Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar

A DEPRESSZIÓS BESZÉD AKUSZTIKAI-FONETIKAI JELLEMZŐINEK VIZSGÁLATA

Ph.D. tézisfüzet
BME-VIK Informatikai Tudományok Doktori Iskola

Kiss Gábor
okl. mérnök-informatikus

Témavezető:
Vicsi Klára, DSc

Budapest, 2019

1. Bevezetés

Az emberi beszéd bonyolult folyamat, amely sokrétű információt hordoz, és a kommunikáció egyik legalapvetőbb eszköze. Az emberi beszédnek mint kommunikációs eszköznek elsődleges célja a gondolatok szimbolikus úton történő közvetítése a nyelvi tartalom segítségével. Ezenfelül további információkat is képes közvetíteni, mint például a beszélő érzelmi állapotát, attitűdjét, és képes utalni a beszélő fiziológiai állapotára is, amennyiben az állapot hatással bír a létrejött beszédproduktumra, vagyis hatással bír a beszédkeltés folyamatára [1][2][3][4][5][6].

A beszédkeltés alapvető fiziológiai szervei: a tüdő, a légcső, a gége, a garat, a száj- és az orrüreg. Működésüket az agy irányítja. Azok a fiziológiai elváltozások detektálhatók a beszédjel feldolgozás alapján, amelyek vagy elváltozást okoznak a beszédkeltésért felelős szervekben (megfázás, rákos daganat), vagy az agyi működésben okoznak elváltozást (Parkinson-kór vagy depresszió), esetleg egyszerre mindkettőben. A kutatásomban a depressziós állapot beszédre gyakorolt hatását és a megváltozott akusztikai-fonetikai jellemzők által lehetséges depressziós állapot detektálását vizsgáltam.

A depresszió korunk egyik legelterjedtebb, gyógyítható betegsége [7], ám diagnosztizálása szaktudást igényel, és így a kórkép felállítása a társadalom egy szűk rétegére hárul. A depresszió súlyossága nagyban befolyásolja az ebben szenvedő beteg életminőségét [8]. Depresszió hatására megváltoznak az emberi beszédproduktum egyes jellemzői, amelyek számszerűsíthetők és mérhetők. Emiatt lehetőség nyílik a depresszió beszédjel alapú gépi detektálásra, ami megkönnyítheti, illetve szélesebb körben lehetővé teheti a betegség diagnosztizálását. Ezen okok miatt fontos kutatási terület a depressziós állapot beszédjel alapú felismerése és súlyosságának becslése.

2. Áttekintés

A depresszió egy pszichiátriai betegség, amely a World Health Organization (WHO) 2012-es felmérése alapján a harmadik leggyakoribb betegség világszerte [9]. A klinikai pszichiátria megkülönbözteti a depresszió különböző fajtáit. A depressziós állapotot két fő kategória szerint sorolják be: unipoláris és bipoláris depresszió. Az unipoláris depresszió, azaz egyfázisú depresszió esetén csak a rossz hangulat irányába van eltérés. A bipoláris depresszió esetén mind a rossz, mind pedig a jó hangulat túlzott felfokozottsága felváltva bekövetkezik. A fő kategóriákon belül további

típusokat definiálnak, ezek közül az unipoláris depresszió esetén főbb típusok a major depresszió (klinikai depresszió), illetve a disztímia. A bipoláris depresszió esetén a főbb típusok az I. típusú bipoláris más néven mániás depresszió, II. típusú bipoláris más néven hipománia, illetve a ciklotímia [10]. A kutatásaim során az unipoláris depresszión belül a klinikai depresszió hatását vizsgáltam az emberi beszédre. Továbbiakban a klinikai depresszió helyett a depressziót használom az egyszerűség kedvéért.

Már 1921-ben megállapította Emil Kraeplin, a modern pszichológia egyik megalapozója, hogy a depressziós állapot hatására megváltozik az emberi beszéd [11]. A depresszió és a beszéd kapcsolata az 1980-as évekre már fontos kutatási területnek számít, és több akusztikai, illetve fonetikai paramétert kapcsolatba hoztak a depresszióval, mint például az átlagos alapfrekvencia érték, az alapfrekvencia tartománya, a beszédtempó [12]. Azonban a depresszió gépi detektálása új kutatási terület, amit az informatika fejlődése tett lehetővé. Cummins és társai 2015-ös folyóiratcikkükben egy átfogó tanulmányt közöltek a beszédjel alapú depresszió detektálásához kapcsolódó fontosabb kutatások legfrissebb eredményeiről angol nyelv esetében [13].

Alapvetően kétféle gépi detektálási módszert alkalmaznak a kutatók: a beszélő depressziós állapotának osztályozását, illetve a depresszió súlyosságának a becslését regressziós eljárással. Kétosztályos gépi tanuló eljárás tanítása esetén (nem depressziós/depressziós) elégséges lehet az orvosi diagnózis megléte a minták címkézéséhez, azonban többosztályos osztályozás esetén, ami képes a depressziós állapot különböző kategóriáit is elkülöníteni súlyosság alapján, illetve a regressziós eljárás esetén, elengedhetetlen valamilyen depressziós állapotot leíró skála.

A klinikai depresszió súlyosságának leírására két széles körben elterjedt skála létezik: a Hammilton skála (HAM-D) [14] és a Beck Depression Inventory (BDI) [15]. Az előbbi értékét a szakorvos határozza meg a diagnózis során, míg az utóbbi egy önkitöltő teszt.

A depressziós állapot az agy motorikus működését befolyásolja, emiatt változik a depressziós ember beszédproduktuma. A nemzetközi irodalom több beszédparamétert is említ, amely a depresszió hatására megváltozik. Azonban abban még nincs általános egyezség, hogy melyek az igazán jellemző paraméterek, és az adott beszédparamétereket hogyan érdemes mérni. Természetesen a mérési lehetőség nagyban függ a beszédadatbázis feldolgozottságától, ami lehet csupán egy egyszerű

beszéd/nem beszéd szerinti [16][17], de akár egy pontos beszédhangszintű szegmentálása is az adatbázisnak. Az utóbbit értelemszerűen lényegesen időigényesebb és költségesebb megvalósítani, emiatt a legtöbb eddigi kutatásban nem alkalmaztak beszédhangszintű szegmentálást, viszont ebből kifolyólag egyes beszédparamétereket csak pontatlanul, nagyobb szórással tudtak megmérni, illetve egyes beszédparaméterek meg sem mérhetők a beszéd szegmentálása nélkül. Talán pont ezen okokból kifolyólag egyes beszédparaméterek értékei esetén eltérő tendenciákat mértek a kutatók a depresszió hatására [13][18][19][20].

Egy másik alapvető probléma a különböző kutatásokban, hogy a női és a férfi beszédmintákat egyben kezelik. A jelenleg létező depressziós beszédadatbázisok viszonylag kisméretűek: 30-160 beszélőtől tartalmaznak beszédmintákat [13], így ha külön vizsgálják a nőket és a férfiakat, akkor a vizsgált halmaz mérete tovább csökken. Ugyanakkor a saját kutatásaimból kiderült, a beszélő neme nagyban befolyásolhatja a beszéd egyes akusztikai-fonetikai jellemző értékeinek a megváltozását a depressziós állapot hatására. Ez az új megfigyelésem egyezik a depresszióval foglalkozó pszichiáterek megfigyelésével, ugyanis a depresszió hatása különbözőképpen mutatkozik meg a férfiak és a nők esetében [21]. Így ha közös modellt használunk a nők és a férfiak esetében, az problémát okozhat a depressziós állapot felismerésében, ami mindenképpen megoldandó feladat.

3. Kutatási célkitűzések

Kutatásaimmal a depresszió beszédre gyakorolt hatásának mélyebb megértéshez, a depressziós állapot beszédalapú detektáláshoz és a depressziós állapot súlyosság szerinti becsléséhez kívánok hozzájárulni. Konkrét céljaim a kutatás során:

1) Megvizsgálni a depressziós állapot hatását az emberi beszédre. Olyan akusztikai-fonetikai paraméterek kiválasztása, amelyek értéke igazolhatóan megváltozik a depressziós állapot hatására.

2) Megvizsgálni gépi tanuló eljárás segítségével, hogy automatikusan, kizárólag a beszédjel feldolgozása alapján, milyen pontossággal lehet osztályozni a bemondókat depressziós, illetve egészséges állapotuk alapján.

3) Megvizsgálni gépi tanuló eljárás segítségével, hogy automatikusan, kizárólag beszédjel feldolgozás alapján, milyen pontossággal lehet becsülni az egyes bemondók depressziós állapotának súlyosságát.

4) Megvizsgálni, hogy a depressziós állapot hatásainak a beszédre gyakorolt realizációi mennyiben függetleníthetők a beszélő által használt nyelvtől. Milyen pontossággal lehetséges automatikusan detektálni, illetve becsülni a depressziós állapotot több nyelvre érvényes modell segítségével.

5) Olyan rendszer kidolgozása és implementálása, amely képes vizsgált személyek depressziós állapotának megfigyelésére beszédjel feldolgozás alapján.

4. Felhasznált adatbázisok és módszertan

Ebben a fejezetben bemutatom a kutatásaim során felhasznált beszédadatbázisokat, eszközöket, illetve az általam használt fontosabb módszereket.

4.1. Módszertan

Kutatásom során először statisztikai összehasonlító elemzést végeztem az egészséges és depressziós emberek beszédéből kinyert akusztikai-fonetikai jellemzők értékein. A statisztikai elemzés segít a depresszió beszédparaméterekre gyakorolt hatásának megértésében, másrészt a gépi tanuló eljárások esetében, mint egy gyors jellemző kiválasztási algoritmust lehet használni.

Gépi tanuló eljárások használatával osztályozási és regressziós feladatokat valósítottam meg annak érdekében, hogy megvizsgáljam a depressziós állapot detektálása, illetve a depresszió súlyossága mekkora pontossággal becsülhető automatikusan a beszédből kinyert paraméterek segítségével.

4.1.1. Kutatásaim során használt programok:

BME-TMIT-LSA nyelvfüggetlen kényszerített beszédhangszintű szegmentáló: beszédhangok és beszédhang határok automatikus jelölésére [22].

Praat: hangelemző és annotáló program, akusztikai-fonetikai alapjellemzők mérésére [23].

LibSVM: SVM/SVR alapú osztályozási és regressziós feladatok megvalósítására [24].

SPSS_20: statisztikai elemző szoftver, statisztikai vizsgálatok megvalósítására [25].

Rapidminer: adatelemző program, aminek segítségével osztályozási és regressziós feladatokat valósítottam meg [26].

