Enyhe kognitív zavar detektálása beszédhangból x-vektor reprezentáció használatával

José Vicente Egas-López[1], Balogh Réka[2], Imre Nóra[2], Tóth László[1], Vincze Veronika[3], Pákáski Magdolna[2], Kálmán János[2], Hoffmann Ildikó[4], Gosztolya Gábor[1,3]

[1] Szegedi Tudományegyetem, Informatikai Intézet
[2] Szegedi Tudományegyetem, Pszichiátriai Klinia
[3] MTA-SZTE Mesterséges Intelligencia Kutatócsoport, Szeged
[4] Nyelvtudományi Intézet, Budapest
{ egasj, ggabor } @ inf.u-szeged.hu

Kivonat Az enyhe kognitív zavar (EKZ) heterogén klinikai szindróma, melyet gyakran tartanak a demencia preklinikai (azaz a demencia diagnózis felállításához nem elegendő, de mérhető kognitív hanyatlással járó) szakaszának is. Az EKZ jellemzői közé tartozik a kognitív funkciók enyhe hanyatlása, beleértve a memóriát, a végrehajtó és a nyelvi funkciókat. Kutatások alapján a nyelvi funkciók megváltozása már azelőtt észlelhető, hogy az EKZ-ra jellemző egyéb kognitív tünetek megjelennének. Az alanyok beszédének elemzése így praktikus, olcsó és nem-invazív eszköze lehetne a betegség korai szűrésének. Jelen munkánkban egy viszonylag friss, mély neurális hálón alapuló eljárást, az x-vektorokat használjuk jellemzőkinyerésre, majd ezen jellemzőket felhasználva osztályozó eljárást (SVM-et) tanítunk az EKZ-s és a kontroll beszélők elkülönítésére. Kísérleti eredményeink alapján az x-vektorokkal pontosabb diszkrimináció érhető el, mint a hagyományos i-vektorok használatával.
Kulcsszavak: demencia, enyhe kognitív zavar, x-vektorok

1. Bevezetés

A demencia krónikus, progresszív klinikai szindróma, amely főként idős személyeket érint világszerte. Jellemzői közé tartozik a memória, a nyelvi készségek és a problémamegoldó képesség romlása. A fenti készségeket érintő hanyatlás olyan mértékű, hogy az megnehezítheti vagy ellehetetlenítheti a páciens mindennapi tevékenységeinek elvégzését (Alzheimer's Association, 2020). A betegség jelenleg kb. 46,8 millió embert érint világszerte, ez a szám pedig a becslések szerint 2050-re megduplázódhat (Prince és mtsai, 2015). Tekintve, hogy a jelenleg elérhető terápiás beavatkozások a betegség korai szakaszában vagy a betegséget megelőző, preklinikai stádiumban mutatják a legnagyobb hatékonyságot (Szatlóczki és mtsai, 2015), a betegség ezen fázisokban történő, korai felismerése kiemelt fontosságú.

A demencia preklinikai szakaszát a szakirodalom enyhe kognitív zavarnak (EKZ) nevezi. Ez az állapot egyfajta határterületnek tekinthető az öregedéshez

társuló, normális mértékűnek tekinthető szellemi hanyatlás és a már kimutatható demencia között (Petersen és mtsai, 2014). Számos kutatási eredmény utal arra, hogy az EKZ a páciensek beszédkészségére is hatással van – ezekre támaszkodva a beszédelemzés költséghatékony, non-invazív eszközt kínálhat a betegség korai felismerésére. Az utóbbi években számos kutatás jelent meg, olyan eszközök és eljárások bemutatásával, amelyek célja egészséges kogníciójú kontroll (K) személyek és EKZ-val vagy Alzheimer-kórral élő páciensek automatizált módszerrel történő megkülönböztetése volt az alanyok beszédének vizsgálata alapján (lásd de Ipiña és mtsai, 2018; König és mtsai, 2018; Themistocleous és mtsai, 2018; Sluis és mtsai, 2020; Themistocleous és mtsai, 2020).

