái alapján a hallgatók egyértelműen meg tudják becsülni a beszélők testmagasságát, míg a későbbi munkák ezt nem erősítették meg [10]. Az újabb kutatások szerint [24] csak a férfiak esetében találtak erős korrelációt a testmagasság és a becsült értékek között, míg más vizsgálatban egyáltalán nem tudtak ilyen jellegű összefüggést kimutatni. Számos akusztikai vonatkozású kutatás vizsgálta a toldalékcső hossza és a testmagasság közötti korrelációt. Fitsh és Giedd [5] erős korrelációt mutatott ki a toldalékcső és a testmagasság között röntgen- és mágnesesrezonancia-alapú vizsgálatok alapján. A testmagasság előrejelzésére a számos akusztikai jellemző közül vizsgálták már a szubglottális rezonanciát [1], az alaphangmagasságot [6], a formánsokat [9,21], az MFCC-t és a lineáris predikciós együtthatókat (LPC) [4,18]. Emellett vannak olyan kutatások, amelyek mindezen jellemzőket használják a testmagasság becsléséhez. Mporas és munkatársa [16] például az openSMILE akusztikai jellemzőkinyerő algoritmust használták, amely több mint 6000 akusztikai paramétert tartalmaz. Tanulmányukban vizsgálták a lineáris és nem-lineáris regressziós modellek hatékonyságát is. Az eredményeik azt mutatták, hogy 0,053 méter volt az átlagos

eltérés a valós testmagasság és prediktált testmagasság között, ha bagging regressziós modellt használtak.

1.4 A beszélő testtömegének becslése

A testtömeg és a beszéd közötti korreláció mögött szintén a vokális traktus és az akusztikai jel közötti korreláció adja [13]. De Sousa és munkatársai [23] kimutatták, hogy a testtömegindex megjelenik az akusztikumban is. Vizsgálatukban 84 nő testtömegindexét (19 alultáplált; 23 normál, 20 túlsúlyos and 22 elhízott) és akusztikai paramétereit hasonlították össze (alaphangmagasság, kitartott magánhangzó [a] és fonációs időtartam az [a], az [i] és az [u] hangnak). A legnagyobb eltérést az alaphangmagasságban találták az egyes csoportok között. Poorjam és munkatársai [19] az i -vektoron (Gaussok középértékének összefűzése) nem-negatív mátrixfaktorizálással (Gaussok súlyain alkalmazva) olyan rendszer hoztak létre, amely a NIST telefonos beszédatbázisán a férfiak estében 0,56-os, míg a nők esetében 0,49-es korrelációt mutat az eredeti és a becsült testtömeg között.

2 Adatbázis: leíró statisztikák a vizsgált paraméterek mentén

A jelen kutatásunkat a BEA (BEszélnyelvi Adatbázis [17]) adatbázisból véletlenszerűen kiválasztott 134 adatközlő hanganyagán végeztük. A BEA protokolljából csak a mondatfelolvasást választottuk ki: 25 mondat minden beszélőtől, így 3350 mondat volt a teljes adatbázisunk. A kiválasztott korpuszban a nemek eloszlása nem volt egyenlő: 54 férfi, 80 női beszélőt tartalmazott. A korpusz leíró statisztikáját az 1. táblázat mutatja.

1. Táblázat: A korpusz leíró statisztikája

Faktorok	Férfi		Női	
	Átlag	Átlagos eltérés	Átlag	Átlagos eltérés
Életkor (év)	38	18	42	21
Testmagasság (cm)	178	10	165	5
Testsúly (kg)	79	15	64	13

Az életkori eloszlás esetében jól látható, hogy a résztvevők többségének életkora 20–30 év között mozog (Fig. 1.). A többi életkori cella közel azonos számosságú. A testmagasság közel normál eloszlású, egy kiugróan alacsony testmagasságú adatközlővel (Fig. 1.). A testtömeg balra ferde, de normál eloszlású (Fig.1.).

