



Budapesti Műszaki és Gazdaságtudományi Egyetem

Villamosmérnöki és Informatikai Kar

Távközlési és Médiainformatikai Tanszék

Fodor Tamás

**DEPRESSZIÓS BESZÉD AKUSZTIKAI ÉS
FONETIKAI JELLEGZETESSÉGEINEK
VIZSGÁLATA**

KONZULENS

Dr. Vicsi Klára

BUDAPEST, 2013

Tartalomjegyzék

Összefoglaló.....	4
Abstract	5
1 Bevezetés	6
2 Irodalmi áttekintés	7
2.1 “Detecting depression – A comparison between spontaneous and read speech” [3].....	7
2.2 “Classification of depression state based on articulatory precision” [4].....	8
2.3 „Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR)” [6]	8
2.4 „Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk” [7]....	9
3 Adatbázis.....	11
3.1 Depressziós csoport adatbázisa	11
3.2 Kontroll csoport adatbázisa	12
3.3 Fonémaszintű szegmentálás	12
4 Jellemző kinyerés.....	13
4.1 Vizsgálandó akusztikai paraméterek	13
4.1.1 Szegmentális paraméterek.....	13
4.1.2 Prozódiai paraméterek.....	15
4.2 Akusztikai paraméterek mérése	16
5 Eredmények	18
5.1 Szegmentális paraméterek.....	18
5.1.1 Alapfrekvencia	18
5.1.2 Első formánsfrekvencia.....	19
5.1.3 Második formánsfrekvencia.....	19
5.1.4 Jitter.....	20
5.1.5 Shimmer	21
5.1.6 Mel-sávós energia	21
5.2 Szupraszegmentális paraméterek	23
5.2.1 Intenzitás szórás és tartomány.....	23
5.2.2 Alaphang szórás és tartomány.....	24
5.2.3 Szünethossz	25
5.2.4 Felolvasás időtartam.....	26

5.2.5 Szünetarány	26
5.2.6 Beszédtempó	27
5.2.7 Artikulációs tempó	27
5.3 Statisztikai analízis	28
5.3.1 A statisztikai próbák általános leírása	28
5.3.2 Normalitás vizsgálat.....	30
5.3.3 Szignifikancia vizsgálat.....	30
5.4 Eredmények értékelése.....	31
6 Konklúzió	35
7 Köszönetnyilvánítás	36
8 Irodalomjegyzék.....	37
9 Melléklet.....	39

Összefoglaló

Dolgozatomban egészséges és diagnosztizáltan depressziós személyek beszédakusztikai paramétereinek közti eltéréseket vizsgálok kötött szöveges felolvasáskor. A depresszió megfelelő időben történő diagnosztizálása emberéleteket menthet. Pszichiáterek szerint hallás alapján is eltérés mutatkozik a depressziós beszédben, mint akusztikai produktumban az egészséges emberek beszédéhez képest. Sokszor használják a depressziós beszédre a fakó, lomha, monoton és fémes jelzőket. Dolgozatom célja a beszéd azon akusztikai jellemzőinek detektálása, melyek az egészséges és a depressziós csoportokat hatékonyan elkülönítik.

Dolgozatomhoz felvételeket készítettem depressziós betegekkel - a Semmelweis Klinika Pszichiátriai osztályán -, továbbá egészséges személyekkel, ügyelve arra, hogy a koreloszlás mindkét csoportban egyforma legyen. A résztvevők az „Északi szél és a Nap” c. fonetikailag kiegyensúlyozott mesét olvasták fel. Minden résztvevő személyt a neurológus orvos besorolt egy szabványosított súlyossági skála (BDI skála) megfelelő fokára. A felvételeket fonémaszinten címkéztem, szegmentáltam. A vizsgált jellemzőket Praat program segítségével nyertem ki. .

Az elemzéskor szegmentális és supra-szegmentális jellemzőket is használtam. A szegmentális paramétereket csak 'E' hangoknál vizsgáltam, melyek: alapfrekvencia, első és második formánsfrekvencia, jitter, shimmer és mel-sávok energiája. A supra-szegmentális vizsgált paraméterek a következők voltak: beszédhangerő dinamika, beszéddallam dinamika, teljes szünethossz, felvétel időtartama, teljes szünet aránya a felvétel időtartamához képest, beszédtempó és artikulációs tempó.

A két csoport akusztikai jellemzőinek elemzését statisztikai analízissel végeztem. A statisztikai elemzéshez Kolmogorov-Szmirnov és kétmintás t-próbát használtam.

A vizsgált szegmentális és supra-szegmentális paraméterek közül szinte mindegyik szignifikáns különbséget mutatott a két vizsgált csoport között. Az eredmények bizakodásra adnak okot, és további kutatásra ösztönöznek.