A szakirodalomban ismertetett eljárások egy részében feladatspecifikus jellemzőket vizsgáltak: olyan paramétereket kerestek tehát, amelyek eltérnek a kontrollszemélyek és az EKZ-s vagy enyhe AK-s alanyok beszédében. Ilyen paraméterek voltak például a szünetek száma és időtartama (Vincze és mtsai, 2020), vagy a beszédtempó és az artikulációs ráta (Meilán és mtsai, 2020). (A jellemzőkinyerést azután természetesen egy standard gépi tanulási lépés követi, például Support Vector Machine-t (SVM) használva.) Egy másik elterjedt megközelítés az, hogy *általános célú* eljárásokat alkalmazva nyernek ki jellemzőket az egyes alanyok hangfelvételeiből. Ezt követően ezeket a jellemzővektorokat felhasználva, statisztikai alapú osztályozó eljárással lehet elkülöníteni a két (vagy esetenként több) beszélőcsoportot. Ilyen általános célú jellemzővektorok lehetnek például az i-vektorok: habár ezeket eredetileg beszélőfelismerés céljára fejlesztették ki, később sikerrel alkalmazták a Parkinson-kór (García és mtsai, 2017; García és mtsai, 2018) és az Alzheimer-kór (Weiner és Schultz, 2018; Egas-López és mtsai, 2019) felismerésére is.

A beszélőfelismerés területén a korábban a legkorszerűbb technikának tartott i-vektorok helyét az utóbbi években egy mély neurális hálóra (Deep Neural Network, DNN) épülő eljárás, az *x-vektorok* vették át (Snyder és mtsai, 2018). A mély tanulás térhódítását tekintve ez nem is meglepő. Ésszerűnek tűnhet, hogy az i-vektorok után az x-vektorokat is alkalmazni kezdik az orvostudományi beszédfeldolgozás területén, vagy a technikai értelemben valamennyire rokon témakörnek számító paralingvisztikai feladatok esetén. Eddig ugyanakkor elég kevés ilyen tanulmány jelent meg: orvostudományi területen csak Botelho és munkatársai, valamint Jeancolas és munkatársai tanulmányairól van tudomásunk. Mindkét fönt említett kutatócsoport a Parkinson-kór felismerésére alkalmazott x-vektorokat (Botelho és mtsai, 2020; Jeancolas és mtsai, 2020) (és mindkét tanulmány csak arXiv preprintként érhető el jelenleg).

Jelen cikkünkben azt vizsgáljuk, hogy milyen hatékonysággal alkalmazhatóak az x-vektor beágyazások az EKZ fölismerésére. Snyder és munkatársai egy előre tanított, letölthető modellt (DNN-t) is a közösség rendelkezésére bocsátottak; emellett a cikkben saját modellel is kísérletezünk, 60 órányi magyar nyelvű spontán beszédre tanítva. Az x-vektorok a háló több rétegéből is kinyerhetők, melyek hatékonyságát szintén megvizsgáljuk, az elért pontosságértékeket pedig összevetjük az i-vektorok használatával elért eredményekkel.

Réteg	**Réteg közvetlen környezete**	**Teljes környezet mérete**	**Be- és kimenetek száma**
Keret #1	[t-2, t+2]	5	120, 512
Keret #2	{ t-2, t, t+2 }	9	1536, 512
Keret #3	{ t-3, t, t+3 }	15	1536, 512
Keret #4	{ t }	15	512, 512
Keret #5	{ t }	15	512, 1500
Összegző	[0, T }	T	1500T, 3000
Szegmens #6	{ 0 }	T	3000, 512
Szegmens #7	{ 0 }	T	512, 512
Szoftmax	{ 0 }	T	512, N