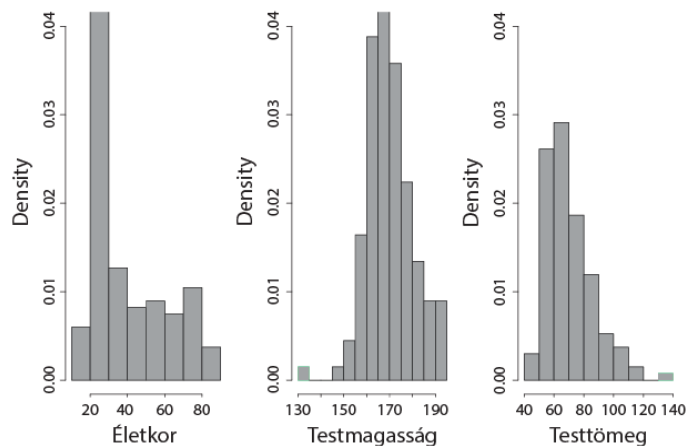


Fig. 1. Az életkor, a testmagasság és a testtömeg eloszlása a korpuszban.

Az egyes testjellemzők közötti korrelációelemzés alapján azt lehet látni, hogy nem meglepő módon a testtömeg és a testmagasság között van közepesen erős korreláció (Fig. 2.). Az életkor és a többi jellemző között viszont nincs ilyen értelemben korreláció.

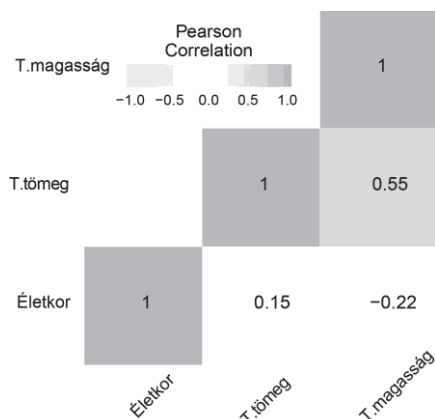


Fig. 2. Korreláció az életkor, a testmagasság és a testtömeg között.

A nemen belül a nők esetében jól látható, hogy a 25 évesek felülreprezentáltak a korpuszban, míg a férfiak esetében ez kevésbé tapasztalható (Fig.3.). Ugyanakkor a nők esetében több olyan adatközlő is van, aki az idősebb korosztályba tartozik.

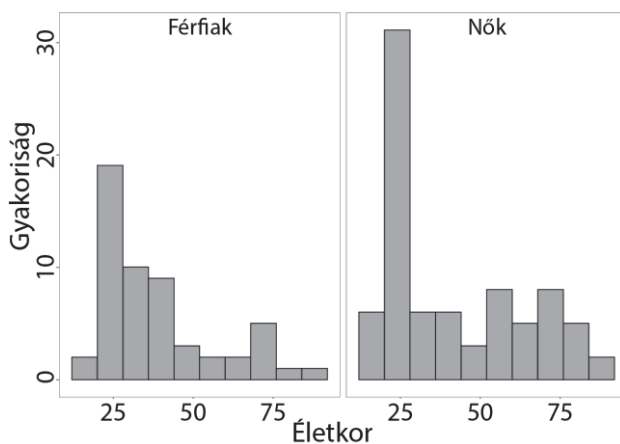


Fig. 3. Az életkori eloszlás a nem tekintetében.

Jelentős különbséget figyelhetünk meg a nők és a férfiak között a testmagasság és a testtömeg esetében is. A férfiak átlagosan 13 cm-el magasabbak, mint a nők (Fig.4.).

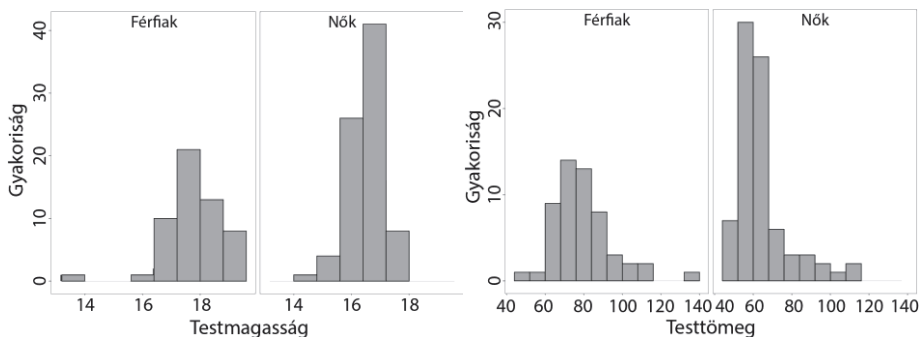


Fig. 4. A testmagasság (balra) és testtömeg (jobbra) eloszlása a nem tekintetében.

