

## Szkizofrénia azonosítása spontán beszéd temporális paraméterein alapján – egy pilot kutatás eredményei

Bagi Anita<sup>1,5</sup>, Gosztolya Gábor<sup>2</sup>, Szalóki Szilvia<sup>3,5</sup>,  
Szendi István<sup>3,5</sup> és Hoffmann Ildikó<sup>1,4,5</sup>

<sup>1</sup> SZTE BTK Magyar Nyelvészeti Tanszék, 6722 Szeged, Egyetem u. 2.  
bagianita88@gmail.com

<sup>2</sup> MTA-SZTE Mesterséges Intelligencia Kutatócsoport, 6701 Szeged, Pf. 652.  
ggabor@inf.u-szeged.hu

<sup>3</sup> SZTE ÁOK Pszichiátriai Klinika, 6725 Szeged, Kálvária sgt. 57.  
szilvi.szaloki@gmail.com, szendi.istvan@med.u-szeged.hu

<sup>4</sup> MTA Nyelvtudományi Intézet, 1394 Budapest Pf. 360.  
i.hoffmann@hung.u-szeged.hu

<sup>5</sup> Mentális Betegségek Megelőzése Interdiszciplináris Kutatócsoport

**Kivonat:** A szkizofrénia olyan neurodegeneratív spektrum zavar, melyet különböző alulműködések együttese alkot. A szkizofréniát, számos tünete mellett, jellemzi például a csökkent információfeldolgozási sebesség és a csökkent verbális fluencia teljesítmény is. Jelen tanulmányunkban a beszédtempó folyamatoságát vizsgáljuk szkizofréniával élők és illesztett egészséges kontrollszemélyek irányított spontán beszéd-felvételeiben. Célunk, hogy rámutassunk a különböző beszédbeli temporális paraméterek (úm. artikulációs tempó, beszédtempó és különböző szünettartási mutatók) segítségével arra, hogy a két csoport között specifikus eltéréseket tudunk meghatározni egy korábban korai demencia felismerésre (enyhe kognitív zavarra és Alzheimer-kórra) kifejlesztett és tesztelt eljárás használatával. Munkánk során ezen temporális mutatók alkalmazhatóságát teszteltük gépi tanulással új betegpopuláción. Eredményeink azt mutatják, hogy a két csoport beszélői 70–80 % közti osztályozási pontosságértékekkel meghatározhatók és az F-értékek 81% és 87% közé esnek. Részletes vizsgálatunk feltárta, hogy a két csoport meghatározására a szünettartási temporális paraméterek közül a leghatékonyabbak azok az elemzési utak, melyek estében mind a néma, mind pedig a kitöltött szünetekkel számolunk.

**Kulcsszavak:** spontánbeszéd, temporális paraméterek, szkizofrénia, kitöltött szünetek

### 1 Bevezetés

Bár számos, több szempontból közelítő, széles körű vizsgálat ismert a szkizofrénia hátterének feltérképezéséhez, ezidáig nem tudtak meghatározni egyetlen specifikus genetikai, neurobiológiai vagy környezeti tényezőt sem, mely a betegség kiala-

kulásának hátterében állhat. Crow elmélete szerint [1] a szkizofrénia (fenomenológiai szempontból) olyan univerzális betegségnek tekinthető, mely a Föld valamennyi populációjában megtalálható. Elméletében feltételezi, hogy a szkizofrénia evolúciós szintű fennmaradásának hátterében a lateralizációt eredményező genetikai változások és a kialakuló pszichológiai struktúrák állhatnak. A szkizofrénia diagnózisának felállításához a következő tüneti kritériumok teljesülése szükséges: (1) téveszmék, (2) hallucinációk, (3) inkoherens beszéd, (4) szembeszökően szétesett vagy katonán viselkedés, (5) negatív tünetek, azaz hangulati üresség, alogia vagy akaratnélküliség. A tüneti kritériumok mellett fontos az időtartam aspektusa is, mely szerint legalább 6 hónapig, de az 5 fő tünet egyikének legalább egy hónapig fenn kell állnia ahhoz, hogy a diagnózis felállítható legyen [2].

A szkizofréniát számos kognitív deficit jellemezheti, ezen deficitek közé tartozik a csökkent információ-feldolgozási sebesség és a munkamemória károsodása [3]. Emlékezeti funkciók alulműködését találták szkizofréniával élőknel neuropszichológiai tesztek eredményeiben is, melyek érintették a munkamemóriát, a verbális fluencia teljesítményt és az epizodikus emlékezetet is [4,5,6]. Más kutatások specifikus károsodást mutattak ki szkizofréniában a munkamemória és a tartós figyelmi funkciók tekintetében is [7,8].

A szkizofréniával élők számos, különböző nyelvi szinteket érintő deficcittel rendelkezhetnek [9]. Pawełczyk és mtsai [10] azt találták, hogy a szkizofréniával élők egészséges kontrollszemélyek eredményeihez képest szignifikánsan alacsonyabb pontszámot értek el az olyan szubtesztek esetében, mint az implicit információ-feldolgozás, a humorfeldolgozás, a metaforák felfejtése, a nem odaillő vagy helytelen észrevételek és megjegyzések felismerése, az érzelmek megkülönböztetésére irányuló feladatokban, melyek a nyelvben használt intonációk felismerésével operáltak; emellett a különböző diskurzusok feldolgozása és értelmezése esetében is jelentős különbségeket találtak. Eltéréseket találtak továbbá a prozódia területén is, míg más kutatások szerint a szkizofrénia negatív tünetei megjelenhetnek a hanghordozás és a hangsúlyozás hiányában is [11,12]. A beszédprodukción felül a spontán beszédet vizsgáló kutatások a kommunikált gondolat összetettségét elemezték, és azt találták, hogy szkizofréniával élőknel ezek a megnyilatkozások kevésbé összetettek, mint az egészséges kontrollszemélyek beszédproduktumai. Ugyanakkor arra is felhívták a figyelmet, hogy azok a páciensek, akik jobb teljesítményt nyújtottak, nagyobb arányban voltak érintettek a depresszió és a szorongás különböző tüneteivel [13].