4.1.2. Depresszió súlyosságának jelölése

A depresszió súlyosságának leírására a Beck Depression Inventory II (BDI-II) skálát használtam [15][27] (továbbiakban BDI-skála), amely az adott személy depressziós súlyosságát írja le. A BDI-skála 0-tól 63-ig terjed, ahol a 0 a teljesen egészséges állapotot jelöli, míg a 63 a legsúlyosabb depressziós állapotot. A BDI-skála alapján 4 féle kategóriát különbözetnek meg (1. táblázat).

1. táblázat. Beck Depression Inventory II kategóriái

BDI-érték	Kategória neve
0-13	Nem depressziós
14-19	Enyhe depresszió
20-28	Közepes depresszió
29-63	Súlyos depresszió

4.1.3. Statisztikai módszerek

Kutatásaim során először független kétmintás t-próbát alkalmaztam annak vizsgálatára, hogy a depressziós állapot a beszéd mely paramétereinek várható értékére van hatással, vagyis a depressziós állapotban lévő személyek mely beszédparaméter értékeinek az átlaga mutat szignifikáns különbséget az egészséges személyek beszédparamétereihez képest. A t-próba alkalmazása előtt minden esetben megvizsgáltam, hogy az adott minta jól közelíthető-e normális eloszlással. Ehhez Kolmogorov-Szmirnov-próbát használtam 99%-os szignifikancia szint mellett. A kétmintás t-próba használata előtt minden esetben ellenőriztem, hogy a két minta szórása egyezőnek tekinthető-e 95%-os szignifikancia szint mellett Levene-próba alapján, és ennek a próbának megfelelő t-próbát alkalmaztam.

Kutatásaim során Pearson korrelációval vizsgáltam, hogy mely beszédparaméterek értékei mutatnak szignifikáns korrelációt a depresszió súlyosságát leíró BDI-értékkel.

4.1.4. Alkalmazott gépi tanuló eljárások

Kutatásaim során kétféle gépi tanuló eljárást adaptáltam a kutatási céljaim elérése érdekében: support vector machine (SVM) és support vector regression (SVR) [28][29]. A kétféle gépi tanuló eljáráson kívül kipróbáltam más eljárásokat (knn, neurális hálók) is, azonban SVM/SVR eljárásokkal kaptam a legjobb eredményeket, feltehetőleg a beszédadatbázis relatív kis mérete miatt.

SVM gépi tanuló eljárás segítségével két- (nem depressziós/depressziós), illetve négyosztályos (nem depressziós/enyhén depressziós/közepesen depressziós/súlyosan depressziós) osztályozási feladatokhoz alkalmas modelleket hoztam létre.

A depresszió súlyosságának becsléséhez pedig SVR eljárás segítségével regressziós feladat megvalósításához alkalmas modelleket hoztam létre.

Az SVM és az SVR eljárások kernel függvényének és hiperparamétereinek megválasztását kimerítő kereséssel végeztem. Kernel függvények közül a lineáris, polinomiális (másod és harmadfokú) és rbf kerneleket teszteltem, míg a cost és gamma hiperparaméterek értékeinek a 2 hatványait teszteltem -10 és 10 közötti tartományban.

A modellépítés során (tanítás) a jellemzővektorban szereplő akusztikai-fonetikai jellemzők értékeit egyenként -1 és 1 közé normáltam.

4.1.5. Kutatás során vizsgált akusztikai-fonetikai jellemzők

Korábbi tapasztalataim alapján, a következő akusztikai-fonetikai jellemzőket továbbiakban **alapjellemzőket** vizsgáltam meg a kutatásaim során. Alapfrekvencia (f0), jitter, shimmer, első és második formánsfrekvenciák (F1 és F2) és azok sávszélességei (B1 és B2), intenzitás, mel-sávos energiaértékek, artikulációs sebesség (AS), relatív szünethossz (RSZH) - szünetek hossza a teljes beszédszakasz hosszához viszonyítva, tranziens szakaszok aránya (RoT) - a beszéd tranziens szakaszainak hossza a teljes beszédszakasz hosszához képest [30].

A beszéd egyes akusztikai-fonetikai jellemzőinek az értéke nagyban függhet attól, hogy a beszéd mely részén mérjük őket. Emiatt két típusba soroltam a jellemzőket, az „**I. típus**”-ba azokat a jellemzőket soroltam, amelyek nagyban függhetnek a mérési helytől (jitter, shimmer, F1, F2, B1, B2, mel-sávos energiaértékek), és ezeket két különböző módszerrel is megmértem, amelyeket lejjebb részletezek. A „**II. típus**”-ba azokat jellemzőket soroltam, ahol nem érdemes a kétféle helyfüggő mérést elvégezni (f0, intenzitás), illetve azokat a jellemzőket, amelyek beszédhangonként nem értelmezettek (AS, RSZH, RoT).

Az **„I. típusú”** jellemzők mérése esetén a következő két módszert használtam:

„a módszer”: beszédhangszintű szegmentálást igénylő módszer, ahol a jellemző értékének a számítását ugyanazon beszédhang közepén elhelyezett megfelelő időablakban mértem, például „E” magánhangzó (SAMPA jelöléssel) közepén. Majd ezekből az értékekből származtattam a statisztikai jellemzőket. Ezekben az esetekben

azért az „E” hangot választottam, mivel ez volt a leggyakoribb magánhangzó a bemondott szövegben.

„b módszer”: egyszerű előfeldolgozást igénylő módszer, ahol a jellemző értékének a számítását az összes lehetséges helyen kimértem 10 ms-os lépésközzel. Majd ezekből az értékekből származtattam a statisztikai jellemzőket.

Mindkét számítási módszernek megvan a maga előnye és hátránya. A **„b módszer”** esetén nem szükséges bonyolultabb előfeldolgozása a beszédmintának, elég csupán egy beszéd/nem beszéd, illetve egy zöngés/nem zöngés meghatározás, ami automatikusan is könnyen és elég pontosan elvégezhető, azonban a beszéd természetéből adódóan a mért értékeknek így nagyobb lesz a szórásuk mind az egészséges, mind a depressziós beszélők esetén. Az **„a módszer”** esetén kisebb lehet a mért paraméter értékének a természetes szórása, így jobban tükrözheti a depressziós állapot okozta változást, azonban az automatikus pontos beszédhangszintű szegmentálás megnehezítheti a paraméterértékek mérését.

A **„II. típusú”** jellemzők mérése esetén az f0-t a zöngés szakaszokon 10 ms-os lépésközzel, az intenzitást a teljes beszédszakaszon 10 ms-os lépésközzel, az AS-t, a RoT-ot a teljes beszédszakaszon, míg az RSZH-t a teljes felvételen (elhagyva a felvétel elején és lévő szüneteket) lett mérve.

A mért alapjellemzők értékeiből (AS, RSZH és RoT-ot leszámítva, hiszen ezek már eleve leíró jellemzők) három származtatott (statisztikai) jellemzőt készítettem: átlag, szórás és százalékos tartomány (a mért paraméter értékeinek alsó és felső 1 százalékának elhagyása után képzett tartománya). A mel-sávos energiaértékeknél csak az átlagot vizsgáltam, míg az intenzitás értéke esetén csak a szórást és a százalékos tartományt. A beszédmintákat amplitúdó szerint csúcsra normáltam. Egy beszédmintából összesen 98 származtatott jellemzőt számoltattam ki (2. táblázat).

2. táblázat. Származtatott akusztikai-fonetikai jellemzők száma beszédmintánként

	Jellemző neve	Kétféle számolási módszer	Átlag	Szórás	Százalékos tartomány	Származtatott jellemzők száma
I. típusú jellemzők	jitter	igen	igen	igen	igen	6
	shimmer	igen	igen	igen	igen	6
	F1	igen	igen	igen	igen	6
	F2	igen	igen	igen	igen	6
	B1	igen	igen	igen	igen	6
	B2	igen	igen	igen	igen	6
	mel-sávos energiaértékek[1]	igen	igen	nem	nem	54
II. típusú jellemzők	f0	nem	igen	igen	igen	3
	intenzitás	nem	nem	igen	igen	2
	RSZH	nem	önmagában leíró jellemző			1
	AS	nem	önmagában leíró jellemző			1
	RoT	nem	önmagában leíró jellemző			1
	Összesen:					98

[1]: 27 mel-sáv került kiszámításra a beszédmintákból

4.1.6. Statisztikai alapú jellemzővektor kiválasztása

Mind az SVM mind az SVR pontosságát befolyásolja a megfelelő bemeneti jellemző vektor kiválasztása [24], vagyis a zajszerű, nem releváns jellemzők elhagyása. Sokféle jellemző kiválasztási módszer létezik [31][32], a kutatásaim során én ezek közül az alábbiakat használtam:

Független kétmintás t-próba alapú jellemző kiválasztást, amikor a depressziós állapottal szignifikáns eltérést mutató paramétereket választottam ki. *Pearson korreláció alapú* jellemző kiválasztást, amikor a depressziós állapot súlyosságával való szignifikánsan korreláló paramétereket választottam be a jellemzővektorba. Ezen eljárások előnye, hogy könnyen megvalósítható és független a később használt gépi tanuló eljárástól, a hátránya, hogy nem vizsgálja az egyes paraméterek egymásra gyakorolt hatását.

A harmadik jellemzővektor kiválasztó algoritmusnak *a Fast Forward Selectio*nt (FFS) használtam. Ennek az eljárásnak a lényege, hogy viszonylag gyorsan kiválaszt egy, az algoritmus által optimálisnak ítélt *n* hosszúságú jellemzővektort, ami az algoritmus által előállított eltérő hosszúságú szuboptimális jellemzővektorok közül a legjobban teljesít az osztályozás, illetve a regressziós eljárás esetében. Az eljárás az üres jellemzővektorból indul ki. Az *i*. lépésben rendelkezésére áll az algoritmus szerinti legjobb *i-1* hosszúságú jellemzővektor, és ezt bővíti ki a legjobb *i* hosszúságú jellemzővektorra, úgy, hogy megvizsgálja, hogy melyik eddig még fel nem használt jellemző hozzáadásával kapható a legjobb pontosan *i* hosszúságú jellemzővektor.

A jellemző kiválasztási módszereket együttesen is használtam, például az FFS eljárás meggyorsítása érdekében, a bemenő alapjellemző halmazt a *Független kétmintás t-próbával* vagy a *Pearson korreláció vizsgálattal* előválogattam.