A férfiak testtömegüket illetően átlagosan 15 kg-mal súlyosabbak, mint a nők (Fig.4.).

Megvizsgáltuk, hogy a nők és férfiak esetében hogyan korrelálnak az egyes testjellemzők (Fig.5.)

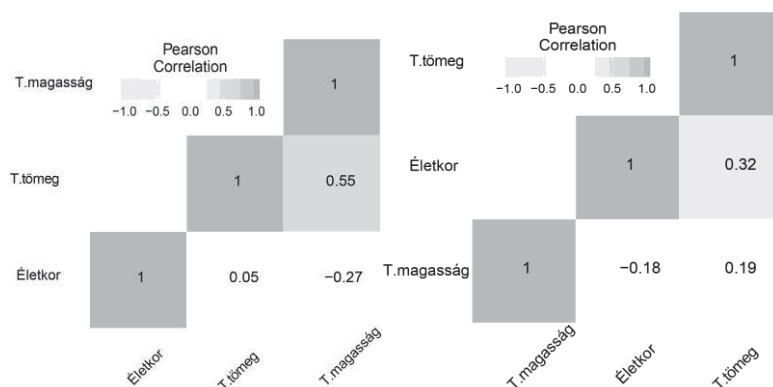


Fig. 5. Korreláció az életkor, a testmagasság és a testtömeg között a férfiaknál (balra) és a nőknél (jobbra).

Látható, hogy a testtömeg és a testmagasság közötti közepes korreláció csak a férfiak esetében igazolható, a nőknél nem. Ugyanakkor a nőknél megfigyelhető egy gyenge korreláció az életkor és a testtömeg között.

Összességében megfigyelhető, hogy az egyes testi jellemzők nem kiegyenlítettek a korpuszban, illetve igen erősen függenek a nemtől, vagyis hogy férfiról vagy nőről van-e szó.

3. Módszer

3.1 Akusztikai jellemzők

Az egyes testi jellemzők osztályozásához, illetve becsléséhez olyan általános jellemzőkészletet kívántunk összeállítani, amely mind a nem, az életkor, a testmagasság és a testtömeg esetében használható. A teljes jellemzőkészletet azonban minden egyes testjellemező esetében specifikáltuk dimenziócsökkentő vagy jellemzőválogató algoritmussal. Az akusztikai jellemzőkinyerést az R programban valósítottuk meg ('seewave', 'sound', 'signal', 'tuneR'), valamint az ezekben nem található függvényeket (HNR, jitter, shimmer) implementáltuk R nyelvre.

3.1.1 Időtartományban mért jellemzők

Tipikusan olyan jellemzők tartoznak ide, mint alaphangmagasságból vagy a hangszínezetből (voice quality) származtatott jellemzők. A jelen kutatásban az alaphangmagasságból számított alapstatisztikai mutatókat vettük ide: átlag, szórás, minimum, maximum, ferdeség, csúcosság. Az alaphangmagasság az egyik legtöbbet idézett akusztikai jellemző, amely összefüggésbe hozható mind a nemmel (a férfiak

f_0 -értéke általában alacsonyabb, mint a nőké), a testmagassággal (az alacsonyabb embereknek általában magasabb az f_0 -értékük, mint a magasabb embereknek [2], illetve a testtömeggel (általában az alultápláltak f_0 értéke magasabb, és akinek magas a testtömegük, azok f_0 értéke alacsonyabb a normál testtömegűekhez képest [23]. Az f_0 mellett olyan akusztikai paramétereket is kerestünk, amelyek utalhatnak a hangszínezetre. Ezek közül kimértük az időbeli rezgésentrópia szintjét, amely a jel rendezettségét kívánja reprezentálni, a nullátmenetek számát (ZRC), az amplitúdó Hilbert-burkolóját, a jittert, a shimmert, a HNR-t, a teljes görbületet (amely az időbeli rezgés második deriváltjának teljes összege), az amplitúdó négyzetes középértékét (RNS), valamint a jel energiáját (annak első két deriváltját, illetve ezekből számított átlagát és szórását).