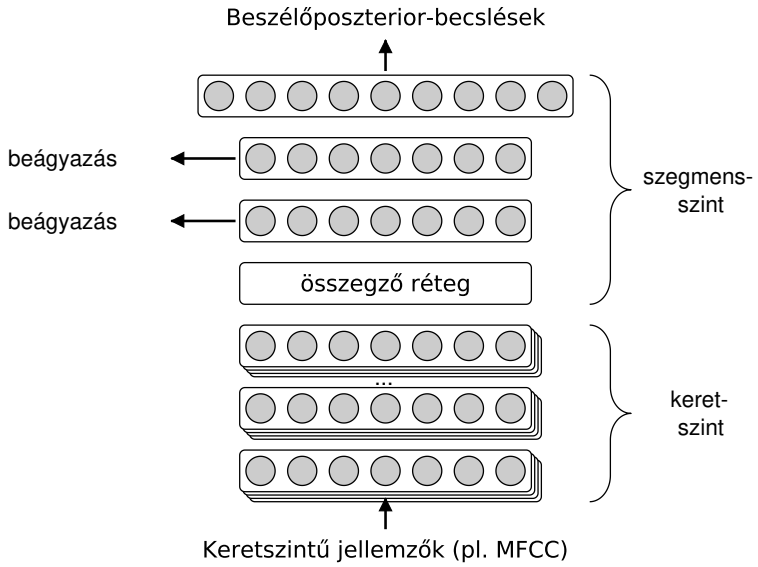
1. táblázat. Az x-vektor rendszer DNN architektúrája, mely öt keretszintű rétegből, egy statisztikai összegző (pooling) rétegből, két szegmensszintű rétegből és egy kimeneti, szoftmax rétegből áll. N-nel jelöltük a háló betanítására használt adatbázis beszélőinek számát (azaz a szoftmax réteg neuronjainak számát). Jelen architektúra megegyezik a Snyder és munkatársai által leírttal (Snyder és mtsai, 2017).

2. X-vektor kódolás

Az x-vektor technika egy neurálisháló-alapú jellemzőkinyerő eljárás, mely a változó hosszú hangfelvételeket fix dimenziószámú jellemzőtérbe képzi. Technikailag egy mély neurális hálóról van szó, melynek bemenetei keretszintű vektorok (például MFCC-k), mélyebben elhelyezkedő rejtett rétegei keret-, magasabb rejtett rétegei pedig szegmensszinten működnek. Az egyes bemondásokhoz tartozó beágyazásokat (azaz az x-vektorokat) a szegmensszintű rétegek aktivációi jelentik.

A legelterjedtebb x-vektor architektúrát Snyder és munkatársai vezették be (Snyder és mtsai, 2018). Ebben a keretszintű rejtett rétegek időeltolásos (*time-delay*) módon működnek: például a második keretszintű rejtett réteg t. kerethez tartozó aktivációinak meghatározásához az első keretszintű réteg három kerethez tartozó aktivációját (t−2, t és t+2) használjuk bemenetként. (Ld. 1. táblázat.) Az ötödik keretszintű réteg után egy összegző réteg (*stats pooling layer*) következik: ennek bemenete az utolsó keretszintű réteg *az aktuális felvétel összes keretén számítva*). (A 1. táblázatban a felvétel kereteinek számát T-vel jelöltük.) Az összegző réteg ezen aktivációk átlagát és szórását számolja ki; ezen két, 1500-1500 elemű vektor egymás után fűzve adja az első szegmensszintű réteg bemenetét. Ezen ponttól kezdve a háló szegmensszintűként működik tovább. A kimeneti, szoftmax réteg a tanító halmazban található beszélők számának megfelelő számú neuront tartalmaz (Snyder és mtsai, 2017, 2018).

A háló tanítása, a fönti struktúrát kihasználva, nem keret-, hanem szegmensszinten történik; címkeként az adott felvétel beszélőjének azonosítóját használjuk (mondjuk keresztentrópia veszteségfüggvénnyel). Tanítás után a beágyazások kinyerésére praktikusan bármelyik szegmensszintű réteg alkalmas; a tapasztalatok alapján a hatodik (a kimeneti rétegtől távolabb eső) réteg aktivációinak használata jobb eredményekhez vezet, mint a hetedik rétegé (Snyder és mtsai, 2018).



1. ábra: Az x-vektort kinyerő mély neurális háló általános struktúrája (Snyder és mtsai, 2018, nyomán).

Megjegyeznénk még, hogy a kimeneti réteg kizárólag tanításkor kap szerepet, így a későbbiekben ez el is dobható.