Számos fentebb említett tünetet számítógépes eszközökkel is elemeztek. Rosenstein és mtsai [14] a verbális munkamemóriát vizsgálták a verbális emlékezeti folyamatok mérésére koncentrálna számítógépes nyelvészeti megközelítésekkel és eszközökkel. Corcoran és mtsai [15] azt találták, hogy az automatizált szemantikai és szintaktikai elemzés jól használható kiindulási alapja lehetne egy diagnosztikai eszköznek. További prozódiai eltéréseket és lehetséges karakterisztikákat [16,17], illetve a beszéd folytonosságát, a megakadásjelenségek és szünetek minőségét és arányát is vizsgálták már hasonló eszköztárakkal [18]. Más kutatások azt találták, hogy a formális gondolkodási zavarral (mely szembetűnő tünete lehet a szkizofréniának) rendelkező páciensek markánsan kevesebb kitöltött szünetet produkáltak, mint az egészséges kontrollszemélyek [19].

Jelen tanulmányunkban a spontán beszédben észlelhető emlékezeti folyamatok deficitére koncentrálnak. Munkánk során irányított spontán beszédet vizsgálunk, mely egyben egy emlékezeti feladat is. A feladat pontos instrukciója a következő: „*Kérem, mesélje el a tegnapi napját!*”. Feltételezzük, hogy a spontán beszéd temporális mutatói különbözni fognak az egészséges és a szkizofréniával élő beszélők felvételeiben. A leginkább eltérő különbségeket a hezitációk számában és típusában várjuk. Elemzésünket automatizált elemzési módszerrel végezzük: automatizált beszédfelismerő-szoftver segítségével (ASR) kinyerjük a temporális mutatókat a felvételekből, majd mérlegre tesszük e kinyert paraméterek felhasználhatóságát statisztikai gépi tanulás alkalmazásával a két beszélői csoport elkülönítésére.

## 2 A beszéd temporális paraméterei

A spontán beszéd vizsgálatához a szkizofréniával élők és az egészséges kontrollszemélyek válaszaiból specifikus temporális paramétereket számítottunk ki. Kutatásunkat korábbi munkáinkra építettük [20, 21, 22], melyekben olyan, a hezitációt központba helyező temporális paramétereket mutattunk be, melyek az enyhe kognitív zavar (EKZ) korai detektálására használhatók. Az EKZ-t gyakran tekintik az Alzheimer-kór prodromális állapotának, mely egyben egy olyan mentális zavar is, amit igen nehéz diagnosztizálni. Az EKZ (spontán) beszédre gyakorolt hatása ismert [23]; e hatások közül jelen tanulmányunkban a verbális fluenciára koncentrálnak, mely szintén érintett lehet szkizofréniával diagnosztizált személyek esetében is [4, 5, 6]. Az EKZ-ban szenvedő betegek verbális fluenciájában gyakran mérhető rosszabbodás, mely megkülönböztető akusztikus változásokat eredményez; a két legfontosabb változást említve ezek tetten érhetők a több, ill. hosszabb hezitációkban és az alacsonyabb beszédtempóban is [24, 25]. Ezen eredmények felhasználására kifejlesztettünk egy olyan temporális paramétereket tartalmazó eszköztárat, mely elsősorban az alanyok beszédében mérhető hezitációk mennyiségére fókuszál.

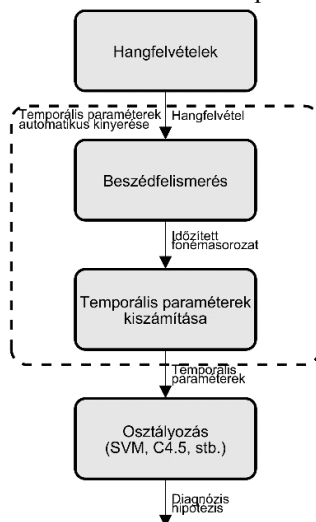
A temporális paramétereket tartalmazó jellemzőkészlet az 1. táblázatban látható. Meg kell jegyeznünk, hogy a paraméterek (4) és (8) között mind az alany spontán beszédében mért hezitáció mértékét írják le, különböző szempontokból fókuszálva a szünetek számára vagy hosszára. Ezen a ponton szükséges definiálnunk a hezitációra vonatkozó meghatározásainkat. A szünet legegyszerűbb formája a *néma* szünet, mely egyenlő a beszéd hiányával. Ugyanakkor a hezitáció megjelenhet *kitöltött* szünetként is, melynek vokalizációi lehetnek például az 'ööö', az 'hmm' vagy az 'ühh'. Mindkét szünettípus hezitációt jelez a spontán beszédprodukciónban. Ahhoz, hogy mindkét szünettípust elemezni tudjunk, a (4)–(8)-as temporális jellemzőket kiszámítottuk csak néma szünetekkel számolva; csak kitöltött szünetekkel számolva és végül minden szünettartással számolva függetlenül a szünet típusától. Ezen elemzési módszer összesen 18 temporális paraméterhez vezetett.