3.1.2 Frekvenciatartományban mért jellemzők

Minden egyes mondatra kiszámoltuk a 12 koefficienset tartalmazó MFCC-t (0–8 kHz között), illetve az energiát, majd ezeknek első két deriváltját, majd ebből számítottunk alapstatisztikai paramétereket (átlag, szórás, ferdeség, csúcosság: $13 \times 2 \times 4$). A MFCC mellett ugyanilyen módon kiszámoltuk a 12 együtthatót tartalmazó PLP-t is (0–4 kHz között). Továbbá kimértük a frekvencia modulációt is 50–8000 Hz között, amelyből szintén leíró paramétereket származtattunk (átlag, szórás, ferdeség, csúcosság). Ezek mellett vettük a 0–4 kHz közötti spektrum alapvető statisztikai paramétereit (átlag, szórás, medián, kvartilisek, centroid, ferdeség, csúcosság, entrópia, flatness, módusz). Összesen 344 dimenziós sorvektorral jellemeztünk minden egyes mondatot.

2.3 Modellek: klasszifikációs és regressziós modellek

A modellek tanításához és teszteléséhez a teljes adatbázist random módon két részre osztottuk, tanító és teszt adatbázisra (70-30%-ban). A tanító halmazba 1940 mondat került, a teszt halmazba pedig 1410 mondat. Az egyes tanuló algoritmusok hyperparamétereit a tanító adatbázison állítottuk be 5-szörös keresztkiértékeléssel. A modellek tanításához az R szoftvert használtuk ('caret', 'h2o').

2.3.1 Klasszifikáció

A jellemzőkészlet nem-specifikus reprezentáláshoz főkomponens-elemzést (PCA) használtunk. A PCA alapján összesen 74 főkomponenst tartottunk meg, ezzel végeztük az osztályozást. A nem osztályozáshoz bináris logisztikus regressziót, random forest algoritmust, szupport vektor gépet, mély neurális hálózatot és gradient boosting machine-t használtunk. Ezen felül készítettünk egy, a modelleket összevonó ún. ensemble modellt is, amelyben az összevonást végző algoritmus a GLM. Az ensemble esetén kinyertük a tanuló és tesztelő adatokra kapott score-okat az öt egyedi tanuló algoritmussal, vagyis az új adatdimenzió egy 5-dimenziós sorvektor minden egyes esetre. Ezekre az adatokra tanítottunk egy új GLM-et, amely így az egyes tanuló algoritmusok előnyeit emeli ki a nem osztályozási feladatában.

2.3.2 Regresszió

Az életkor, a testtömeg és a testmagasság becsléséhez a teljes jellemzőkészletet használtuk. Az egyes testi jellemzők becsléséhez általános lineáris modellt, random forest-et, gradient boosting machine-et és mély neurális hálózatot használtunk. Ezen felül kísérletet végeztünk a modellek összevonásával, úgynevezett ensemble modellel. Az első ilyen modell esetén mind a négy tanuló algoritmus kimeneti score-ját átlagoltuk. A második ensemble modell esetén a négy kimentre egy újabb általános lineáris modellt tanítottunk.

3 Eredmények

3.1 Az nem osztályozásának eredménye

Az általunk elvégzett klasszifikáció a nem tekintetében igen jó eredményt adott. Ez megegyezik a korábbi kutatások eredményeivel. Mindegyik algoritmussal közel 98-99%-os pontosságot értünk el (Táblázat 2.). Az eredmények azt mutatják, hogy a legjobb osztályozási metrikákat az SVM, a két rejtett réteget tartalmazó DNN, illetve az Ensemble modell adta.

2. Táblázat: A nem osztályozásának eredménye

	Acc	Kappa	AUC
LogisticRegression	0,953	0,902	0,967
RandomForest	0,894	0,772	0,983
SupportVectorMachine	0,966	0,928	0,996
GradientBoosting	0,939	0,872	0,984
DNN [100]-Rectifier	0,953	0,902	0,992
DNN [100-100]-TahnDropout 50%	0,961	0,919	0,993
Ensemble	0,965	0,926	0,996

3.2 Az életkor becslésének eredménye

Az eredményekből az látszik, hogy a feladat lineáris megközelítése a GLM-mel sikertelen volt, ugyanakkor a többi modell eredménye jónak mondható. A becsült életkor közel 10-12 évet téved átlagosan. Az önálló modellek közül a DNN (2 rejtett réteget 200-200 unit, Tanh-függvény) adta a legjobb eredményt. A legjobb eredményt összességében a GLM-mel összevont ensemble modell adta, ebben az esetben volt a legkisebb a hiba: 10 éven belüli (Táblázat 3.).