3. Kísérleti körülmények

3.1. Az EKZ-s és kontroll alanyok felvételei

A EKZ felismerésére vonatkozó kísérleteinket saját hangadatbázison végeztük, melyet a Szegedi Tudományegyetem Pszichiátriai Klinikáján rögzítettünk. A rögzítés digitális diktafonnal történt, külső mikrofon használatával; a felvételeket utólag monó, 16 kHz-es mintavételezésű formátumra konvertáltuk. Az alanyok spontán beszédét rögzítettük, az instrukció a következő volt: „Kérem, részletesen mesélje el az előző napját!”. (A felvételekről további részletekért ld. Vincze és mtsai, 2020). Az elkészült felvételekből hangminőség alapján válogattunk; jelen tanulmányunkban 25 EKZ-s és 25 kontroll alany felvételeit használtuk fel. A két csoport életkorbeli, nembeli és (elvégzett iskolai években mért) képzettségbeli eloszlása nem mutatott statisztikailag szignifikáns különbséget. Sajnos a felvételi körülmények miatt sok bemondás még a válogatás ellenére is visszhangos vagy jelentős háttérzajjal rendelkező volt; a jel-zaj-arány (Signal-to-Noise Ratio, SNR) 14 és 35 dB közé esett.

3.2. Keretszintű jellemzők

Keretszintű jellemzőkészletként standard MFCC vektorokat használtunk. 20 MFCC együtthatót nyertünk ki a felvételekből 25 milliszekundum hosszú keretekből, 10

milliszekundumos lépésközzel, a Kaldi eszköz segítségével (Povey és mtsai, 2011), melyhez hozzátettük még a lokális energiát mint jellemzőt. Bár terveink között szerepelt az egyéb típusú (keretszintű) jellemzők kipróbálása, az eszköz korlátai miatt nem volt lehetőségünk sem frekvencia-szűrősorok energiaértékeinek („filter banks") használatára, sem az első- vagy másodrendű deriváltak felhasználására.

3.3. Az x-vektor DNN-ek tanítása

Saját x-vektor-kinyerő neurális háló modelljeinket a BEA Spontánbeszéd-adatbázis egy részhalmazán tanítottuk (Neuberger és mtsai, 2014). 165 beszélőt választottunk ki; a felvételekből automatikusan kivágtuk azokat a részeket, melyekben a felvételvezető hangja is hallható, így 10636 hangfelvételt kaptunk, összesen körülbelül 60 órányi terjedelemben. Az eredeti sztereó, 44,1 kHz-en mintavételezett bemondásokat monó, 16 kHz-es formátumra konvertáltuk.

A DNN modelleket a Kaldi rendszerrel (Povey és mtsai, 2011) tanítottuk be, és ezt használtuk a jellemzővektorok kinyerésére is. A tanítás során szokásos eljárás a tanító adat méretét mesterségesen megnövelni úgy, hogy az eredeti hangfelvételekhez zajt adnak és/vagy visszhangosítják azokat (Snyder és mtsai, 2018). Mivel az EKZ-s és kontroll alanyainktól gyűjtött hangfelvételeink eleve elég rossz minőségűek voltak, míg a BEA adatbázis szinte stúdióminőségű felvételeket tartalmaz, kíváncsiak voltunk, hogy ez a fajta augmentáció segíti-e az osztályozási lépést. Emiatt két DNN modellt tanítottunk: egyet augmentációval, egyet pedig ennek a lépésnek a kihagyásával. (Az augmentáció 52636 felvételre (293 órányira) növelte a tanítóanyag méretét.)

A fönti két saját modell mellett kipróbáltunk egy előre betanított, szabadon elérhető modellt is (Snyder és mtsai, 2018). Ez a modell a Switchboard 2 Phases 1, 2 és 3, a Switchboard Cellular, valamint a NIST SRE adatbázisokon lett tanítva (összesen kb. hétezer beszélőn). További kisebb eltérés, hogy ez a modell 23 dimenziós MFCC-n (plusz az energián) működik, míg az általunk tanítottak, az i-vektoroknál szokásos méretet követve, 20 dimenzióson (korábbi tesztjeinkben azonban nem találtunk különbséget a 20 és a 23 dimenziós MFCC-t használó modellek között).