4.1.7. Tesztelési eljárások

Az adatbázisokban szereplő viszonylag alacsony mintaszám miatt az ilyenkor szokásos, teljes keresztvalidációs eljárást (LOOCV – leave-one-out cross validation) használtam a tesztelések, az FFS alkalmazása és a hiperparaméterek optimalizálása során is. Az eljárás lényege, hogy a rendszer pontosságának a leírására mindegyik mintát pontosan egyszer használja fel, mint teszthalmaz, míg a maradék mintákat, mint tanítóhalmaz, és így a tesztelő mintákon kapott eredmények írják le a rendszer teljes mintahalmazon számított pontosságát. Természetesen az eljárás során a teszthalmaz és a tanítóhalmaz metszete mindig üres.

Emellett egyes esetekben (ekkor mindig külön kiemelem az adott esetnél) használtam a szokásos tanító-/tesztelőhalmaz felosztást is, hogy az általam elért eredményeket összehasonlíthassam más kutatók által közölt eredményekkel.

4.1.7. Kiértékelési módszerek

Az osztályozások kiértékeléséhez tévesztési mátrixot adtam meg, ahol az egyes cellákban lévő esetek számát százalékosan adtam meg az eredeti kategória szerint, így az átlóban a fedés (recall) értékek kerültek. Ennek a megoldásnak a fő előnye, hogy a kétosztályos osztályozás esetén az átlóban a specificitás és szenzitivitás értékek olvashatók ki, amelyek az orvosi gyakorlatban az egyik leggyakrabban használt leíró jellemzői egy diagnózis pontosságának. Emellett még a pontosságot (accuracy) - jól osztályozott minták száma osztva az összes mintával - használtam.

Regressziós feladatok pontosságának jellemzésére három leíró jellemzőt használtam, amelyek széles körben elterjedtek a regressziós eljárások pontosságának jellemzésére: az átlagos hibaértékét (MAE – mean absolute error), az átlagos négyzetes hibaértékének a gyökét (RMSE – root mean square error), illetve az eredeti BDI-értékeknek és az eljárás által becsült BDI-értékeknek a Pearson korrelációs együttható értékét.

4.2. Felhasznált beszédadatbázisok

Magyar, német és olasz nyelvű depressziós beszédadatbázisokat használtam a kutatásaim során.

A magyar nyelvű beszédminták gyűjtését a Semmelweis Egyetem Pszichiátriai és Pszichoterápiás Klinikájával együtt végeztük. A beszédminták gyűjtésénél törekedtünk arra, hogy a beszélők lefedjék a depresszió súlyosságának különböző fokozatait, az egészséges állapottól az egészen súlyos depresszióig. Erre az adatbázisra a későbbiek során mint „Magyar Depressziós Beszédadatbázis” hivatkozom.

A német nyelvű depressziós beszédadatbázis, egy része az „Audio-visual depressive language corpus”-nak (AViD-Corpus) [16]. Erre az adatbázisra a későbbiek során, mint „Német Depressziós Beszédadatbázis” hivatkozom.

Az olasz nyelvű depressziós beszédadatbázist a nemzetközi Coala projekt kereteiben gyűjtötték az olasz kollégák. Erre az adatbázisra a későbbiek során, mint „Olasz Depressziós Beszédadatbázis” hivatkozom.

Mindhárom adatbázisban a bemondóknak egy rövid mesét ("Az északi szél és a Nap") kellett felolvasniuk az anyanyelvükön, ami széles körben elterjedt a különböző beszéd vizsgálatokban. A felvételek csendes helyiségben kerültek rögzítésre 44,1 kHz mintavételi frekvenciával, 16 bites kvantálással, csiptetős mikrofonnal.

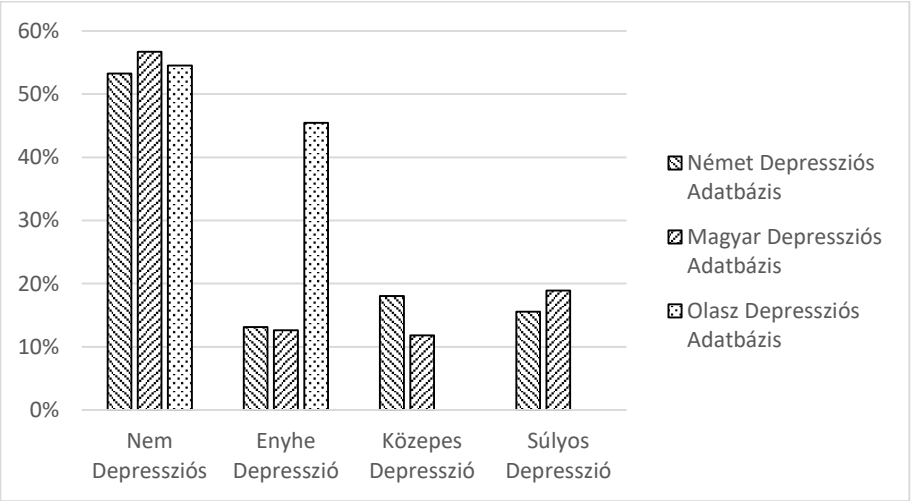
Az adatbázisba gyűjtött felvételekhez elkészítettem az egyes felvételekhez tartozó beszédhangszintű szegmentálást, az általam fejlesztett automatikus szegmentáló program (BME-TMIT-LSA nyelvfüggetlen kényszerített beszédhangszintű szegmentáló) segítségével.

Emellett minden felvételhez rögzítésre került a bemondó BDI-skála szerinti értéke. A BDI-értékek 0-tól 50-ig fordultak elő a vizsgált adatbázisokban.

A 3. táblázatban láthatók a három adatbázisban található bemondók főbb leíró jellemzői, míg az 1. ábra mutatja a három adatbázisban szereplő bemondók depresszió súlyossága szerinti eloszlásukat a BDI-skála alapján.

3. táblázat. A három depressziós adatbázis fő jellemzői

Nyelv	Bemondók száma (nők/férfiak)	Bemondók BDI szerinti átlaga és szórása	Bemondók életkorának átlaga és szórása
Magyar	127 (79/48)	14.2 (+-13.5)	42.2 (+-14.4)
Német	150 (99/51)	14.9 (+-11.7)	31.5 (+-12.3)
Olasz	11 (8/3)	12 (+-5.2)	nem állt rendelkezésre adat



1. ábra. BDI-értékek eloszlása a három depressziós adatbázisban.

5. Új eredmények

I. Téziscsoport: A Depressziós állapot hatásának statisztikai vizsgálata a beszéd akusztikai-fonetikai jellemzőire, a Magyar Depressziós Beszédadatbázis segítségével.

I.1. tézis: [B1, C4, C6, C7, J1] Statisztikai úton bizonyítottam, hogy a depressziós állapot szignifikáns változást okoz a beszéd egyes akusztikai-fonetikai alapjellemzőiből származtatott jellemzők átlagos értékében, az egészséges átlaghoz képest. 99%-os szignifikancia szinten az alábbi alapjellemzők származtatott értékei mutattak szignifikáns eltérést mindkét nem esetén: shimmer, F1, F2, B1, B2, mel-sávos energiaértékek (65-390 Hz és 1330-5190 Hz közöttiek), intenzitás.

Megvizsgáltam, hogy mely akusztikai-fonetikai alapjellemzők származtatott értékei mutatnak szignifikáns eltérést a depressziós és nem depressziós emberek beszédéből kinyert paraméterhalmazokat összehasonlítva. A vizsgálatokat nemek szerint külön és együttesen is elvégeztem. Mivel a beszéd legtöbb alapjellemzőjének értékét jelentősen befolyásolja a beszélő neme, így a nemek szerinti együttes elemzésnél, nemek szerint az egészséges minták alapján z-normalizáltam a paraméterek értékeit. Az egyes származtatott jellemzők átlagát hasonlítottam össze csoportok szerint független kétmintás t-próbával.

A 4. táblázatban láthatók az **„I. típusú”** jellemzők közül azok a jellemzők, amelyek legalább az egyik nem szerint, legalább az egyik statisztikai függvény szerint szignifikáns eltérést mutattak a két csoport között.

Az 5. táblázatban láthatók a **„II. típusú”** jellemzők esetén a szignifikáns eltérést mutató jellemzők, amelyek legalább az egyik nem szerint, legalább az egyik statisztikai függvény szerint szignifikáns eltérést mutattak a két csoport között.

A vizsgálat eredményeként elmondható, hogy több akusztikai-fonetikai alapjellemző származtatott értéke is szignifikáns eltérést mutatott a két csoportot összehasonlítva az **„I típusú”** jellemzőhalmaz esetében. A számítási módszereket **(„a és b” módszer”**) összehasonlítva kijelenthető, hogy jellemzően az **„a módszerrel”** számított jellemzők nagyobb szignifikáns eltérést mutattak, mint a **„b módszerrel”** számítottak. Emiatt a későbbi kutatásaimban az **„I. típusú”** jellemzőknél az **„a módszert”** használtam a jellemző értékeinek kinyeréséhez.

4. táblázat. Kétmintás t-próba eredményei az I. típusú akusztikai-fonetikai jellemzők esetében

Jellemző	Számítási módszer	Statisztikai függvény	Szignifikancia | Változás iránya					
			Együtt		Nők		Férfiak	
jitter	"a"	Átlag	**	↑	**	↑	-	↑
		Szórás	**	↑	**	↑	-	↑
		Sz. tartomány	*	↑	-	↑	*	↑
	"b"	Átlag	**	↑	*	↑	-	↑
		Szórás	*	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
shimmer	"a"	Átlag	***	↑	**	↑	*	↑
		Szórás	-	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
	"b"	Átlag	***	↑	***	↑	**	↑
		Szórás	-	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
F1	"a"	Átlag	***	↓	***	↓	***	↓
		Szórás	***	↑	**	↑	*	↑
		Sz. tartomány	-	↑	*	↑	-	↑
	"b"	Átlag	**	↓	-	↓	**	↓
		Szórás	-	↓	-	↓	-	↓
		Sz. tartomány	-	↓	-	↓	-	↓
F2	"a"	Átlag	***	↓	**	↓	**	↓
		Szórás	-	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
	"b"	Átlag	*	↓	-	↓	-	↓
		Szórás	*	↑	-	↑	-	↑
		Sz. tartomány	-	↑	*	↓	-	↑
B1	"a"	Átlag	***	↑	**	↑	***	↓
		Szórás	***	↑	***	↑	***	↓
		Sz. tartomány	***	↑	**	↑	**	↓
	"b"	Átlag	-	↑	-	↑	-	↓
		Szórás	-	↑	-	↑	-	↓
		Sz. tartomány	-	↑	-	↑	-	↓
B2	"a"	Átlag	***	↑	**	↑	**	↑
		Szórás	*	↑	-	↑	**	↑
		Sz. tartomány	*	↑	-	↑	***	↑
	"b"	Átlag	***	↑	***	↑	-	↑
		Szórás	-	↑	-	↑	-	↑
		Sz. tartomány	-	↑	-	↑	-	↑
„alacsony" mel-sávos energiaértékek[1]	"a"	Átlag	** (65-492Hz)	↑	** (65-603 Hz)	↑	** (65-390 Hz)	↑
	"b"	Átlag	** (65-390 Hz)	↑	** (65-390 Hz)	↑	** (65-214 Hz)	↑
„magas" mel-sávos energiaértékek[1]	"a"	Átlag	** (1330-6330 Hz)	↓	** (1330-6330 Hz)	↓	** (1330-5187 Hz)	↓
	"b"	Átlag	** (1330-5734 Hz)	↓	** (1330-5734 Hz)	↓	** (1330-4688 Hz)	↓