3. Táblázat: Az életkor becslésének eredménye

	RMSE	Rsquare	MAE
GLM	15,8	0,426	12,6
RandomForest	13,2	0,665	9,8
GBM	12,7	0,634	9,3
DNN	10,2	0,763	7,2
Ensemble_Avg	12,3	0,720	9,3
Ensemble_GLM	9,8	0,770	6,8

A regresszió szerint a nők esetén illeszkednek jobban a valós értékekre a prediktált értékek (0,782), az illeszkedés férfiak estén kisebb (0,768). Ugyanakkor a várt értéktől való eltérés az RMSE és a MAE alapján a férfiak estében kisebb (RMSE: 8,96; MAE: 6,3), mint a nők esetében (RMSE: 10,3; MAE: 7,1) (Fig. 6).

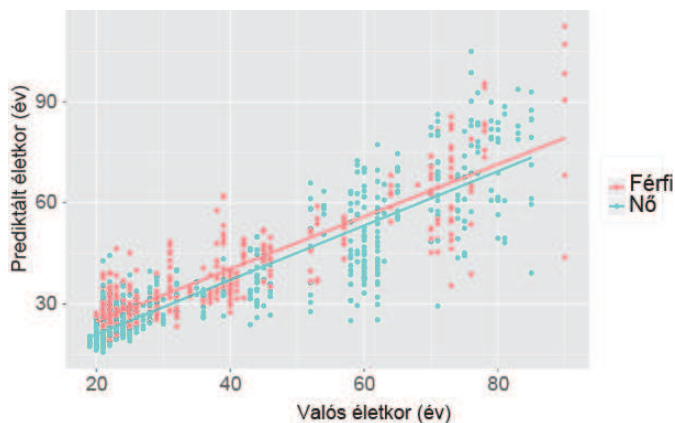


Fig. 6. A valós életkor és a becsült életkor közötti összefüggés a férfiak és nők esetében

3.4 A testmagasság becslésének becslése

A testmagasság becslése az eredmények szerint jobban megközelíthető az általunk készített módszerrel. A GLM-et leszámítva 0,8 feletti korrelációs értéket kaptunk a valós testmagasság, illetve a becsült testmagasság között, emellett az átlagos hiba 3–7 cm között mozgott. A legjobb eredményt itt a DNN-nel, illetve a GLM-mel egyesített modellel kaptuk (Táblázat 4.).

A regresszió szerint a férfiak esetén jobban illeszkednek a prediktált értékek a valós értékekre (0,788), mint a nők estén (0,750). Ugyanakkor a várt értéktől való eltérés az RMSE és a MAE alapján a nők estében kisebb (RMSE: 2,5; MAE: 2,0), mint a férfiak esetében (RMSE: 9,2; MAE: 3) (Fig. 7.).

4. Táblázat: A testmagasság becslésének eredménye

	RMSE	Rsquare	MAE
GLM	6,6	0,500	5,2
RandomForest	4,2	0,814	3,1
GBM	4,3	0,785	3,2
DNN	3,8	0,824	2,9
Ensemble_Avg	4,2	0,812	3,2
Ensemble_GLM	3,2	0,873	2,4

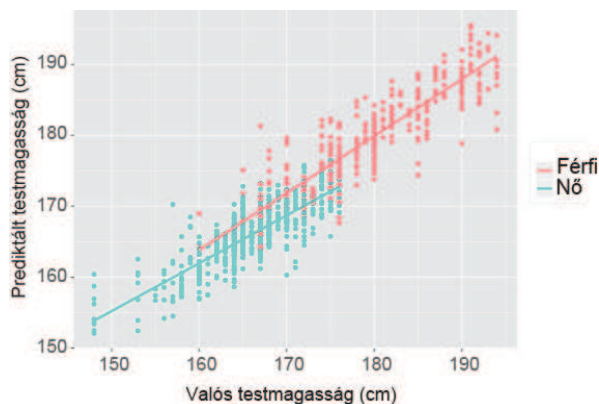


Fig. 7. A valós testmagasság és a becsült életkor közötti összefüggés a férfiak és nők esetében