3.4. Jellemzőkinyerés

A beszélőtípusok azonosítására jellemzővektorként használt x-vektorokat a fönt ismertetett három DNN modellből nyertük ki. Az 1. táblázatban leírt struktúrájú DNN-ből három ponton nyerhető ki felvételszintű vektor: az általában használt *Szegmens #6* réteg mellett a *Szegmens #7* és az *Összegző* réteg is alkalmas arra, hogy aktivációit (felvételszintű) jellemzőként használjuk. (Ezekben a rétegekben sorban 512, 512 és 3000 neuron található, így ezeknek megfelelő méretű jellemzővektort kapunk.) Mivel kíváncsiak voltunk, hogy enyhe kognitív zavar detektálására melyik réteg a legalkalmasabb, kísérleteinkben összesen kilenc variációt (három DNN modell és három réteg) próbáltunk ki. Emellett viszonyítási alapként i-vektorokat is használtunk jellemzőkként (128 komponenst

alkalmazva); az i-vektorok általános háttérmodellje (Universal Background Model, UBM) az összehasonlíthatóság érdekében szintén a BEA adatbázis 3.3. fejezetben bemutatott részhalmazán lett tanítva. Az i-vektorok kiszámítására is a Kaldi rendszert használtuk.

3.5. Beszélőosztályozás

A jellemzőkinyerési lépés után a beszélőket Support Vector Machine (SVM, Schölkopf és mtsai, 2001) alkalmazásával, ötszörös keresztvalidációval osztályoztuk, a Python scikit-learn csomagját (Pedregosa és mtsai, 2011) használva. Minden SVM modell tanítása 20 EKZ-s és 20 kontroll alany hangfelvételén történt. Kiértékelési metrikáink a következők voltak: osztályozási pontosság (classification accuracy, *Pont.*), pontosság (precision, *Prec.*), fedés (recall), F_1-érték (F-measure), valamint a ROC görbe alatti terület (AUC). (Pontosság (precision), fedés és F_1 esetén az EKZ beszélőkategóriát tekintettük pozitív osztálynak; mivel csak két beszélőkategóriánk (EKZ és kontroll) volt, a két osztályra kapott AUC-értékek megegyeztek.) A túltanulás elkerülése érdekében lineáris kernelt használtunk, így egyetlen hiperparaméterünk az SVM C (complexity) értéke volt; ezt az $10^{-5}, 10^{-4}, \ldots, 10^{1}$ értékeket végigpróbálva (grid search), a legnagyobb AUC értéket megcélozva választottuk ki. Előzetes tesztjeink eredményeit követve az x-vektorok esetén nem volt szükség a vektorok standardizálására vagy normalizálására, míg az i-vektorokat standardizáltuk (azaz minden jellemzőt nulla átlagra és egység szórásra transzformáltunk).

4. Eredmények

A 2. táblázat tartalmazza a különböző DNN modellekből és rétegekből kinyert x-vektor jellemzőket használva kapott pontosságértékeket. Látható, hogy a három használt DNN-réteg közül mindig az irodalomban általában ajánlott *Szegmens #6*-os réteg használatával kinyert jellemzőkkel kaptuk a legjobb eredményeket. Ennek oka feltehetőleg az, hogy az összegző réteg még nem foglalja össze a keretszintű információkat elég precízen, míg az utolsó rejtett réteg (*Szegmens #7*) már túlságosan feladatspecifikus információkat tárol (azaz túl specifikus a tanítóhalmazban szereplő beszélőkre).

A BEA adatbázison tanított két DNN modell közül az augmentálás használata valamivel jobb eredményekhez vezetett. Ez feltehetőleg annak köszönhető, hogy az augmentálási lépés amellett, hogy megnöveli a tanítóadat mennyiségét, zajtűrőbbé is teszi a modellt (mivel az extra tanítóadat az eredeti felvételek zajosított, illetve visszhangosított változataiból áll), amely hasznosnak bizonyulhat, amennyiben az osztályozandó felvételek nem éppen ideális körülmények között lettek rögzítve. Figyelembe véve, hogy a legtöbb beszédtechnológiai alkalmazás esetén nem várhatunk el stúdióminőséget, a modell felkészítése a zajos körülményekre mindenképpen a gyakorlati használhatóság felé tett lépés, melyet akár az i-vektorok háttérmodelljének (az UBM-nek) a tanítása során is érdemes lenne alkalmazni. (Sajnos itt megint könnyű technikai akadályokba ütközni.)