- (1) Artikulációs tempó (hezitációk nélkül számított másodpercenkénti beszédhang-szám)
- (2) Beszédtempó (másodpercenkénti beszédhang-szám osztva a megnyilatkozás teljes hosszával)
- (3) Megnyilatkozás teljes hossza (ezredmásodpercben mérve)
- (4) Szünetek száma (a szünetek előfordulásának száma)
- (5) Szünetek hossza (a szünetek összhossza)
- (6) Szünetek hosszának aránya (szünetek összhossza osztva a megnyilatkozás hosszával)
- (7) Szünetgyakoriság (szünetek előfordulásának száma osztva a megnyilatkozás hosszával)
- (8) Átlagos szünethossz (szünetek összhossza osztva a szünetek számával)

1. Táblázat: A nyolc vizsgált temporális jellemző, Hoffmann és mtsai [21] és Tóth és mtsai [22] nyomán

## 2.1 A beszéd felismerés-alapú temporális paraméterek kiszámítása

A fentebb bemutatott akusztikus-temporális paraméterek manuális feldolgozása meglehetősen hosszadalmas, drága és munkaigényes. Bár korai munkáink során még ezt a kinyerési utat alkalmaztuk [25], jelen tanulmányunkban már automatikusan nyertük ki azokat. Ezt az automatizált utat választva kézenfekvő megoldásnak tűnhet a jelfeldolgozásra támaszkodni [26]. Azonban ez a jelfeldolgozási technikákat alkalmazó megoldás, bár viszonylag egyszerűen és nagy hatékonysággal képes megkülönböztetni a *csendet* az emberi beszéd más hangzó részeitől; más, itt megkövetelt különbségtételeket nem képes megtenni. Így például, pusztán jelfeldolgozási eszközökre támaszkodva képtelenek lennének megkülönböztetni a kitöltött szüneteket a normál beszéd-től, illetve nem tudnánk kiszámítani az artikulációs tempót és a beszédtempót sem.



1. ábra: Automatizált folyamat a temporális paraméterek kiszámítására és elemzésére Tóth és mtsai [20] nyomán

A fenti szempontokat figyelembe véve, az automatikus beszédfelismerési technikák (Automatic Speech Recognition, ASR) mellett döntöttünk, melynek folyamatábrája az 1. ábrán látható. Sajnos egy készen kapott ASR eszköz várhatóan alkalmatlannak bizonyulna erre a feladatra – annál is inkább, mert a szabványos beszédfelismerőket a szószintű átírási hibák minimalizálására tanítják, miközben mi most éppen olyan nem-verbális akusztikus jellemzőket igyekszünk kinyerni, mint a beszédtempó vagy épp a néma és kitöltött szünetek időtartama. Szerencsére azonban az 1. táblázatban bemutatott beszédparaméterek nem követelik meg a hangok azonosítását, csak azok megszámlálását. Továbbá, míg a kitöltött szünetek nem jelennek meg expliciten egy standard beszédfelismerő rendszer kimenetében, a mi jellemzőkészletünk ezek detektálását kimondottan megköveteli. Mindezen megfontolások okán egy standard beszédfelismerő rendszert úgy módosítani, hogy az képes legyen kezelni az ilyen típusú „hibákat”, ha nem is lehetetlen, de mindenképpen nagyon munkaiigényes lenne.

Mindezen okokból egy olyan beszédfelismerő használata mellett döntöttünk, amely a bemenetként megadott hangfelvételhez kimenetként nem annak szószintű, hanem fonémaszintű átíratát adja meg. (A kitöltött szüneteket, az általánosság megkövetése nélkül, kezelhetjük egy speciális „fonémaként”). Természetesen a szószint teljes elhagyása (a szószintű nyelvi modellel és a teljes kiejtési szótárral együtt) várhatóan növelni fogja a fonémaszintű hibák számát is. Azonban, amint arra fentebb rámutattunk, nem minden típusú hangfelismerési hiba „rontja le” a temporális paraméterek kinyerését; jelen esetben csak a fonémák száma és a két szünet típusa (úm. néma és kitöltött) fontos.

### 3 Adatbázis

Tanulmányunkhoz folyó kutatásunk jelenleg is bővülő adatbázisából véletlenszerűen kiválasztottunk tíz szkizofréniával élő személyt, majd hozzájuk korban és nemben illesztettünk nyolc egészséges kontrollszemélyt. A két csoport nemenkénti megoszlása 50-50% volt, tehát a szkizofréniával élők csoportjában (SZ) 5 férfi és 5 nő, míg az egészséges kontrollcsoportban (K) 4 férfi és 4 nő volt. Jelenlegi kontrollcsoportunk száma meglehetősen alacsony, de folyamatosan dolgozunk az adatbázis bővítésén.