3.5 A testtömeg becslésének eredménye

A testtömeg becslése szintén nehezebb feladatnak bizonyult. Az eredmények szerint a becsült testmagasság közel 4–6 kg közötti átlagos hibával becsülhető. Ugyanúgy, ahogy az előző regressziós feladatokban, a legjobb eredményt ebben az esetben is a DNN, illetve a GLM-mel egyesített modellel adta (Táblázat 5.).

A regresszió szerint a nők esetén jobban illeszkednek a valós értékekre a prediktált értékek (0,783), mint a férfiak estén (0,754). A becslés pontosság a két esetben közel ugyanolyan mértékű volt (nők: RMSE: 6,7; MAE: 4,8; férfiak: RMSE: 6,5; MAE: 4,8) (Fig. 8.).

4 Következtetések

A jelen kutatásban azt vizsgáltuk, hogy milyen eredményességgel lehet a beszédből becslést adni a beszélő fizikai adottságaira a nem, az életkor a testmagasság, illetve a testtömeg tekintetében. Az eredményeink szerint az olvasott mondatokon a nem

meghatározása igen jó minőségben megvalósítható. Az életkor becslése esetében a feladat jóval komplexebb, hiszen a beszédben észlelhető életkor, illetve a valós életkor között (vö. [15]) sok esetben ugrás észlelhető. Ez a különbség tovább nő az idő előrehaladtával, vagyis a becslés pontossága a fiatalok esetében jobb, mint az idősek esetében (vö. [11]) (lásd Fig. 6). A testmagasság esetében a modell pontossága igen jónak mondható, vagyis az általunk kinyert akusztikai paraméterek hordozzák azon jegyeket, amelyek alkalmasak ezen fizikai paraméter becslésére. Ugyanez elmondható a testtömeg esetén is, jóllehet ennek automatikus becslése valamivel pontatlanabbnak bizonyult.

5. Táblázat: A testtömeg becslésének eredménye

	RMSE	Rsquare	MAE
GLM	12,9	0,359	9,9
RandomForest	8,9	0,706	6,6
GBM	8,6	0,709	6,4
DNN	7,9	0,754	6,0
Ensemble_Avg	8,9	0,750	6,8
Ensemble GLM	6,6	0,817	4,8

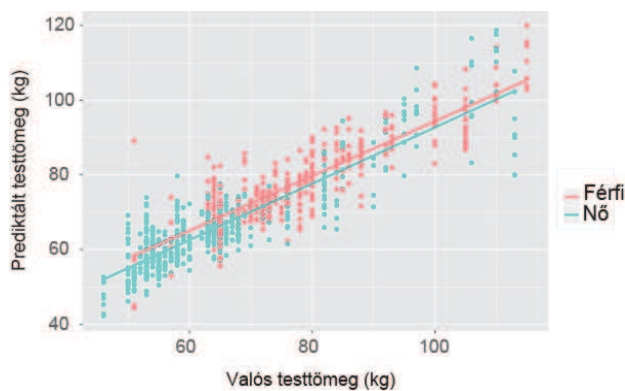


Fig. 8. A valós testtömeg és a becsült életkor közötti összefüggés a férfiak és nők esetében

Ebben a kutatásban egy alap, ún. baseline rendszert hoztunk létre, amely felolvasott mondatokból kinyert akusztikai jellemzők alapján tesz becslést automatikusan a beszélő fizikai paramétereire. Ezen baseline rendszert kívánjuk a jövőben tovább fejleszteni, kiterjeszteni más beszéd típusra, más korpuszra más/komplexebb modellek bevonásával, továbbá az akusztikai jellemzők mellett szöveges jellemzők felhasználásával.

Köszönetnyilvánítás

A jelen kutatás a Bolyai János Kutatási Ösztöndíj támogatásával készült.