[1]: a „Szignifikancia" oszlopban zárójelek között feltűntettem, hogy a szignifikáns eltérését ($p<0.01$) mekkora tartományban kaptam

- : nem szignifikáns; * : $p < 0.05$; ** : $p< 0.01$; *** : $p<0.001$

Továbbiakban a vizsgálatból az is megállapítható, hogy a nagyobb átlag, szórás és százalékos tartomány értékek jellemezték a depressziós csoportot a jitter, shimmer, B1 és B2 esetében, míg kisebb átlag érték jellemezte a depressziós csoportot F1 és F2 esetében. A mel-sávos energiaértékek esetében az alsó tartományokban (65-390 Hz

között) magasabb értékek jellemezték a depressziós csoportot, míg a felső tartományokban (1330-5190 Hz között) ennek ellenkezője volt megfigyelhető.

5. táblázat. Két mintás t-próba eredményei az II. típusú akusztikai-fonetikai jellemzők esetében

Jellemző	Statisztikai függvény	Szignifikancia | Változás iránya					
		Együtt		Nők		Férfiak	
f0	Átlag	**	↓	**	↓	*	↓
	Szórás	*	↓	*	↓	*	↓
	Sz. tartomány	*	↓	*	↓	*	↓
intenzitás	Szórás	***	↓	***	↓	***	↓
	Sz. tartomány	***	↓	***	↓	**	↓
AS	nincs	**	↓	*	↓	**	↓
RSZH	nincs	*	↑	*	↑	**	↑
RoT	nincs	*	↓	*	↓	*	↓

- : nem szignifikáns; * : $p < 0.05$; ** : $p< 0.01$; *** : $p<0.001$

A **„II. típusú”** jellemzők esetében az összes vizsgált jellemző, mindkét nem esetében szignifikáns eltérést mutatott a két csoport között. A depressziós csoportra a kisebb artikulációs sebesség és tranziens arány és a magasabb relatív szünetek hossza volt jellemző. Emellett a depressziós csoportot még az alapfrekvencia kisebb átlagos értéke, szórása és kisebb százalékos tartománya, illetve az intenzitás kisebb szórás értéke és százalékos tartománya jellemezte.

I.2. tézis: [J1] Statisztikai úton bizonyítottam, hogy a beszéd egyes akusztikai-fonetikai alapjellemzőinek származtatott értékei szignifikáns korrelációt mutatnak mindkét nem esetében a beszédminta depresszió súlyosságát leíró Beck Depression Inventory (BDI) skálán mért értékével. 99%-os szignifikancia szinten az alábbi alapjellemzők mutattak szignifikáns korrelációt mindkét nem esetén: shimmer, F1, B2, mel-sávos energiaértékek (65-390Hz; 1518-4688Hz), intenzitás.

Korreláció vizsgálatot végeztem a beszédminták BDI-skálán mért értékei és az akusztikai-fonetikai alapjellemzők azon származtatott értékei között, amelyeknél szignifikáns különbséget kaptam a depressziós és nem depressziós csoportok között az **I.1. tézisben**. A vizsgálatot Pearson korreláció módszerrel nemek szerint együtt és elkülönítve végeztem. Az együttes vizsgálat esetén nemek szerint normalizáltam a jellemzőket az **I.1. tézisben** leírtak szerint. Azok az alapjellemzők, amelyek legalább 95%-os szignifikancia szint mellett szignifikáns korrelációt mutattak a depresszió súlyosságát leíró BDI-értékkel legalább az egyik nem esetén, a 6. táblázatban láthatók.

6. táblázat. Akusztikai-fonetikai alapjellemzők és a BDI-értékek Pearson féle korrelációja

Jellemző	Korreláció értéke | Szignifikanciaszint					
	Együtt		Nők		Férfiak	
jitter átlaga	0.22	*	0.25	*	0.2	*
jitter szórása	0.22	*	0.26	*	0.19	*
shimmer átlaga	0.34	**	0.3	**	0.43	**
shimmer szórása	0.24	**	0.35	**	0.2	*
F1 átlaga	-0.6	**	-0.62	**	-0.62	**
F1 szórása	-0.25	**	-0.29	**	-0.2	-
F2 átlaga	-0.29	**	-0.27	*	-0.36	*
B1 átlaga	-0.31	**	-0.27	*	-0.39	**
B1 szórása	-0.37	**	-0.44	**	-0.36	*
B1 sz. tartománya	-0.33	**	-0.34	**	-0.36	*
B2 átlaga	0.32	**	0.28	**	0.41	**
B2 szórása	0.23	**	0.1	-	0.44	**
B2 sz. tartománya	0.25	**	0.05	-	0.5	**
„alacsony” mel-sávos energiaértékek[1]	0.54 -0.38 (65-390 Hz)	**	0.51 - 0.36 (65-492 Hz)	**	0.58 - 0.41 (65-390 Hz)	**
„magas” mel sávos energiaértékek[1]	-0.47 - -0.37 (1330-5735 Hz)	**	-0.43 - -0.37 (1518-4688 Hz)	**	-0.53 - -0.37 (1518-4688 Hz)	**
f0 átlaga	-0.25	**	-0.32	**	-0.15	-
f0 Sz. tartománya	-0.35	**	-0.55	**	-0.23	*
intenzitás szórása	-0.51	**	-0.54	**	-0.46	**
intenzitás sz. tartománya	-0.53	**	-0.58	**	-0.44	**
AS	-0.29	**	-0.22	*	-0.54	**
RSZH	0.25	**	0.23	*	0.28	*
RoT	-0.28	**	-0.27	*	-0.34	*

[1]: a „Korreláció értéke” oszlopban zárójelek között feltűntettem, hogy a szignifikáns korrelációt ($p<0.01$) mekkora tartományban kaptam

-: nem szignifikáns; *: $p < 0.05$; **: $p< 0.01$; ***: $p<0.001$

A vizsgált alapjellemzők közül 99%-os szignifikancia szinten a B2 átlaga, a shimmer átlaga, az F1 átlaga, az intenzitás szórása és százalékos tartománya, és a mel-sávos energiaértékek mutattak szignifikáns korrelációt a BDI-értékekkel mindkét nem esetében, bár maga a korreláció mértéke nem mondható kifejezetten erősnek (±0.37 - ±.58). A kapott eredményeket összehasonlítva az **I.1 tézisben** kapott eredményekkel, azt a következtetést lehet levonni, hogy bár több alapjellemző származtatott értéke is szignifikánsan megváltozik a depresszió hatására, ugyanakkor a depresszió súlyosságával való lineáris kapcsolatuk már nem olyan erős.

II. Téziscsoport: Depressziós állapot detektálásának vizsgálata beszédjel feldolgozása alapján, support vector machine (SVM) gépi tanuló eljárás használatával, a Magyar Depressziós Beszédadatbázis segítségével.

II.1. tézis: [C4, J1] Kimutattam, hogy SVM gépi tanuló eljárás alkalmazásával depressziós és egészséges beszélőktől származó beszédminták egymástól elkülöníthetők, továbbá a depresszió négyfokozatú súlyossága szerinti osztályozás is megvalósítható. Kétosztályos osztályozás esetén 68%-os pontosságot, négyosztályos osztályozás esetén 52%-os pontosságot értem el, a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel és nemek szerinti közös modellel.

Két- (nem depressziós/depressziós), illetve négyosztályos (nem depressziós/enyhén depressziós/közepesen depressziós/súlyosan depressziós) osztályozási modelleket hoztam létre SVM gépi tanuló eljárást használva.

A jellemzővektor kialakításában az összes akusztikai-fonetikai alapjellemzőt felhasználtam, és közös modellt alakítottam ki nőkre és férfiakra. Mivel a nők és a férfiak egyes beszédjellemzőinek értékei nagyban különbözhetnek, a jellemzők értékeit nemek szerint normáltam, a nem depressziós minták alapján az **I.1 tézisben** leírtaknak megfelelően.

A modellek tesztelésével előállított tévesztési mátrixok a 7. táblázatban láthatók a kétosztályos modell esetén, és a 8. táblázatban láthatók a négyosztályos osztályozási modell esetén.

7. táblázat. Kétosztályos osztályozás tévesztési mátrixa, nemek szerinti közös modellel, az akusztikai-fonetikai alapjellemzők alapján

		Osztályozott kategória	
		Nem depressziós	**Depressziós**
Eredeti kategória	**Nem depressziós**	**65%**	35%
	Depressziós	27%	**73%**

68%-os pontosságot értem el (7. táblázat) kétosztályos osztályozás esetén nemek szerinti közös modellel az összes alapjellemzőt felhasználva.

8. táblázat. Négyosztályos osztályozás tévesztési mátrixa, nemek szerinti közös modellel, az akusztikai-fonetikai alapjellemzők alapján

		Osztályozott kategória			
		Nem depressziós	**Enyhén depressziós**	**Közepesen depressziós**	**Súlyosan depressziós**
Eredeti kategória	**Nem depressziós**	**64%**	28%	7%	1%
	Enyhén depressziós	29%	**21%**	21%	29%
	Közepesen depressziós	14%	15%	**43%**	28%
	Súlyosan depressziós	11%	4%	45%	**40%**

52%-os pontosságot értem el (8. táblázat) négyosztályos osztályozás esetén nemek szerinti közös modellel az összes akusztikai-fonetikai alapjellemzőt felhasználva.

Ezekre az eredményekre (7. táblázat és 8. táblázat) a későbbi kutatásaimban, mint „kiindulási eredmények” hivatkozom.

II.2. tézis: [J1] Kimutattam, hogy statisztikai alapú jellemzővektor kiválasztással akár 12% relatív hibacsökkenés is elérhető a kétosztályú osztályozás esetében, és akár 11% relatív hibacsökkenés is elérhető a négyosztályú osztályozás esetében a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel kapott eredményekhez képest.