A résztvevőktől származó megnyilatkozásokat 2016 februárja és 2017 márciusa között rögzítettük a Szegedi Tudományegyetem Általános Orvostudományi Karának Pszichiátriai Klinikáján. A kutatást jóváhagyta a Szegedi Tudományegyetem Etikai Bizottsága; a kutatás teljes folyamatát a Helsinki Nyilatkozat szellemében végeztük. A kutatásban résztvevő minden beszélő magyar anyanyelvű volt. Az elemzés során a beszéd temporális jellemzőit mértük. A résztvevőktől irányított spontán beszédet rögzítettünk: arra kértük őket, hogy meséljenek a tegnapi napjukról. A pontos instrukció elhangzása után („Kérem, mesélje el a tegnapi napját!”) a résztvevőknek hozzávetőleg öt perc állt rendelkezésükre, hogy teljesítsék a feladatot – természetesen, ha egy-két perccel rövidebb vagy hosszabb időt vett igénybe a feladat megoldása, akkor sem szakította félbe őket a vizsgálatvezető. A hangfelvételek elkészítéséhez Roland R-05 típusú diktafont használtunk.

A csoportonkénti kormegoszlás az SZ-csoport esetében 39,9 év volt, míg a K-csoport esetében 40,2. Az iskolázottságot években számolva ( $t=-1,82$ ,  $df=18$ ,  $p=0,09$ )

és az életkor ( $t=0,06$ ,  $df=18$ ,  $p=0,96$ ) tekintetében nem volt szignifikáns különbség a két csoport értékei között. A hangfelvételek mellett minden résztvevővel elvégeztük a Módosított Mini-Mentál Tesztet is (MMSE [27]), melynek eredményeiben a két csoport szignifikáns eltérést mutatott ( $t=2,55$ ,  $df=10,55$ ,  $p=0,028$ ). A szkizofréniával élő személyek legtöbbször a *Felidőző emlékezés* altesztben veszítettek pontot: ez azonban nem feltétlenül jelez emlékezeti deficitet – a szórt figyelem eredménye is lehet az altesztben nyújtott csökkent teljesítmény.

## 4 Kísérleti elrendezés

### 4.1 A temporális paraméterek kinyerése

A beszédfelismerő rendszer akusztikus modelljének tanítására a Magyar Beszéltnyelvi Adatbázist (BEA, [28]) használtuk. A BEA adatbázisa tartalmaz spontán beszédet, így jelen kutatásunk szempontjából az egyik leghasznosabbnak tűnt, annál is inkább, mert kitöltött szünetekkel csak spontán beszédben találkozunk. A tanuláshoz közel 7 órányi spontán beszédet használtunk fel. Előzetesen megbizonyosodtunk arról, hogy az átiratokban fonémaszinten következetes módon volt jelölve a kitöltött szünet, a be- és kilégzés, a nevetés, a köhögés és a zihálás.

A beszédfelismerő rendszert arra tanítottuk, hogy felismerje a megnyilatkozásokban lévő beszédhangokat – a fonémakészlet természetesen tartalmazta ezeket a speciális nonverbális címkéket (kitöltött szünet, be- és kilégzés, nevetés stb.) is. Az akusztikai modellezéshez egy standard mély neurális hálót (Deep Neural Network, DNN) alkalmaztunk előre-csatolt (feed-forward) topológiával, melynek három rejtett rétege egyenként ezer ún. ReLU aktivációs függvényt használó neuront tartalmazott. Munkánk során saját implementációinkat használtuk, mellyel korábban kutatócsoportunk érte el a legalacsonyabb publikált szószintű hibaaarányt a TIMIT adatbázison [29]. Az alkalmazott nyelvi modell egy egyszerű fonéma bigram volt, mely (még egyszer kihangsúlyozva) tartalmazta a fentebb felsorolt nonverbális hangjelenségeket is. A beszédfelismerő rendszer kimenete egy időzített fonetikus átirat volt; ezekből az átiratokból (melyek a kitöltött szünetet mint speciális fonémát is tartalmazták) az 1. táblázatban felsorolt temporális paraméterek már könnyen kinyerhetők és kiszámíthatók.

### 4.2 Kiértékelési mutatók

A közelmúlt számos orvosi biológiai tanulmánya, mely ASR-alkalmazásokat használt, egyszerű osztályozási pontosságra támaszkodott (vö. [26,30]). Esetünkben azonban a vizsgált csoportok mindkét típusának gyakorisága meglehetősen kiegyensúlyozatlan: a szkizofrénia a populáció 1-1,5%-át érinti. Az ilyen kiegyensúlyozatlan osztályeloszlás miatt a pontosság egyáltalán nem működne megbízható mutatóként. Emiatt jelen kutatásban standard információ-visszakeresési kiértékelési metrikákat használtunk: pontosságot (precision), fedést (recall), és e kettő harmonikus középértékét, az F-

értéket (vagy az F1-értéket; F-measure; F1 score). Ezen felül kiszámítottuk a ROC-görbe alatti terület nagyságát (azaz az AUC mutatót) is az SZ osztályra.

### 4.3 Osztályozási folyamat

Osztályozási folyamatunk alapvetően az orvosbiológiai szokásokat követi, és hasonlít azokhoz a korábbi tanulmányainkhoz, amelyek az EKZ kimutatására koncentráltak (vö. [20,22]). A fentebb bemutatott temporális paraméterekre mint jellemzőkre szupport-vektor gépet (Support Vector Machine, SVM, [31]) tanítottunk, a LibSVM [32] implementációt használva. A nu-SVM metódust használtuk lineáris kernelfüggvény-nyel; a  $C$  értékét a  $10^{\{-5, \dots, 1\}}$  tartományban teszteltük.