Tanító adatbázis	Jellemző-kinyerési réteg	Pontosságértékek				
		Pont.	Prec.	Fedés	F_1	AUC
BEA (augmentáció nélkül)	Összegző	58%	60,0%	48,0%	53,3%	0,562
	Szegmens #6	64%	65,2%	60,0%	62,5%	0,628
	Szegmens #7	56%	57,1%	48,0%	52,2%	0,576
BEA (augmentálva)	Összegző	60%	63,2%	48,0%	54,5%	0,595
	Szegmens #6	64%	68,4%	52,0%	59,1%	0,645
	Szegmens #7	58%	61,1%	44,0%	51,2%	0,602
Előtanított modell	Összegző	62%	63,6%	56,0%	59,6%	0,640
	Szegmens #6	70%	72,7%	64,0%	68,1%	0,673
	Szegmens #7	58%	61,1%	44,0%	51,2%	0,527
i-vektor (BEA, augmentáció nélkül)		60%	63,2%	48,0%	54,5%	0,597

2. táblázat. A különböző x-vektorok, valamint a viszonyítási alapként megvizsgált i-vektorok használatával EKZ-azonosításra kapott kiértékelési metrikák. (Pont.: osztályozási pontosság; Prec.: pontosság (precision).)

A három modell közül a legjobb eredményekhez az előretanított modell használata vezetett. Ez véleményünk szerint egyrészt azt támasztja alá, hogy az x-vektorok a gyakorlatban (legalábbis ezen nyelvek esetén) nyelvfüggetlen módon képesek kódolni a beszélőket. Másrészt azt is jelzi, hogy még hatvan órányi felvétel (illetve 165 beszélő) sem képes azt a varianciát nyújtani, amely kellően robosztus x-vektor beágyazások kinyerését lehetővé tevő DNN-ek tanításához szükséges. Kétségtelen, hogy a Snyder és munkatársai által használt korpusz a mintegy hétezer beszélővel lényegesen változatosabb, mint amit akár a teljes BEA adatbázissal (tehát 500 beszélővel) lehetséges lenne elérni.

Összességében elmondható, hogy a kapott pontosságértékek nem különösebben magasak: még a legjobb esethez is csupán 70%-os osztályozási pontosság, és 0,673-es AUC érték tartozik. Ez véleményünk szerint elsősorban a felvételek zajosságának tudható be: a 14 és 35 dB közé eső SNR elég alacsonynak mondható (viszonyításképpen: egy hagyományos analóg telefonvonalhoz 40 dB-es érték tartozik (Aude, 1998)). Ugyanakkor még ezen hátráltató tényező ellenére is jobban el tudtuk különíteni az enyhe kognitív zavarral rendelkező alanyokat az egészséges kontroll személyektől az x-vektorok használatával, mint az i-vektorokra építve.

5. Összegzés

Az enyhe kognitív zavar egy krónikus klinikai szindróma, melynek korai detektálása kulcsfontosságú a kezelés minél hamarabb történő megkezdéséhez. Jelen tanulmányunkban egy viszonylag új jellemzőkinyerési eljárást, az x-vektorokat teszteltük ebben a feladatban. Az x-vektort szolgáltató mély neurális hálókat a

BEA adatbázis egy 60 órás részhalmazán, 165 beszélő adatain tanítottuk két variációban (zaj hozzáadásával és anélkül), valamint egy angol beszédre előtanított modellt is kipróbáltunk. Az x-vektorokat a DNN modellek több rejtett rétegéből is kinyertük.