Abstract

I examine in my paper the differences between acoustics of speech parameters of healthy persons and persons diagnosed with depression, in case of reading a fixed text. Diagnosis of depression in a proper time may even help lives. According to the opinion of psychiatrists, a difference can occur in depressed speech on the basis of hearing, too, as in an acoustic product. Depressed speech often is described by psychiatrists as dull, slow and monotonous. Purpose of my paper is to detect acoustic features of speech, which can effectively separate healthy and depressed groups from each other.

I made records for my paper with patients suffering from depression at Department of Psychiatry of Semmelweis Clinic and with healthy persons keeping the age distributions identical. The participants read a phonetically balanced tale aloud, called “The North Wind and the Sun”. All participants were ranked by a neurologist using a standardized severity scale called BDI. I labeled and segmented the recordings at a phoneme level. The selected parameters were obtained by using Praat program.

I used segmental and supra-segmental parameters in the analysis. The segmental parameters such as pitch, first and second formant frequency, jitter, shimmer and mel-filter energies were only examined at the phoneme “E”. The supra-segmental parameters were the following: intensity and pitch fluctuation, full pause length, full duration of record, ratio of full pause compared to duration of record, speech rate and articulation rate.

The statistical analysis of acoustic features was carried out using Kolmogorov-Smirnov test and two-patterned T-test.

Nearly all of the segmental and supra-segmental parameters showed significant differences between the examined groups. As the results are promising, they incite me for further research.

1 Bevezetés

A WHO (World Health Organization) Depression – A Global Public Health Concern című kiadványában 350 millióra becsülte a depresszióban szenvedő betegek számát 2012-ben [1]. A WHO előrejelzései szerint 2030-ra az unipoláris depresszió az első három legsúlyosabb betegség között lesz világviszonylatban a HIV/AIDS vírus és a szívbetegségek mellett [2]. Annak ellenére, hogy a betegségben szenvedők száma igen magas, a diagnózis felállítása még mindig egy kisszámú képzett szakorvos rétegre hárul, ráadásul a használt kérdőíves tesztek is sok tekintetben szubjektívek. Szükség lenne egy olyan objektív, robosztus értékelő rendszer kialakítására, amelyet akár házi orvosok is használhatnának a depresszió felismerésére és követésére. Orvosok gyakran használják a depressziós beszédre a fakó, lomha, monoton, élettelen és fémes jelzőket. Ezek az érzeti tulajdonságok összekapcsolhatók egyes akusztikai jellemzőkkel, ide értve az alaphérvenciát, formáns hérvenciákat, a beszéd alatt tartott szünetek hosszát, továbbá a beszédtempót. Az elmúlt évek tapasztalatai szerint a beszéd bizonyos paraméterei összefüggésben állnak a depresszió súlyával. Az említett kutatások két részre oszthatóak, egyrészt depressziós és egészséges csoportok hangmintái közötti eltéréseket kerestek, másrészt követéses vizsgálatokkal, depressziós személyek kezelésre adott reakciójuk ismeretében értelmezték a személyen belüli változásokat.

A dolgozat második fejezetében ismertetem a témával kapcsolatos aktuális irodalmakat. Ezt követően a harmadik fejezetben mutatom be a dolgozathoz használt hangadatbázist. A negyedik fejezetben kitérek a kiválasztott akusztikai paraméterekre, és röviden bemutatom a jellemzők kinyeréséhez használt Praat programot. A dolgozat ötödik fejezetében írok az egészséges és depressziós csoportok között tapasztalt eltérésekről, és összefoglalom a statisztikai elemzés eredményeit. A hatodik fejezetben az általam levont következtetéseket, és a további előrelépési lehetőségeket ismertetem.

2 Irodalmi áttekintés

A depresszió hatására megváltozott beszéd már régóta fontos kutatási terület, és több évtizede jelennek meg cikkek a témában. Ebben a fejezetben négy tanulmány kerül bemutatásra, melyek az utóbbi évek fontosabb eredményeit összegzik. A kutatásokban számos akusztikai paramétert azonosítottak, melyek alkalmasak lehetnek a depressziós beszéd megkülönböztetésére. A magyar nyelv sajátosságai miatt azonban fontos egy magyar depressziós hangadatbázis kiépítése és vizsgálata.

2.1 “Detecting depression – A comparison between spontaneous and read speech” [3]

A kutatásban depressziósnak diagnosztizált és egészséges személyek spontán és olvasott beszédét hasonlították össze. A felolvasások 20 pozitív és 20 negatív érzelmű mondatot tartalmaztak, a spontán szakaszban 8 különböző érzelmet kiváltó kérdésre vártak választ. Spontán beszéd alatt nagyobb eltérésre számítottak, mint a felolvasás alatt, illetve feltették, hogy a depressziós és egészséges csoport több akusztikai jellemző mentén is eltérést mutat.