Gépi tanulási szempontból rendkívül kicsi adathalmazon dolgoztunk, hiszen a kontrollcsoportba tartozó résztvevők száma korlátozott volt. Ebből adódóan nem láttuk értelmét külön tanító és tesztalmazok definiálásának, hanem beszélők szerinti keresztvalidációt (cross-validation, CV) alkalmaztunk: az osztályozó modellünket mindig 17 fős korpusz adatain tanítottuk, és mindig a fennmaradó egyre értékeltük ki azokat. Az SVM  $C$  meta-paraméterét beágyazott keresztvalidációban határoztuk meg [33]: a 17 beszélő esetében végzett tanításnál a tényleges CV lépésben újabb (beszélő szerinti) keresztvalidációt végeztünk. Azt a  $C$  értéket választottuk, amely a legmagasabb AUC pontszámot eredményezte ebben a saját „belső” CV tesztben. Ezt követően az SVM modellt ennek a 17 beszélőnek az adataira tanítottuk, és ezt a modellt értékeltük ki a 18. beszélő adatain. Ez az eljárás garantálja, hogy semmilyen szinten ne használjuk az aktuális tesztadatot az aktuális modell tanítására – ez ugyanis pontszámainkban torzulást eredményezett volna például standard keresztvalidáció használata esetén.

### 4.4 Az adatok előzetes feldolgozása

Kísérleteinkben egy-egy hangfelvételt használhatunk 18 beszélőtől. Adathalmazunk méretének növelése érdekében úgy döntöttünk, hogy kísérleteinkben rövidebb megnyilatkozás-egységeket használunk. Hipotézisünk az volt, hogy temporális beszédparamétereink akkor is értelmezhetőek maradnak, ha viszonylag rövid megnyilatkozásokból számoljuk őket. Ezt szem előtt tartva, a megnyilatkozásokat 30 másodperces szegmensekre osztottuk fel 10 másodperces átfedéseket hagyva (függetlenül a tényleges fonetikai határoktól), és a továbbiakban ezeket a szegmenseket önállóan kezeltük. Ezen lépések után 96 viszonylag rövid, de egyenlő méretű szegmensből álló adathalmazt kaptunk, amely jelentősen növelte SVM tanulókészletünk méretét. Természetesen az osztályozást ezek után is a már bemutatott beszélők szerinti beágyazott keresztvalidációs sémával végeztük; azaz egy-egy fold mindig egy beszélő összes szegmenséből állt.

Bár az eddig használt osztályozási metrikák logikus választásnak tűnnek a 30 másodperces szegmensek esetén is, a pontszámok jobban értelmezhetővé válnak, ha lefordítjuk őket a résztvevőkre. Egyszerű megoldás lehet erre az egyes beszélők kategóriájának (SZ vagy K) meghatározása az egyes szegmensekre adott hipotéziseinkből egyszerű többségi szavazással. Ezt azonban meglehetősen nehéz lenne értelmezni.

Ezért úgy döntöttünk, hogy előjelzéseinket egy másik megközelítéssel vonjuk össze a beszélő-szintű értékek meghatározásakor: egy beszélőre normalizált tévesztési mátrixot számítottunk ki az egyes beszédsgzemensek új súlyozásával:  $1/k$ , ahol  $k$  az adott beszélő szegmenseinek száma. Például egy egészséges beszélő 10 beszédsgzemennel (melyek közül 7 lett helyesen azonosítva) 0,7 valódi negatív és 0,3 hamis pozitív esetnek számít. A beszélők szerinti beágyazott keresztvalidálás befejezése után az osztályozási pontosság valamint az információ-visszakeresési metrikák könnyen kiszámíthatóak a beszélők szerint normalizált tévesztési mátrixból. Sajnos az AUC értékeket ebben a megközelítésben nem tudtuk meghatározni, mivel ahhoz az egyes példákra adott poszteriorbecslések is szükségesek lennének, míg most csak egy (normalizált) tévesztési mátrixsal rendelkezünk.

## 5 Eredmények

A 2. táblázat tartalmazza a kiszámított metrikáinkat a **szegmensek szintjén**. Ha mind a 18 temporális beszédparamétert be vesszük a jellemzőkészletbe, a 70,8%-os osztályozási pontosság viszonylag jó teljesítményt mutat. Az  $F_1$  81,3%-os értéke véleményünk szerint meglehetősen magasnak tűnik, különösen, ha figyelembe vesszük a tanítópéldák alacsony számát. A pontossági és fedési mutatókat vizsgálva láthatjuk, hogy a teljesítmény meglehetősen kiegyensúlyozatlan: a szkizofréniával élő betegek által produkált szegmensek mindössze 74%-át találta meg az eljárás, ám ezt megközelítőleg 90%-os pontossággal tette. Ez a probléma a kimeneti poszteriorbecslések küszöbértékelésével kezelhető [34], ugyanakkor úgy véljük, hogy e probléma tárgyalása már szétfeszítené jelen tanulmányunk kereteit.