A **II.1. tézisben** bemutatott modelleket úgy módosítottam, hogy a jellemzővektor kialakítása során, azokat az akusztikai-fonetikai jellemzőket használtam fel csupán, amelyek 95%-os szignifikáns eltérést mutattak mindkét nem esetén. Fontos megjegyeznem, hogy a 95%-os szignifikancia szint megválasztása önkényes volt, azonban az eredmények alapján egyértelműen látszik, hogy egy jó kiindulási alap egy későbbi optimalizáláshoz.

A modellek tesztelésével előállított tévesztési mátrixok a 9. táblázatban láthatók a kétosztályos modell esetén, és az 10. táblázatban láthatók a négyosztályos modell esetén.

9. táblázat. Kétosztályos osztályozás tévesztési mátrixa, nemek szerinti közös modellel, válogatott akusztikai-fonetikai jellemzők alapján

		Osztályozott kategória	
		Nem depressziós	**Depressziós**
Eredeti kategória	**Nem depressziós**	**72%**	28%
	Depressziós	29%	**71%**

72%-os pontosságot értem el (9. táblázat) kétosztályos osztályozás esetén nemek szerinti közös modellel statisztikai jellemző kiválasztási módszerrel. A „kiindulási eredményhez” viszonyítva 12,5%-os relatív csökkenést értem el a hibásan osztályozott személyek számában (7. táblázat és 9. táblázat).

10. táblázat. Négyosztályos osztályozás tévesztési mátrixa, nemek szerinti közös modellel, válogatott akusztikai-fonetikai jellemzők alapján

		Osztályozott kategória			
		Nem depressziós	**Enyhén depressziós**	**Közepesen depressziós**	**Súlyosan depressziós**
Eredeti kategória	**Nem depressziós**	**72%**	21%	7%	0%
	Enyhén depressziós	36%	**36%**	21%	7%
	Közepesen depressziós	24%	19%	**36%**	21%
	Súlyosan depressziós	18%	22%	19%	**41%**

57%-os pontosságot értem el (10. táblázat) négyosztályos osztályozás esetén nemek szerinti közös modellel statisztikai jellemző kiválasztási módszerrel. A „kiindulási eredményhez” viszonyítva 11%-os relatív csökkenést értem el a hibásan osztályozott személyek számában (8. táblázat és 10. táblázat).

A modellek tesztelése alapján kijelenthető, hogy a statisztikai jellemző kiválasztási módszerrel akár jelentős javulás is elérhető.

II.3. tézis:[J1] Kimutattam, hogy nemek szerint külön létrehozott és statisztikai jellemző kiválasztással optimalizált modellek alkalmazása jelentősen javítja a depresszió osztályozási pontosságát. Az optimalizált modellel 56% relatív csökkenést értem el a hibásan osztályozott személyek számában a kétosztályú osztályozás esetében, és 35% relatív csökkenést értem el a hibásan osztályozott személyek számában a négyosztályú osztályozás esetében, a Magyar Depressziós Beszédadatbázison, az akusztikai-fonetikai alapjellemzőkkel kapott eredményekhez képest.

A **II.2. tézisben** bemutatott modelleket úgy módosítottam, hogy nemek szerint eltérő jellemzővektorokat alakítottam ki az **I.1 tézisben** bemutatott statisztikai vizsgálat eredménye alapján, vagyis az adott nem szerint szignifikáns (95%) különbséget mutató paramétereket választottam ki, és így nemek szerint eltérő modelleket hoztam létre.

Az osztályozási modellek tesztelésével előállított tévesztési mátrixok a 11. táblázatban láthatók a kétosztályos modell esetén, és az 12. táblázatban láthatók a négyosztályos modell esetén.

11. táblázat. Kétosztályos osztályozás tévesztési mátrixa, nemek szerinti eltérő modellel, válogatott akusztikai-fonetikai jellemzők alapján

		Osztályozott kategória	
		Nem depressziós	**Depressziós**
Eredeti kategória	**Nem depressziós**	**88%** (nők:84%, férfiak: 93%)	12% (nők: 16%, férfiak: 7%)
	Depressziós	16% (nők: 20%, férfiak: 10%)	**84%** (nők: 80%, férfiak: 90%)

86%-os pontosságot értem el (11. táblázat) kétosztályos osztályozás esetén nemek szerinti eltérő modellel statisztikai jellemző kiválasztási módszerrel, 92%-os volt a pontosság a férfiak esetében, 82%-os volt a pontosság a nők esetében. A „kiindulási eredményhez” viszonyítva 56%-os relatív javulást értem el a hibásan osztályozott személyek számában (7. táblázat és 11. táblázat).

12. táblázat. Négyosztályos osztályozás tévesztési mátrixa, nemek szerinti eltérő modellel, válogatott akusztikai-fonetikai jellemzők alapján

		Osztályozott kategória			
		Nem depressziós	**Enyhén depressziós**	**Közepesen depressziós**	**Súlyosan depressziós**
Eredeti kategória	**Nem depressziós**	**87%** (nők: 89%, férfiak: 85%)	10% (nők:10%, férfiak: 10%)	3% (nők:2%, férfiak: 4%)	0%
	Enyhén depressziós	33% (nők: 30%, férfiak: 40%)	**40%** (nők: 40%, férfiak 40%)	27% (nők: 30%, férfiak 20%)	0%
	Közepesen depressziós	7% (nők11%, férfiak 0%)	20% (nők:11%, férfiak 33%)	**53%** (nők:55%, férfiak 50%)	20% (nők:22%, férfiak: 17%)
	Súlyosan depressziós	8% (nők:6%, férfiak 11%)	12% (nők13%, férfiak 11%)	36% (nők:50%, férfiak: 11%)	**44%** (nők:31%, férfiak: 67%)

69%-os pontosságot értem el (12. táblázat) négyosztályos osztályozás esetén nemek szerinti eltérő modellel statisztikai jellemző kiválasztási módszerrel. 74%-os volt a pontosság a férfiak esetében, 66%-os volt a pontosság a nők esetében. A „kiindulási eredményhez” viszonyítva 35%-os relatív javulást értem el a hibásan osztályozott személyek számában (8. táblázat és 12. táblázat).

Az eredmények alapján kijelenthető, hogy az eltérő modell használata nemek szerint jelentős javulást eredményez.

III. Téziscsoport: A depresszió súlyosság szerinti becslésének vizsgálata a beszédjel különböző típusú előfeldolgozása alapján, support vector regression (SVR) gépi tanuló eljárás használatával, a Magyar Depressziós Beszédadatbázis segítségével.

III.1. tézis: [C5] Kimutattam, hogy SVR gépi tanuló eljárás alkalmazásával a beszélő depressziós állapotának súlyossága becsülhető egyszerű (beszédhangszintű szegmentálást nem igénylő) akusztikai-fonetikai jellemzőkkel. A modelleket nemek szerint elkülönítve, statisztikai jellemző kiválasztással optimalizált módon hoztam létre. Az elért átlagos négyzetes hiba gyökére (RMSE) 8,4, míg az átlagos hibára 6,7 (MAE) értékeket kaptam.

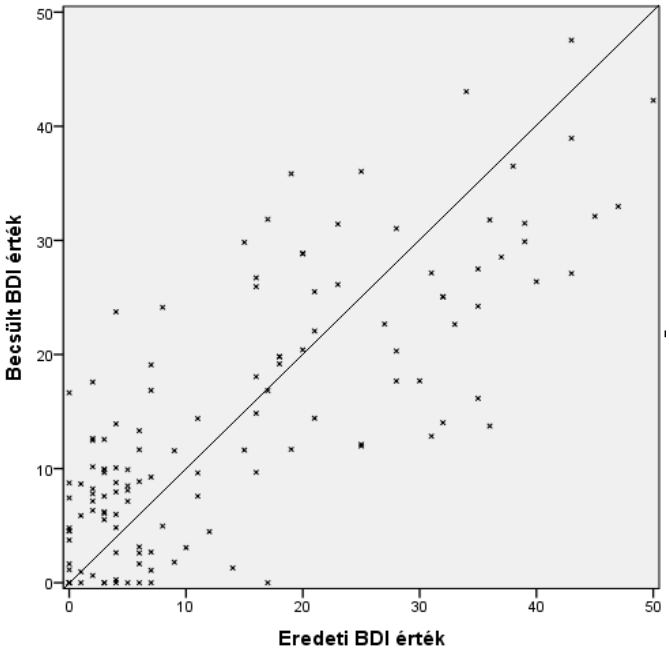
SVR gépi tanuló eljárás használatával regressziós modellt hoztam létre külön női és férfi beszélőkre, ami a beszéd jellemzői alapján becsli a beszélő depressziós súlyosságának mértékét a BDI-skála szerint.

A jellemzővektor kialakításában kizárólag az olyan alapjellemzőket használtam fel, amelyek szignifikáns korrelációt mutattak a BDI-skálával, külön a női vagy a férfi beszélők esetében, és az értékük beszédhangszintű szegmentálás nélkül kiszámíthatók, így az **I.2. tézisben** bemutatott jellemzők értékeit a **„b módszerrel”** számíttattam ki. Az így kialakított jellemzőhalmazból FFS jellemző kiválasztási algoritmussal választottam ki a végleges jellemzővektort.

A beszédhangszintű szegmentálás nélkül kinyert akusztikai-fonetikai jellemzőkkel épített modell pontosságnak leíró jellemzői a 13. táblázatban láthatók. A 2. ábra mutatja az az egyes beszélők eredeti és becsült BDI-értékeit.

13. táblázat. Depresszió súlyosság becslésének hibája beszédhangszintű szegmentálás használata nélkül

	Együtt	**Nők**	**Férfiak**
MAE	6,76	7,10	6,20
RMSE	8,37	8,70	7,80
Pearson korreláció	0,79	0,73	0,79



2. ábra. A regressziós modellel becsült, és az eredeti BDI-értékek összehasonlítása, ha a modell építéséhez használt akusztikai-fonetikai jellemzők a „b" módszer alkalmazásával lettek számítva.