Eredményeink alapján az x-vektorok valamivel alkalmasabbak az enyhe kognitív zavar detektálására, mint az i-vektorok hasonló méretű adatokon és hasonló (akusztikai) körülmények között. A három tesztelt rejtett réteg közül egyértelműen a mélyebben fekvő szegmensszintű réteg (*Szegmens #6*) vezetett a legjobb eredményekhez mindhárom DNN modell esetében. Az augmentációval tanított modell a legtöbb esetben eredményesebb volt, mint az augmentációs lépés kihagyásával tanított; mindkettő alulmaradt ugyanakkor Snyder és munkatársai előtanított modelljével szemben, melyben valószínűleg a lényegesen nagyobb tanítóadat játszhatott szerepet. Bár kíváncsiak lettünk volna, hogy más keretszintű jellemzők használata hogyan alakítja az eredményeket, a Kaldi beépített x-vektor eszköze meglepően sok korláttal bír: sem a Δ / $\Delta\Delta$ értékek, sem például frekvenciasávok energiaösszegeinek mint jellemzőknek a használata nem könnyen megoldható. Ugyanígy kíváncsiak lettünk volna rá, hogy a tanítófelvételek „zajosítása" számszerűen mennyit segíthet az i-vektorok teljesítményén, azonban ez az augmentációs lépés is az x-vektor DNN modell tanításához van kötve.

Az osztályozáskor kapott eredményeink nem voltak különösebben átütőek, aminek több oka is lehet. Egyrészt az EKZ-s és kontroll alanyainktól származó felvételek sajnos elég zajosak, melyen utólag már nehéz segíteni (ugyanakkor így talán jobban tükrözik egy valós környezetben lefolytatott EKZ-szűrővizsgálat akusztikai körülményeit). Másrészt érdemes szem előtt tartani, hogy az enyhe kognitív zavart elsősorban a memória és bizonyos nyelvi készségek romlása jellemzi, melyeket sokkal nehezebb kimutatni a beszédből, mint például a Parkinson-kór tüneteit. Mégis, Jeancolas és munkatársai az x-vektorok használatával is „csupán" 70% körüli osztályozási pontosságokat kaptak Parkinson-kór felismerésére (Jeancolas és mtsai, 2020). Természetesen az x-vektor beágyazások lehetséges felhasználása lehet még, hogy kombináljuk azokat (vagy a használatukkal kapott predikciókat) más jellegű jellemzőkkel (például temporális paraméterekkel, lásd Gosztolya és mtsai, 2018), melyet tervezünk kipróbálni a közeljövőben.

Köszönetnyilvánítás

A kutatást részben a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal (projektkód: FK-124413), részben az Innovációs és Technológiai Minisztérium (projektkód: TUDFO/47138-1/2019-ITM) támogatta. Gosztolya Gábor kutatásait az MTA Bolyai János ösztöndíja és az Új Nemzeti Kiválóság Program Bolyai+ pályázata (azonosító: ÚNKP-20-5-SZTE-649) támogatta. A publikációban szereplő kutatást (amelyet a Szegedi Tudományegyetem valósított meg) az Innovációs és Technológiai Minisztérium és a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal is támogatta a Mesterséges Intelligencia Nemzeti Laboratórium keretében.

Hivatkozások

Alzheimer's Association: 2020 Alzheimer's disease facts and figures. Alzheimer's & Dementia 16(3), 391–460 (2020)

Aude, A.: Audio quality measurement primer (1998)

Botelho, C., Teixeira, F., Rolland, T., Abad, A., Trancoso, I.: Pathological speech detection using x-vector embeddings (2020)

Egas-López, J.V., Tóth, L., Hoffmann, I., Kálmán, J., Pákáski, M., Gosztolya, G.: Assessing Alzheimer's Disease from speech using the i-vector approach. In: SPECOM. pp. 289–298. Isztambul, Törökország (2019)

García, N., Orozco-Arroyave, J.R., D'Haro, L.F., Dehak, N., Nöth, E.: Evaluation of the neurological state of people with Parkinson's Disease using i-vectors. In: Interspeech. pp. 299–303. Stockholm, Svédország (2017)

García, N., Vásquez-Correa, J., Orozco-Arroyave, J.R., Nöth, E.: Multimodal i-vectors to detect and evaluate Parkinson's Disease. pp. 2349–2353. Hyderabad, India (2018)

Gosztolya, G., Hoffmann, I., Tóth, L., Vincze, V., Pákáski, M., Kálmán, J.: Az enyhe kognitív zavar és korai alzheimer-kór automatikus azonosítása spontán beszédből akusztikus jellemzők segítségével. In: MSZNY. pp. 219–230. Szeged (2018)

de Ipiña, K.L., de Lizarduy, U.M., Calvo, P.M., Beitia, B., García-Melero, J., Fernández, E., Ecay-Torres, M., Faundez-Zanuy, M., Sanz, P.: On the analysis of speech and disfluencies for automatic detection of mild cognitive impairment. Neural Computing and Applications 9, 437 (2018)