Jellemzőkészlet	Osztályozási pontosság (%)				AUC
	Pont.	Prec.	Fedés	$F_1$	
Teljes	70,8	89,7	74,4	81,3	0,514
Néma szünetek	76,0	94,1	77,1	84,8	0,599
Kitöltött szünetek	75,0	97,1	75,0	84,6	0,435
Minden hezitáció	79,2	92,6	80,8	86,3	0,726
Tempó + néma szünetek	80,2	97,1	79,5	87,4	0,641
Tempó + kitöltött szünetek	70,8	91,2	73,8	81,6	0,602
Tempó + minden hezitáció	78,1	91,2	80,5	85,5	0,694

2. táblázat. A szegmensszintű pontossági értékek a különböző jellemző-részhalmozatok használata esetén

A temporális paraméterek egy részhalmozát felhasználó elemzések eredményeit vizsgálva megfigyelhetjük, hogy az osztályozási pontszámok szinte minden esetben javultak. A néma vagy kitöltött szünetekkel kapcsolatos időbeli paraméterek összehasonlításával megállapíthatjuk, hogy a szkizofréniára azonosítására a kitöltött szünetek kevésbé hasznosak, mint a néma szünetek értékei: a 71-75%-os osztályozási pontossági értékek elmaradnak a 76-80%-os értékek mögött, melyek a néma szünetekre koncentrálnak – az F-érték és az AUC pontszám is magasabb az utóbbi két esetben. A



kapott értékek tendenciáit vizsgálva, véleményünk szerint, a vizsgált temporális paraméterek leghasznosabb részhalmazai azok voltak, amelyek a hezitálások alapján számított indikátorokból álltak – függetlenül attól, hogy ezek néma vagy kitöltött szünettel operáltak-e. Bár a néma szünethez tartozó paraméterek az artikulációs tempóval és a beszédtempóval kombinálva valamivel nagyobb pontossághoz és magasabb  $F_1$  értékhez vezettek, abban a két esetben, amikor mindkét szünettípust figyelembe vettük, konzisztensen magasabb pontosságértékeket kaptunk, valamint a legmagasabb AUC értékek is ekkor adódtak.

Az osztályozási eredmények értelmezésével a megnyilatkozások számának normalizálásával, az egyes szegmensek külön-külön való számbavétele helyett a mutató értékének enyhe csökkenését láthatjuk (3. táblázat).

Jellemzőkészlet	Osztályozási pontosság (%)			
	Pont.	Prec.	Fedés	$F_1$
Teljes	60,8	60,0	88,0	71,4
Néma szünetek	68,3	65,0	93,0	76,6
Kitöltött szünetek	65,7	62,1	98,3	76,1
Minden hezitáció	77,2	74,4	90,0	81,5
Tempó + néma szünetek	73,4	68,7	95,6	80,0
Tempó + kitöltött szünetek	61,0	59,9	89,7	71,9
Tempó + minden hezitáció	76,5	74,1	88,6	80,7

3. táblázat. A beszélőszintű pontossági értékek a különböző jellemző-részhalmazok használata esetén

Ami még érdekesebbnek tűnik, hogy a pontossági és visszakeresési eredmények tendenciáját nézve, az eredmények éppen ellenkező irányú tendenciát mutatnak, mint a szegmensek szintjét vizsgálva – itt már alacsonyabb pontosságot (precision), de viszonylag magas fedés értékeket láthatunk. Ez valószínűleg azért van, mert a szkizofréniával élők sokkal részletesebben írták le a tegnapi napjukat, mint az egészséges kontrollok; ebből következően az SZ csoportba tartozók felvételi szignifikánsan hosszabbak voltak, mint az egészséges kontrolloké. Ez azt eredményezte, hogy a megnyilatkozások száma is kiegyensúlyozatlanul alakult: számszerűsítve 68 (SZ) és 28 (K). A felhasznált temporális beszédparaméterek különböző alcsoportjainak vizsgálatát tekintve, minden valószínűség szerint a két beszélői csoportot a leghatékonyabban úgy azonosíthatnánk, ha figyelembe vennénk mindkét szünettípust. Ez következik abból is, hogy a 77,2% és 76,5%-os osztályozási pontossági pontszámok szignifikánsan magasabbak, mint csak a néma szünetekkel (68,3% és 73,4%), vagy csak a kitöltött szünetek használatával (65,7% és 61,0%) kapott értékek. Az így kapott  $F_1$ -értékek (81,5% és 80,7%) is messze a legmagasabbnak mértek (76,6-80,0% és 76,1%-71,9%, a néma és a kitöltött szüneteket külön-külön vizsgálva).

A különböző temporális paraméterek hasznosíthatóságát illetően tény, hogy a szkizofréniával élő résztvevők felvételi lényegesen hosszabbak voltak, mint az egészséges kontrollok hangfelvételei. A jelenség háttérben felvetődhet lehetséges magyarázatként az olyan pozitív tünetek jelenléte, mint a circumstantialitás, mely a kommunikálni kívánt tartalom túlzott részletességgel való kifejtését jelenti, de hasonlóan e pozitív tünethez, a gondolatrohanások és a szisztematikus önhivatkozások is vezet-

hetnek a hosszabb megnyilatkozásokhoz. A szkizofréniával élők néma szüneteinek magasabb száma további tünetek beszédre gyakorolt hatásával is magyarázható, melyek egyaránt érintik a végrehajtó és emlékezeti funkciókat is, s gyakran eredményeznek zavart gondolkodást, mely a beszédben válik tetten érhetővé. A szkizofréniával élőknek gyakran okoz problémát a gondolatok szervezése, rendszerezése, ami tükröződhet a spontán beszéd temporális paramétereiben is (például a néma vagy kitöltött szünetek számában).