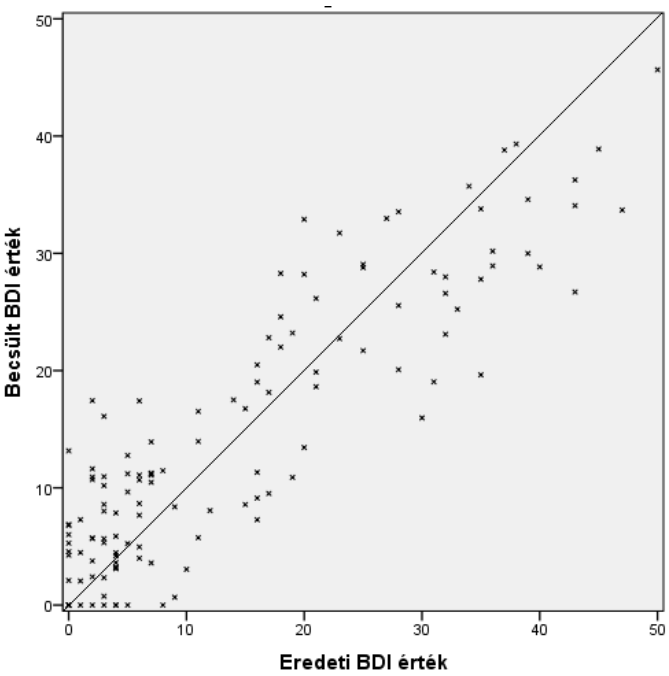
III.2. tézis: [C5] A beszédhangszintű feldolgozással végzett jellemzőkinyerés javít a becslés pontosságán. Az egyszerű előfeldolgozással mérhető (beszédhangszintű szegmentálás nélküli) akusztikai-fonetikai jellemzőkkel kialakított modellel elért eredményhez képest 25%-os relatív csökkenést értem el az RMSE hibaértékben, és 29%-os relatív csökkenést értem el a MAE hibaértékben.

A **III.1. tézisben** bemutatott regressziós eljárást úgy módosítottam, hogy az **I. típusú** akusztikai-fonetikai jellemzőket az **„a módszerrel"** számítottam ki. Vagyis a jellemzők kiszámíttatásához felhasználtam az automatikus beszédhangszintű szegmentáló kimenetelét.

A beszédhangszintű szegmentálás feldolgozásával nyert akusztikai-fonetikai jellemzőkkel épített modell pontosságnak leíró jellemzői a 14. táblázatban láthatók. A 3. ábra mutatja az az egyes beszélők eredeti és becsült BDI-értékeit.

14. táblázat. Depresszió súlyosság becslésének hibája beszédhangszintű szegmentálás felhasználásával

	Együtt	Nők	Férfiak
MAE	5,1	5,7	4,1
RMSE	6,3	7	5
Pearson korreláció	0,89	0,87	0,92



3. ábra. A regressziós modellel becsült és az eredeti BDI-értékek összehasonlítása, ha a modell építéséhez használt akusztikai-fonetikai jellemzők az „a módszer" alkalmazásával lettek számítva.

A **III.1. tézisben** bemutatott eredményekhez képest a MAE esetében 29%-os relatív csökkenést értem el, a RMSE estében 25%-os relatív csökkenést értem el, míg a Perason korrelációs együttható esetében 11%-os relatív javulást értem el a pontosabb előfeldolgozó eljárás használatával (beszédhang szintű szegmentálás használatával).

IV. Téziscsoport: A Depressziós állapotot tükröző akusztikai-fonetikai jellemzők nyelvfüggőségének vizsgálata különböző nyelvű depressziós beszédadatbázisokon

IV. 1. tézis: [B1, C3, J1] Statisztikai úton kimutattam, hogy az alkalmazott akusztikai-fonetikai jellemzőkészletből kiválasztható olyan részhalmaz, amely magyar és német nyelv esetén is szignifikáns korrelációt mutat a beszélők depressziós állapotuk szerinti súlyosságával. Az alábbi jellemzők mutattak 95%-os szignifikancia szinten szignifikáns korrelációt mindkét nem esetén: f0, jitter, shimmer, B2, intenzitás, AS, RSZH, RoT.

Korreláció vizsgálatot végeztem a beszélők BDI-skálán mért értékei és a beszédükből kinyert akusztikai-fonetikai alapjellemzők értékei között. A vizsgálatot Pearson korreláció módszerrel a Magyar Depressziós Adatbázison és a Német Depressziós Adatbázison külön-külön és a két adatbázis egyesítésén is elvégeztem nemek szerint elkülönítve. Azok az alapjellemzők, amelyek legalább 95%-os szignifikancia szint mellett szignifikáns korrelációt mutattak a depresszió súlyosságát leíró BDI-értékkel mindhárom vizsgált adatbázis esetében a 15. táblázatban láthatók.

15. táblázat. Az akusztikai-fonetikai alapjellemzők és a BDI-értékek közötti korreláció elemzés eredményei a magyar és a német beszélők esetében

	Korrelációs együttható mértéke | Szignifikancia szint											
	Német				Magyar				Német + Magyar			
	Nők		Férfiak		Nők		Férfiak		Nők		Férfiak	
B2 átlaga	0.19	*	0.39	*	0.28	**	0.41	**	0.32	**	0.41	**
B2 szórása	0.12	-	0.24	*	0.10	-	0.44	**	0.13	-	0.35	**
jitter átlaga	0.19	*	0.18	*	0.25	*	0.2	*	0.2	*	0.2	*
jitter szórása	0.17	*	0.38	*	0.26	*	0.19	*	0.17	*	0.22	*
shimmer átlaga	0.24	*	0.19	*	0.3	**	0.43	**	0.25	**	0.22	**
shimmer szórása	0.21	*	0.19	*	0.35	**	0.2	*	0.22	**	0.19	*
AS	-0.64	**	-0.23	*	-0.22	*	-0.54	**	-0.28	*	-0.41	**
f0 sz. tartománya	-0.38	*	-0.45	**	-0.55	**	-0.23	*	-0.49	**	-0.32	**
intenzitás szórása	-0.23	*	-0.32	*	-0.54	**	-0.46	**	-0.26	*	-0.27	**
RSZH	0.34	**	0.21	*	0.23	*	0.28	*	0.19	*	0.26	*
RoT	-0.17	*	-0.32	*	-0.27	*	-0.34	*	-0.20	*	-0.21	*

- : nem szignifikáns; * : $p < 0.05$; ** : $p < 0.01$; *** : $p<0.001$

A következő akusztikai-fonetikai alapjellemzők mutattak szignifikáns korrelációt a BDI-értékkel (11. táblázat): B2, shimmer, jitter, artikulációs sebesség (AS), alapfrekvencia (f0), intenzitás, relatív szünethossz (RSZH), RoT.

A vizsgálat alapján kijelenthető, hogy a depressziós állapot hatása hasonló módon jelenik meg a fenti akusztikai-fonetikai jellemzők értékében (15. táblázat) a vizsgált két nyelv (német és magyar) esetében.

IV.2. tézis: [B1, C1, C2, C3, J1, J2] Megmutattam, hogy lehetséges olyan SVR alapú regressziós modell létrehozása, amely megfelelő jellemző kiválasztási algoritmus használatával, a tanító beszédmintáktól eltérő anyanyelvű beszélők (vizsgált nyelvek: magyar, német, olasz) depressziós állapotának becslésére is alkalmas.

SVR gépi tanuló eljárás használatával több regressziós modellt is létrehoztam minden esetben külön női és férfi beszélőkre, a depresszió nyelvek közötti becslésének pontosságának vizsgálatára. Vizsgálataim során kétféle tesztelési eljárást alkalmaztam, az egyik esetben az aktuálisan vizsgált adatbázist (adatbázisokat) felosztottam tanító- és tesztelőhalmazra (normál), a másik esetben a már korábban is használt teljes keresztvalidációs kiértékelést használtam (LOOCV). A megvalósított modellek tesztelésének eredményeit a 16. táblázatban láthatók.

Létrehoztam három kiindulási modellt, ahol jellemző kiválasztási eljárás nélkül az összes akusztikai-fonetikai alapjellemzőt felhasználtam a modell építése során (16. táblázat 1-3 sora). A modellek kiértékelését LOOCV eljárással valósítottam meg, és egynyelvű magyar és német, illetve többnyelvű magyar+német modelleket hoztam létre. Ezzel a három modellel elért eredményeket mint kiindulási eredményeket használtam fel a később létrehozott modelljeim eredményeinek kiértékeléséhez.

A kiindulási eredmények esetében a rendszer csak nagy hibával volt képes becsülni a beszélő depressziós súlyosságát, és ahogy várható volt, az egyesített magyar és német nyelvű modell esetében (16. táblázat 3. sor) kaptam a legnagyobb hibaértékeket.

16. táblázat. Regressziós eredmények egy- és többnyelvű modellek esetében különböző tanító és tesztelő adathalmazok esetében

Eljárás neve	Jellemző kiválasztás	Tanító adathalmaz	Tesztelő adathalmaz	RMSE	MAE
1. Kiindulási_Német	nem	Német Adatbázis	LOOCV	14.56 nők: 12.62 férfiak: 17.69	11.58 nők: 9.99 férfiak: 14.62
2. Kiindulási_Magyar	nem	Magyar Adatbázis	LOOCV	12.43 nők: 12.65 férfiak: 12.04	9.8 nők: 10.13 férfiak: 9.26
3. Kiindulási_Magyar+Német	nem	Magyar+Né met Adatbázis	LOOCV	19.7 nők: 22.57 férfiak: 13.13	15.13 nők: 17.86 férfiak: 10.28
4. Német _Normál	igen	Német Adatbázis Tanítóalmaza	Német Adatbázis Tesztelőhalmaz a	8.99 nők: 9.04 férfiak: 8.92	7.54 nők: 7.86 férfiak: 7.02
5. Magyar_LOOCV	igen	Magyar Adatbázis	LOOCV	8.07 nők: 7.78 férfiak: 8.65	6.31 nők: 6.07 férfiak: 6.8
6. Német_ LOOCV	igen	Német adatbázis	LOOCV	8.55 nők: 8.89 férfiak: 7.88	7.02 nők: 7.24 férfiak: 6.99
7. Magyar+Német_LOOCV	igen	Magyar+Né met Adatbázis	LOOCV	8.59 nők: 8.5 férfiak: 8.66	7.05 nők: 7.07 férfiak: 7.02
		Magyar+Né met Adatbázis	Magyar Adatbázis	8.27 nők:7.92 férfiak: 8.86	6.65 nők: 6.35 férfiak: 7.18
		Magyar+Né met Adatbázis	Német Adatbázis	8.92 nők: 8.99 férfiak: 7.96	7.33 nők: 7.52 férfiak: 7.12
8. Magyar_Német_Normál	igen	Magyar Adatbázis	Német Adatbázis	9.52 nők: 9.45 férfiak: 9.67	7.46 nők: 7.61 férfiak: 7.16
9. Német_Magyar_Normál	igen	Német Adatbázis	Magyar Adatbázis	9.83 nők: 9.75 férfiak: 9.98	7.72 nők: 7.36 férfiak: 8.29
10. Magyar+Német_Olasz_Nor-mál	igen	Magyar+Né met Adatbázis	Olasz Adatbázis	5.09 nők: 4.42 férfiak: 6.55	4.19 nők: 3.65 férfiak: 5.66

A további modellek építése során csak azokat az akusztikai-fonetikai jellemzőket használtam fel, amik **IV.1. tézisben** 95%-os szignifikancia szinten a Magyar Depressziós Adatbázis, a Német Depressziós Adatbázis és a két adatbázis egyesítése esetén is szignifikáns korrelációt mutattak a depresszió súlyosságával, és külön

modellt hoztam létre a női és férfi bemondókra. A jellemző kiválasztási algoritmus használata után az eredmények jelentősen javultak.