Jeancolas, L., Petrovska-Delacrétaz, D., Mangone, G., Benkelfat, B., Corvol, J., Vidailhet, M., Lehéricy, S., Benali, H.: X-vectors: New quantitative biomarkers for early Parkinson's Disease detection from speech. arXiv preprint arXiv:2007.03599 (2020)

König, A., Satt, A., Sorin, A., Hoory, R., Derreumaux, A., David, R., Robert, P.H.: Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. Current Alzheimer Research 15(2), 120–129 (2018)

Meilán, J.J.G., Martínez-Sánchez, F., Martínez-Nicolás, I., Llorente, T.E., Carro, J.: Changes in the rhythm of speech difference between people with nondegenerative mild cognitive impairment and with preclinical dementia. Behavioural Neurology 2020, 4683573 (2020)

Neuberger, T., Gyarmathy, D., Gráczi, T.E., Horváth, V., Gósy, M., Beke, A.: Development of a large spontaneous speech database of agglutinative Hungarian language. In: Proceedings of TSD. pp. 424–431. Brno, Czech Republic (Sep 2014)

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830 (2011)

Petersen, R.C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., Fratiglioni, L.: Mild cognitive impairment: a concept in evolution. Journal of Internal Medicine 275(3), 214–228 (2014)

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlíček, P., Qian, Y., Schwarz, P., Silovský, J., Stemmer, G., Veselý, K.: The Kaldi speech recognition toolkit. In: Proceedings of ASRU (2011)

Prince, M., Wimo, A., Guerchet, M., Ali, G.C., Wu, Y.T., Prina, M.: World Alzheimer Report 2015. The Global Impact of Dementia. Alzheimer's Disease International, London, UK (2015)

Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., Williamson, R.: Estimating the support of a high-dimensional distribution. Neural Computation 13(7), 1443–1471 (2001)

Sluis, R.A., Angus, D., Wiles, J., Back, A., Gibson, T.A., Liddle, J., Worthy, P., Copland, D., Angwin, A.J.: An automated approach to examining pausing in the speech of people with dementia. American Journal of Alzheimer's Disease & Other Dementias 35, 1533317520939773 (2020)

Snyder, D., Garcia-Romero, D., Povey, D., Khudanpur, S.: Deep Neural Network embeddings for text-independent speaker verification. In: Interspeech. pp. 999–1003. Stockholm, Svédország (2017)

Snyder, D., Garcia-Romero, D., Sell, G., Povey, D., Khudanpur, S.: X-vectors: Robust DNN embeddings for speaker verification. In: ICASSP. pp. 5329–5333. Calgary, Alberta, Kanada (2018)

Szatlóczki, G., Hoffmann, I., Vincze, V., Kálmán, J., Pákáski, M.: Speaking in Alzheimer's Disease, is that an early sign? Importance of changes in language abilities in Alzheimer's Disease. Frontiers in Aging Neuroscience 7, 104943 (2015)

Themistocleous, C., Eckerström, M., Kokkinakis, D.: Identification of Mild Cognitive Impairment from speech in Swedish using Deep Sequential Neural Networks. Frontiers in Neurology 9, 975 (2018)

Themistocleous, C., Eckerström, M., Kokkinakis, D.: Voice quality and speech fluency distinguish individuals with Mild Cognitive Impairment from Healthy Controls. PloS one 15(7), e0236009 (2020)

Vincze, V., Szatlóczki, G., Tóth, L., Gosztolya, G., Pákáski, M., Hoffmann, I., Kálmán, J.: Telltale silence: temporal speech parameters discriminate between prodromal dementia and mild Alzheimer's disease. Clinical Linguistics & Phonetics közlésre elfogadva (2020)

Weiner, J., Schultz, T.: Selecting features for automatic screening for dementia based on speech. In: SPECOM. pp. 747–756. Lipcse, Németország (2018)