Összegezve az eddigieket, vizsgálatunkban szignifikáns különbséget találtunk a két beszélői csoport (SZ és K) spontán beszédének temporális paramétereiben. A vizsgált temporális paraméterek közül az artikulációs arányra, a beszédtempóra és a hezitációkra koncentrálva, meglehetősen pontosan tudtunk különbséget tenni a két beszélői csoport között. A jövőben további résztvevőket kívánunk bevonni jelenleg is folyó kutatásunkba, hogy megerősíthessük és árnyalhassuk eddigi eredményeinket. Tervezzük továbbá a spontán beszéd fentebb bemutatott elemzését a teljes pszichózis spektrumon is, beteg-kontrollcsoportként együtt vizsgálva a szkizofréniát a bipoláris zavarral és a szkizoaffektív zavarral.

## 6 Összegzés

Jelen tanulmányunkban feltételeztük, hogy különbséget találunk az egészséges kontroll személyek és a szkizofréniával élők spontán beszédének temporális paramétereiben. Automatikus beszédelemzéssel és gépi tanulási technikákkal hatékonyan meg tudtuk különböztetni a két beszélői csoport tagjait. A hezitációs jelenségeket a legfontosabb megkülönböztető jegyeknek feltételeztük, mely feltételezésünket a vizsgálat eredményei igazoltak is: a 77%-os osztályozási pontszámok szignifikánsan magasabbak voltak, mintha csak a néma szüneteket (68-73%) vagy csak a kitöltött szüneteket vizsgáltuk volna (61-66%).

Munkánk pilotkutatás volt: arra kerestük a választ, hogy vajon az automatikus beszédelemzési folyamat használható lenne-e a szkizofréniával élők spontán beszédének temporális elemzésében. Törekedtünk továbbá arra is, hogy kutatásunk hozzájáruljon a neurodegeneratív rendellenességekről alkotott ismereteink bővítéséhez, s ezzel együtt pontosítsa a kapcsolódó szupraszegmentális jegyek leírását is. Természetesen az erősebb kijelentések megtételéhez szükség van a kutatásainkban résztvevők számának növelésére. Jelenleg is folyamatosan vonunk be résztvevőket a pszichózis-spektrum egyéb betegcsoportjaiból is.

## Köszönetnyilvánítás

A kutatást az EFOP-3.6.1-16-2016-00008 a.sz., EU társfinanszírozású projekt támogatta.

## Bibliográfia

1. Crow, T.J.: Is schizophrenia the price that Homo sapiens pays for language? *Schizophrenia Research* **28** (2–3) (1997) 127–141
2. American Psychiatric Association: Diagnostic and statistic manual of mental disorders (DSM-5). American Psychiatric Publishing (2013)
3. Kochunov, P., Coyle, T.R., Rowland, L.M., Jahanshad, N., Thompson, P.M., Kelly, S., Du, X., Sampath, H., Bruce, H., Chiappelli, J., Ryan, M., Fisseha, F., Savransky, A., Adhikari, B., Chen, S., Paciga, S.A., Whelan, C.D., Xie, Z., Hyde, C.L., Chen, X., Schubert, C.R., O'Donnell, P., Hong, E.: Association of White Matter With Core Cognitive Deficits in Patients With Schizophrenia. *JAMA Psychiatry* **74** (9) (2017) 958–966
4. Heinrichs, R.W., Zakzanis, K.K.: Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* **12** (3) (1998) 426–445
5. McCleery, A., Ventura, J., Kern, R.S., Subotnik, K.L., Gretchen-Doorly, D., Green, F.M., Helleman, G.S., Nuechterlein, K.H.: Cognitive functioning in first-episode schizophrenia: MATRICS consensus cognitive battery (MCCB) profile of impairment. *Schizophrenia Research* **157** (1–3) (2014) 33–39
6. Zhang, T., Li, H., Stone, W.S., Woodberry, K.A., Seidman, L.J., Tang, T., Guo, Q., Zhuo, K., Qian, Z., Cui, H., Zhu, Y., Jiang, L., Chow, A., Tang, Y., Li, C., Jiang, K., Yi, Z., Xiao, Z., Wang, J.: Neuropsychological impairment in prodromal, first-episode, and chronic psychosis: assessing RBANS performance. *PLoS One* **10** (5) (2015) 33–39
7. Chan, R., Chen, E., Cheung, E., Cheung, H.: Executive dysfunction in schizophrenia: relationships to clinical manifestation. *European Archives of Psychiatry and Clinical Neuroscience* **254** (4) (2004) 256–262
8. Huang, J., Tan, S.P., Walsh, S.C., Spriggs, K., Neumann, D.L., Shum, D.H., Chan, R.C.: Working memory dysfunctions predict social problem solving skills in schizophrenia. *Psychiatry Research* **220** (1–2) (2014) 96–101
9. Nagels, A., Kircher, T.: Symptoms and Neurobiological Models of Language in Schizophrenia. In: Hickok, G., Small, S. (eds.) *Neurobiology of Language*. Academic Press (2016) 887–897
10. Pawelczyk, A.M., Kotlicka-Antczak, M., Lojek, E., Ruszpel, A., Pawelczyk, T.: Schizophrenia patients have higher-order language and extralinguistic impairments. *Schizophrenia Research* **192** (2017) 274–280
11. Covington, M.A., Congzhou, H., Brown, C., Naci, L., McClain, J.T., Fjordbak, B.S., Semple, J., Brown, J.: Schizophrenia and the structure of language: The linguist's view. *Schizophrenia Research* **77** (1) (2005) 85–98
12. Rapcan, V., D'Arcy, S., Yeap, S., Afzal, N., Thakore, J.H., Reilly, R.B.: Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. *Medical Engineering & Physics* **32** (9) (2010) 1074–1079
13. Moe, A.M., Breitborde, N.J.K., Shakeel, M.K., Gallagher, C.J., Docherty, N.M.: Idea density in the life-stories of people with schizophrenia: Associations with narrative qualities and psychiatric symptoms. *Schizophrenia Research* **172** (1–3) (2015) 201–205
14. Rosenstein, M., Diaz-Asper, C., Foltz, P.W., Elvevag, B.: A computational language approach to modeling prose recall in schizophrenia. *Cortex* **55** (2014) 148–166
15. Corcoran, C.M., Carrillo, F., Fernández-Slezak, D., Bedi, G., Klim, C., Javitt, D.C., Bearden, C.E., Cecchi, G.A.: Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry* **17** (1) (2018) 67–75
16. Bedwell, J.S., Cohen, A.S., Trachik, B.J., Deptula, A.E., Mitchell, J.C.: Speech prosody abnormalities and specific dimensional schizotypy features: Are relationships limited to males. *The Journal of Nervous and Mental Disease* **202** (10) (2014) 745–751