Bibliográfia

1. Arsikere, H., Leung, G., Lulich, S., and Alwan, A. (2012). "Automatic height estimation using the second subglottal resonance," in IEEE International Conference on Acoustics, Speech and Signal Processing 2012 (ICASSP) (2012) 3989–3992.
2. Barsties, B., Verfaillie, R., Dicks, P., & Maryn, Y. „Is the speaking fundamental frequency in females related to body height?“. *Logopedics Phoniatrics Vocology*, 41(1), (2016) 27–32.
3. Bóna, J. "A spontán beszéd sajátosságai az időskorban." ELTE Eötvös Kiadó, Bp. (2013).
4. Dusan, S. "Estimation of speakers height and vocal tract length from speech signal," in INTERSPEECH (ISCA) (2005) 1989–1992.
5. Fitch, W. T., and Giedd, J., "Morphology and development of the human vocal tract: A study using magnetic resonance imaging", *J. Acoust. Soc. Amer.* (1999)106(3): 1511–1522.
6. Ganchev, T., Mporas, I., and Fakotakis, N. "Audio features selection for automatic height estimation from speech," in *Lecture Notes in Computer Science. Artificial Intelligence: Theories, Models and Applications*, edited by S. Konstantopoulos, S. Perantonis, V. Karkaletsis, C. Spyropoulos, and G. Vouros (Springer, Berlin) (2010) 6040, 81–90.
7. Gocsál Á. Életkorbecslés a beszélő hangja alapján. *Beszédkutatás* (1998), 122–134.
8. Gósy M. A testalkat és az életkor becslése a beszéd alapján. *Magyar Nyelvőr*; (2001) 25/4: 478–487.
9. Greisbach, R. (1999). "Estimation of speaker height from formant frequencies," *Forensic Ling.* 6, (1999) 265–277.
10. Gunter, C. D., & Manning, W. H. (1982). Listener estimations of speaker height and weight in unfiltered and filtered conditions. *Journal of Phonetics*.
11. Huckvale, M. and Aimee W. "A Comparison of Human and Machine Estimation of Speaker Age." *International Conference on Statistical Language and Speech Processing*. Springer, Cham, 2015.
12. Huszár Á. Bevezetés a gendernyelvészetbe: Miben különbözik és miben egyezik a férfiak és a nők nyelvhasználata és kommunikációja? Budapest: Tinta Könyvkiadó (2009) 123.
13. Lass N. J. and Davis M. "An investigation on speaker height and weight identification," *Journal of the Acoustical Society of America*, vol. 60 (1976) 700–703.
14. Lass N. J, Hughes KR, Bowyer MD, Waters LT, Bourne VT: Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America* (1976) 59:675–678.
15. Moyse, E. "Age estimation from faces and voices: a review." *Psychologica Belgica* 54.3 (2014).
16. Mporas, I., and Todor G.. "Estimation of unknown speaker's height from speech." *International Journal of Speech Technology* 12.4 (2009) 149–160.
17. Neuberger, T., Gyarmathy, D., Grácsi, T. E., Horváth, V., Gósy, M., Beke, A.: Development of a large spontaneous speech database of agglutinative Hungarian language. In: *TSD2014x*. (2014) 424–431
18. Pellom, B. L., and Hansen, J. H. L. "Voice analysis in adverse conditions: The Centennial Olympic Park Bombing 911 call," in *Proceedings of the 40th Midwest Symposium on Circuits and Systems* 2. (1997), 873–876.

19. Poorjam, A. H., Mohamad H. B., and Hugo Van H. "Speaker weight estimation from speech signals using a fusion of the i-vector and NFA frameworks." *Artificial Intelligence and Signal Processing (AISP), 2015 International Symposium on. IEEE (2015).*
20. Ptacek P. H, Sander E. K: Age recognition from voice. *Journal of Speech and Hearing Research (1966) 9/2:273–277.*
21. Rendall, D., Kollias, S., Ney, C., and Lloyd, P. "Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," *J. Acoust. Soc. Am. 117, (2005) 944–955.*
22. Scherer K. R, Banse R, Wallbott H: Emotional inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology (2001) 32/1:76–92.*
23. Souza de, Lourdes Bernadete R., and Marquiony M. dos S. "Body mass index and acoustic voice parameters: is there a relationship." *Brazilian Journal of Otorhinolaryngology (2017).*
24. Van Dommelen, W. A., Moxness, B. H: Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech (1995) 38/3:267–287.*