A Német Depressziós adatbázist használták az AVEC 2013 versenyben. A versenyben az adatbázist felosztották tanító- és tesztelőhalmazra és ez alapján tanítottam be én is a modellemet, és valósítottam meg a tesztelést (16. táblázat 4. sor) így lehetőségem volt összehasonlítani az akusztikai-fonetikai jellemző számítási és a többnyelvű jellemző kiválasztási módszeremet más kutatók nyelvspecifikus módszerével. A vizsgálat eredményét (RMSE: 8,99; MAE: 7,54) összehasonlítva az AVEC 2013 verseny „baseline" eredményével [16] (RMSE: 14,12; MAE: 10,35), jelentős javulás látható a depressziós állapot becslésében. Az általam elért eredmények alig maradnak el a verseny győztesének nyelvspecifikus eredményétől [17] (RMSE 8,68; MAE: 7,12).

További egy-, illetve többnyelvű modelleket hoztam létre, és teszteltem azokat, hogy megvizsgáljam a többnyelvű jellemző kiválasztási módszerem hatását a depressziós állapot becslésére (16. táblázat 5-7 sora). Az eredményeket összehasonlítva a kiindulási eredményekkel a legjelentősebb javulás a magyar+német többnyelvű modell esetében figyelhető meg, ahol az RMSE értékében 57%-os relatív csökkenést értem el, tehát kijelenthető, hogy az általam alkalmazott statisztikai jellemző kiválasztási módszerrel jelentős javulás érhető el többnyelvű modellek esetében.

Megvizsgáltam még azt is, hogy az általam használt módszerekkel milyen pontosságot lehet elérni, ha a tanító- és a tesztelőhalmazban más anyanyelvű beszélők vannak (16. táblázat 8-10 sora). Az így elért eredmények elmaradnak attól, mint amikor a tanításnál mindkét nyelv mintái szerepelnek (16. táblázat 5-7 sora), de az eredmények összehasonlíthatók, és magyar nyelvű tanítás német nyelven való tesztelésének (16. táblázat 8. sora) eredménye még mindig jelentősen jobb, mint az AVEC 2013 „baseline" eredménye.

Amikor az egyesített magyar+német adatbázissal tanítottam a modellt, és az olasz adatbázison teszteltem, kifejezetten jó eredményeket kaptam, azonban fontos megjegyeznem, hogy az olasz adatbázis beszédmintáinak száma alacsony, illetve a BDI-értékek szerinti eloszlása lényegesen eltérő a másik két adatbázistól (csak nem depressziós és enyhén depressziós személyek beszédmintáit tartalmazza az adatbázis).

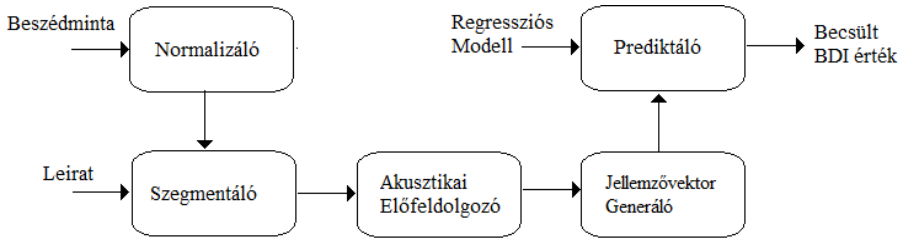
6. Az eredmények gyakorlati alkalmazhatósága

Részt vettem a European Space Agency (ESA) Coala projektjében. A projekt célja a Concordia kutatóállomáson áttelelő kutatók pszichológia állapotának beszédhang alapú megfigyelése volt, különösképpen a depressziós állapot kialakulásának jelzésére.

Kidolgoztam és megvalósítottam a beszélő depressziós állapotát beszédminta alapján becsülő rendszert. A megvalósított rendszer többnyelvű modell segítségével francia, magyar, német és olasz nyelven használható.

A megvalósított rendszer segítségével a Concordia kutatóállomáson dolgozó személyek depressziós állapotát monitoroztam. Az állomáson dolgozó francia és olasz kutatók nagyjából kétheti rendszerességgel felmondtak és rögzítettek egy rövid mesét („Az északi szél és a Nap) az anyanyelvükön. A felvételek rögzítésére 2013-ban, 2014-ben és 2016-ban került sor. Minden évben 10-12 kutató önkéntesen vett részt a kísérletben.

A feladatom volt a hangminták elemzése, és ez alapján a beszélők depressziós állapotának becslése. Mivel a kutatás során nem állt rendelkezésemre megfelelő francia és olasz nyelvű depressziós adatbázis, így fontos szempont volt, hogy az általam tervezett rendszer alkalmas legyen több nyelven is a depressziós állapot becslésére. Az elkészült rendszer folyamatábrája a 4. ábrán látható.



4. ábra. Depressziós állapot becslésére készült rendszer folyamatábrája

A rendszer bemenetként megkapja a beszédmintát, a beszélő nemét és anyanyelvét, és kimenetként megadja a beszélő becsült depressziós súlyosságát a BDI-skála szerint. A *Normalizáló* egység a vizsgált beszédmintát normalizálja amplitúdó szerint. A *Szegmentáló* egység a normalizált beszédminta és a beszélő nyelvének megfelelő leirat alapján elkészíti a beszédminta beszédhangszintű szegmentálását kényszerített

illesztéssel. A szegmentálás alapján az *Akusztikai Előfeldolgozó* egység kinyeri a megfelelő akusztikai-fonetikai jellemzőket. A *Jellemzővektor Generáló* egység a kiszámított jellemző értékeket a beszélő neme szerinti modell alapján transzformálja (mivel a modell tanítása során a jellemző értékek -1 és 1 közötti értékek közé lettek normalizálva), és előállítja a modellnek megfelelő jellemzővektort. A *Prediktáló* egység az elkészített jellemzővektor és a beszélő neme szerinti modell alapján elvégzi a depressziós állapot becslését a BDI-skála szerint, ami a rendszer végső kimenete.

A modellek tanítását a IV.2 tézisben bemutatott eljárással készítettem a Magyar és Német Depressziós Beszédadatbázisok beszédmintái alapján, annyi módosítással, hogy a statisztikai jellemző kiválasztási algoritmus után FFS algoritmust alkalmaztam a becslés pontosságának növelésére, amivel a tanító modell LOOCV eljárással való tesztelése esetén a hibaértékeket tovább tudtam csökkenteni (RMSE: 8,4; MAE 6,7).

A kialakított végleges rendszert 12 egészséges, különböző anyanyelvű személyen teszteltem (3 francia, magyar, német és olasz). A tesztelés során a hiba RMSE értékre 9,1-et, míg a MAE értékre 7,3-at kaptam.

A végleges rendszer nem csak a Concordia kutatóállomáson dolgozó kutatók pszichológiai állapotának a becslésére alkalmas, hanem általánosságban egy többnyelvű, a depresszió súlyosságát becslő olyan diagnosztikai rendszer megvalósításra, ami nem invazívan – egy mese felolvasásával – a depressziós állapot meglétére figyelmeztet. Például jól alkalmazható lehet előszűrésre körzeti orvosi rendelőkben.

7. Köszönetnyilvánítás

Szeretném megköszönni témavezetőmnek, Vicsi Klárának sok éves útmutatását, türelmét, bíztatását. Beszédakusztikai laboratórium csoport összes dolgozójának – elsősorban Sztahó Dávid, Tulics Miklós Gábriel, Nagy Ildikó, Tündik Máté Ákos, Szaszák György - és jelenlegi illetve korábbi hallgatójának észrevételeit, kritikáit és segítségét munkám során.

Szeretném megköszönni Beszédkommunikáció és Intelligens Interakciók Laboratórium több dolgozójának a kritikáit és ötleteit téziseim megvitatása során, névszerint: Németh Géza, Mihajlik Péter, Zainkó Csaba, Csapó Tamás Gábor, Tarján Balázs.

Szeretném megköszönni Simon Lajos és Dér Bálint pszichiátereknek a segítségét a felvételek gyűjtése során.

Szeretném megköszönni édesanyámnak, hogy lektorálta az értekezésemet. Szeretném megköszönni feleségemnek és fiaimnak, hogy mindenben támogattak kutatásaim elvégzése során.

8. Rövidítések, jelölések

AS	artikulációs sebesség
B1	első formánsfrekvencia sávszélessége
B2	második formánsfrekvencia sávszélessége
BDI	Beck Depression Inventory / Beck depressziós skála
DSM	The Diagnostic and Statistical Manual of Mental Disorders
f0	alapfrekvencia
F1	első formánsfrekvencia
F2	második formánsfrekvencia
FFS	fast forward selection
HAM-D	Hamilton Depression Rating Scale / Hamilton depressziós skála
LOOCV	leave one out crossvalidation / teljes keresztvalidáció
MAE	mean absolute error / átlagos abszolút hiba
RMSE	root mean square error / átlagos négyzetes hiba gyöke
RoT	ratio of transients / tranziensek aránya
RSZH	relatív szünethossz
SAMPA	Speech Assessment Methods Phonetic Alphabet / fonetikus abc
SVM	support vector machine / szupport vektor gép
SVR	support vector regression / support vector regresszió
WHO	World Health Organization / Egészségügyi Világszervezet

9. Irodalomjegyzék

[1] Godino-Llorente, J. I., Gomez-Vilda, P. (2004). Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors. *IEEE Transactions on Biomedical Engineering*, *51*(2), 380-384.

[2] Vicsi, K., Sztahó, D. (2012). Recognition of Emotions on the Basis of Different Levels of Speech Segments. *JACIII*, *16*(2), 335-340.

[3] El Ayadi, M., Kamel, M. S., Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*, *44*(3), 572-587.

[4] Cummins, N., Epps, J., Breakspear, M., Goecke, R. (2011). An investigation of depressed speech detection: Features and normalization. *In Proceedings of INTERSPEECH 2011* (pp. 2997-3000).

[5] SztahóD., Tulics, M. G., Vicsi, K., Valálik, I. (2017). Automatic estimation of severity of Parkinson's disease based on speech rhythm related features. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 11-16).