17. Martínez-Sánchez, F., Muela-Martinez, J.A., Cortés-Soto, P., Meilán, J.J.G., Ferrándiz, J.A.V., Caparrós, A.E., Valverde, I.M.P.: Can the acoustic analysis of expressive prosody discriminate schizophrenia? *The Spanish Journal of Psychology* **18** (86) (2015) 1–9
18. Alpert, M., Kotsaftis, A., Pouget, E.R.: At Issue: Speech fluency and schizophrenic negative signs. *Schizophrenia Bulletin* **23** (2) (1997) 171–177
19. Matsumoto, K., Kircher, T.T.J., Paul R.A.: Stokes and Michael J. Brammer and Peter F. Liddle and Philip K. McGuire Frequency and neural correlates of pauses in patients with formal thought disorder. *Frontiers in Psychiatry* **4** (2013) 67–75
20. Tóth, L., Gosztolya, G., Vincze, V., Hoffmann, I., Szatlóczki, G., Biró, E., Zsura, F., Pákáski, M., Kálmán, J.: Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech using ASR. *Proceedings of Interspeech Dresden, Germany* (2015) 2694–2698
21. Hoffmann, I., Tóth, L., Gosztolya, G., Szatlóczki, G., Vincze, V., Kárpáti, E., Pákáski, M., Kálmán, J.: Beszédfelismerés alapú eljárás az enyhe kognitív zavar automatikus felismerésére spontán beszéd alapján. *Általános Nyelvészeti Tanulmányok* **29** (2017) 385–405
22. Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatlóczki, G., Bánréti, Z., Pákáski, M., Kálmán, J.: A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech. *Current Alzheimer Research* **15** (2) (2018) 130–138
23. Laske, Ch., Sohrabi, H.R., Frost, Sh.M, López-de-Ipina, K., Garrard, P., Buscema, M., Dauwels, J., Soekadar, S.R., Mueller, S., Linnemann, Ch., Bridenbaugh, S.A., Kanagasingam, Y., Martins, R.N., O'Bryant, S.E.: Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimer's & Dementia* **11** (2015) 561–578
24. Roark, B., Mitchell, M., Hosom, J.P., Hollingshead, K., Kaye, J.: Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech, and Language Processing* **19** (7) (2011) 2081–2090
25. Hoffmann, I., Németh, D., Dye, C.D. and Pákáski, M., Irinyi, T., Kálmán, J.: Temporal parameters of spontaneous speech in Alzheimer's disease. *International Journal of Speech-Language Pathology* **12** (1) (2010) 29–34
26. López-de-Ipina, K., Alonso, J.B., Solé-Casals, J., Barroso, N., Henriquez, P., Faundez-Zanuy, M., Travieso, C.M., Ecay-Torres, M., Martinez-Lage, P., Eguiraun, H.: On Automatic Diagnosis of Alzheimer's Disease Based on Spontaneous Speech Analysis and Emotional Temperature. *Cognitive Computation* **7** (1) (2015) 44–55
27. Folstein, M.F., Folstein, S.E., McHugh, P.R.: Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* **12** (3) (1975) 189–198
28. Gósy, M.: BEA A multifunctional Hungarian spoken language database. *The Phonetician* **105** (106) (2012) 50–61
29. Tóth, L.: Phone Recognition with Hierarchical Convolutional Deep Maxout Networks. *EURASIP Journal on Audio, Speech, and Music Processing* **25** (2015) 1–13
30. Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., Gorno-Tempini, M.L.: Machine learning approaches to diagnosis and laterality effects in semantic dementia discours. *Cortex* **55** (2014) 122–129
31. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the Support of a High-Dimensional Distribution. *Neural Computation* **13** (7) (2001) 1443–1471
32. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2** (3) (2011) 1–27
33. Cawley, G.C., Talbot, N.L.C.: On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *Journal of Machine Learning Research* **11** (2010) 2079–2107

XV. Magyar Számítógépes Nyelvészeti Konferencia      Szeged, 2019. január 24–25.

34. Waegeman, W., Dembczynski, K., Jachnik, A., Cheng, W., Hüllermeier, E.: On the Bayes-Optimality of F-Measure Maximizers. *Journal of Machine Learning Research* **1** (15) (2014) 3333–3388.