[6] Orozco-Arroyave, J. R., Hönig, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Daqrouq, K., Skodda, S., ... Nöth, E. (2016). Automatic detection of Parkinson's disease in running speech spoken in three different languages. *The Journal of the Acoustical Society of America*, *139*(1), 481-500.
[7] World Health Organization. (2017). *Depression and other common mental disorders: global health estimates* (No. WHO/MSD/MER/2017.2). World Health Organization.
[8] Friedrich, M. J. (2017). Depression is the leading cause of disability around the world. *Jama*, *317*(15), 1517-1517.
[9] Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D., Saxena, S. (2012). Depression: A global public health concern (www.who.int).
[10] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
[11] Nilsonne, A. (1988). Speech characteristics as indicators of depressive illness. *Acta Psychiatrica Scandinavica*, *77*(3), 253-263.
[12] Kraepelin, E., 1921. Manic depressive insanity and paranoia. J. Nerv. Ment. Dis. 53, 350.
[13] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10-49.
[14] Hamilton, H., 1960. HAMD: a rating scale for depression. Neurosurg. Psych. 23, 56–62.
[15] Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety
[16] Valstar, M., Schuller, B., Smith, K., Eyben, F., Jiang, B., Bilakhia, S., Schnieder, S., Cowie, R. and Pantic, M. (2013, October). AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, ACM, pp. 3-10.
[17] Williamson, J. R., Quatieri, T. F., Helfer, B. S., Horwitz, R., Yu, B., Mehta, D. D. (2013, October). Vocal biomarkers of depression based on motor incoordination. In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, ACM, pp. 41-48.
[18] Hönig, F., Batliner, A., Nöth, E., Schnieder, S., Krajewski, J., 2014. Automatic modelling of depressed speech: relevant features and relevance of gender. In: Proceedings of Interspeech. Singapore, pp. 1248–1252.
[19] Alpert, M., Pouget, E.R., Silva, R.R., 2001. Reflections of depression in acoustic measures of the patient's speech. J. Affect. Disord. 66, 59–69.
[20] Yang, Y., Fairbairn, C., Cohn, J., 2012. Detecting depression severity from vocal prosody. IEEE Trans. Affect. Comput. 4, 142–150.
[21] dr. Simon Lajos pszihiáter szakorvos személyes közlése alapján. Semmelweis Orvostudományi Egyetem Pszichiátriai és Pszichoterápiás Klinika
[22] Kiss, G., Sztahó, D., Vicsi, K. (2013, December). Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. In *2013 IEEE 4th international conference on cognitive infocommunications (CogInfoCom)* (pp. 579-582).
[23] IEEE.Boersma, Paul (2001). Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341-345.
[24] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
[25] IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.
[26] Hofmann, M., Klinkenberg, R. (Eds.). (2013). RapidMiner: Data mining use cases and business analytics applications. CRC Press.

[27] Beck, A.T., Steer, R.A., Ball, R., Ranieri, W.F., 1996. Comparison of beck depression inventories-ia and-ii in psychiatric outpatients. J. Pers. Assess. 67, 588–597.
[28] Suykens, J. A., Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, *9*(3), 293-300.
[29] Drucker, H., Burges, C. J., Kaufman, L., Smola, A. J., Vapnik, V. (1997). Support vector regression machines. In *Advances in neural information processing systems* (pp. 155-161).
[30] Németh Géza, Olaszy Gábor (szerk.) (2010) A magyar beszéd., Budapest: Akadémiai Kiadó.
[31] Hall, M. A. (1999). Correlation-based feature selection for machine learning.
[32] Yu, L., Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th international conference on machine learning (ICML-03)* (pp. 856-863).

10. A szerző tudományos közleményei

A tézispontokhoz kapcsolódó tudományos közlemények

Cikkek lektorált külföldi folyóiratban [12 p]

[J1] Kiss, G., Vicsi, K. (2017). Mono-and multi-lingual depression prediction based on speech processing. *International Journal of Speech Technology*, *20*(4), 919-935.
(BME-PA pontszám: 100% *6p = 6p.) Scopus / Web of Science

[J2] Kiss, G., Vicsi, K. (2015). Seasonal affective disorder speech detection on the base of acoustic phonetic speech parameters. *Acta Universitatis Sapientiae Electrical and Mechanical Engineering*, 7, 62-79.
(BME-PA pontszám: 100% *6p = 6p.)

Cikkek szerkesztett könyvben [1,5 p]

[B1] Kiss, G., Tulics, M. G., Sztahó, D., Esposito, A., Vicsi, K. (2016). Language independent detection possibilities of depression by speech. In *Recent Advances in Nonlinear Speech Processing* (pp. 103-114). Springer, Cham.
(BME-PA pontszám: 25% *6p = 1,5p.)

Konferenciacikkek [15,5 p]

[C1] Kiss, G., Módos, A., Vicsi, K. (2018) Comparison of the predicted depressed state of crew members with the results of their subjective psychological test at Concordia research station. *In Selected papers of 4th International Conference on Resarch, Technology and Educaion of Space* (p. 11).

(BME-PA pontszám: 50% *3p = 1,5p.)

[C2] Kiss, G., Vicsi, K. (2017). Follow-up Psychological Status Monitoring of the Crew Members of Concordia Research Station at Antarctica Based on their Speech. In *Selected papers of 3rd International Conference on Research, Technology and Education of Space* (p. 37).
(BME-PA pontszám: 100% *3p = 3p.)

[C3] Kiss, G., Vicsi, K. (2017). Investigation of cross-lingual depression prediction possibilities based on speech processing. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 97-102).
(BME-PA pontszám: 100% *3p = 3p.)

[C4] Kiss, G., Vicsi, K. (2017). Comparison of read and spontaneous speech in case of automatic detection of depression. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 213-218).
(BME-PA pontszám: 100% *3p = 3p.)

[C5] Kiss, G., Simon, L., Vicsi K. (2017). Depresszió súlyosságának becslése beszédjel alapján magyar nyelven. In: Vincze Veronika (szerk.) *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2017)*. pp. 125-135.
(BME-PA pontszám: 50% *1p = 0,5p.)

[C6] Kiss, G., Vicsi, K. (2014). Physiological and cognitive status monitoring on the base of acoustic-phonetic speech parameters. In *International Conference on Statistical Language and Speech Processing* (pp. 120-131). Springer, Cham.
(BME-PA pontszám: 100% *3p = 3p.)

[C7] Vicsi, K., Sztahó, D., Kiss, G. (2012). Examination of the sensitivity of acoustic-phonetic parameters of speech to depression. In *3rd International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 511-515). IEEE.
(BME-PA pontszám: 50% *3p = 1,5p.)

Egyéb a tézispontokhoz nem kapcsolódó tudományos közlemények

Cikkek lektorált külföldi folyóiratban [3 p]

[J3] Sztahó, D., Kiss, G., Vicsi, K. (2018). Computer based speech prosody teaching system. *Computer Speech Language, 50,* 126-140.
(BME-PA pontszám: 50% *6p = 3p.) Scopus / Web of Science

Cikkek lektorált hazai folyóiratban [3,5 p]

[J4] Sztahó, D., Kiss, G., Tulics, M. G., Czap, L., Vicsi K. (2014) Számítógéppel támogatott prozódiaoktató program. *Alkalmazott Nyelvészeti Közlemények* 9:(1) 144-153. (2014)
(BME-PA pontszám: 25% *2p = 0,5p.)

[J5] Kiss, G., Vicsi, K. (2013). Akusztikai hangosztályok felismerésén alapuló, nemlineáris idővetemítés megvalósítása a mondathanglejtés és a szóhangsúlyozás oktatásához. *Beszédkutatás* 21: 247-260.
(BME-PA pontszám: 100% *2p = 2p.)

[J6] Vicsi, K., Sztahó, D., Kiss, G. (2011). Nem verbális hangjelenségek spontán társalgásban. *Beszédkutatás* 19: 134-147.
(BME-PA pontszám: 50% *2p = 1p.)

Konferenciacikkek [7,58 p]

[C8] Sztahó, D., Kiss, G., Tulics, M. G., Vicsi, K. (2018). Automatic Separation of Various Disease Types by Correlation Structure of Time Shifted Speech Features. In *41st International Conference on Telecommunications and Signal Processing (TSP)* (pp. 1-4).
(BME-PA pontszám: 33% *3p = 1p.)

[C9] Tündik, M. Á., Kiss, G., Sztahó, D., Szaszák, Gy. (2017) Automatikus frázisdetektáló módszereken alapuló patológiás beszédelemzés magyar nyelven. *XIII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2017)* (pp. 113-124)
(BME-PA pontszám: 25% *1p = 0,25p.)

[C10] Tündik, M. Á., Kiss, G., Sztahó, D., Szaszák, Gy. (2017). Assessment of pathological speech prosody based on automatic stress detection and phrasing approaches. In *8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 67-72).

(BME-PA pontszám: 25% *3p = 0,75p.)

[C11] Kovács, A., Kiss, G., Vicsi, K., Winkler, I., Coath, M. (2015). Comparison of skewness-based salient event detector algorithms in speech. In *6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 285-290).
(BME-PA pontszám: 25% *3p = 0,75p.)

[C12] Sztahó, D., Kiss, G., Vicsi, K. (2015). Estimating the severity of Parkinson's disease from speech using linear regression and database partitioning. *In Proceedings of INTERSPEECH 201*5 (pp. 498-502)
(BME-PA pontszám: 50% *3p = 1,5p.)

[C13] Kiss, G., Sztahó, D., Vicsi, K., Golemis, A. (2014). Connection between body condition and speech parameters-especially in the case of hypoxia. In *5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 333-336).
(BME-PA pontszám: 33% *3p = 1p.)

[C14] Sztahó, D., Kiss, G., Czap, L., Vicsi, K. (2014). A computer-assisted prosody pronunciation teaching system. In *WOCCI 2014* (pp. 45-49).
(BME-PA pontszám: 33% *3p = 1p.)

[C15] Kiss, G., Sztahó, D., Vicsi, K. (2013). Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features. In *4th international conference on cognitive infocommunications (CogInfoCom)* (pp. 579-582).
(BME-PA pontszám: 33% *3p = 1p.)

[C16] Vicsi, K., Sztahó, D., Kiss, G., Czira, A. (2010) Spontán beszédben rejlő nem verbális hangjelenségek – érzelmek, hanggesztusok – vizsgálata. In: Tanács Attila, Vincze Veronika (szerk.) *VII. Magyar Számítógépes Nyelvészeti Konferencia MSZNY 2010*. pp. 249-260.
(BME-PA pontszám: 33% *1p = 0,33p.)