A használt adatbázis 30 depressziós és 30 egészséges férfi és női felvételt tartalmaz fele-fele arányban nemeként. A depressziós személyeket HAM-D skála segítségével, míg a kontroll csoportban résztvevő személyeket a depressziós csoport koreloszlásának figyelembevételével választották ki. A résztvevő személyek 21-75 évesek voltak.

A vizsgált akusztikai jellemzők a következők voltak: alapfrekvencia, alapfrekvencia-ingadozás, MFC együtthatók, intenzitás, első három formánsfrekvencia, jitter, shimmer és HNR (Harmonic to Noise Ratio). A jellemzők kinyeréséhez az openSMILE programot használták, az osztályozáshoz pedig SVM-et (Support Vector Machine).

Spontán beszéd során nagyobb különbségeket tapasztaltak, mint a kötött szöveges felolvasások alatt, és sikeresen azonosítottak olyan akusztikai paramétereket, melyek magas differenciáló erővel bírnak az osztályok elkülönítésére. A legjobbnak bizonyuló paraméterek a következők voltak: jitter, shimmer, intenzitás és MFC együtthatók.

2.2 “Classification of depression state based on articulatory precision” [4]

A kísérleti személyek mind tizennyolcadik életévüket betöltött, gyógyszeres és pszichoterápiás kezelésben részesülő diagnosztizáltan depressziós személyek voltak – 20 nő, 15 férfi. Harmincöt pszichológus által ajánlott páciens depressziókezelését indították el a kutatással párhuzamosan, és a kezelésre adott reakciójuk ismeretében elemezték a beszédakusztikájuk változását. Az anyagokat a résztvevőktől egy automata IVR (Interactive Voice Response) telefonos interfészen keresztül gyűjtötték. A hangfelvételeket 8 biten kvantálták, és 8 kHz-en mintavételezték. A hangelemzéshez Praat programot használtak [5]. A betegek telefonon keresztül, heti rendszerességgel elvégezték a HAM-D (Hamilton Depression Rating Scale) és QIDS (Quick Inventory of Depressive Symptomatology) teszteket.

Folyamatos beszéden és kitartott hangokon mért formánsfrekvenciák segítségével próbálták elkülöníteni a depressziós és egészséges csoportokat. A formánsok karakterisztikájának és dinamikájának leírásához mindenegyed formánsra 9 jellemzőt számoltak 20 ms-os keretekkel és 10 ms-os időléptékekkel. Mind a három formánst három részre osztották – nyers, aluláteresztő és felüláteresztő szűrővel szűrt -, és ezeknek külön-külön számolták a formánsfrekvenciát, az első és második deriváltat. A deriváltakat a megelőző és következő keretek eltéréséből számolták.

Az osztályozáshoz SVM-et és GMM-et (Gaussian Mixture Model) is használtak. Az SVM egy pár százalékkal jobban teljesített, de mindkét módszerrel 70-80%-os hatékonyságú volt az osztályozás.

2.3 „Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR)” [6]

A tanulmányban kezelés alatt álló depressziós személyek gyógyulása, és a beszédakusztikai jellemzőik megváltozása között kerestek összefüggéseket. A használt hangadatbázisok megegyeznek a 2.2. pontban bemutatottal. Számos paramétert találtak melyek változtak a betegek javulásával. Azon betegeknél, akik reagáltak a kezelésre lényegesen nagyobb volt az alaphang ingadozása, csökkent a szünetek hossza beszéd közben, továbbá gyorsult a beszédtempójuk. Azoknál a pácienseknél, akiknél nem mutatkozott javulás, ezen jellemzők változása sem volt megfigyelhető.

A vizsgált akusztikai jellemzők a következők voltak: alapfrekvencia, első és második formánsfrekvencia, teljes felvételi időtartam, teljes szünetidő, szünetek száma és beszédtempó.

Ahogy az várható volt, a kutatás előrehaladtával a depresszió súlyossága a legtöbb esetben csökkent. Az első formánsfrekvencia ingadozása nem bizonyult statisztikailag alkalmasnak a depresszió súlyának meghatározására, azonban az alaphang és a második formáns ingadozása már igen. A felvétel időtartama, a teljes szünet ideje, a szünetek ingadozása és a beszédsebesség is jelentős korrelációt mutatott a HAM-D értékekkel.

2.4 „Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk” [7]

Pszichiáterek előzetesen a betegség súlyának függvényében egészséges, depressziós illetve magas kockázatú öngyilkos hajlamú csoportokba sorolták a betegeket. Feltételezték, hogy a kezelés hatására a betegeknél javulás lesz megfigyelhető, emiatt a javulás előtti beszédet is vizsgálták. A nők esetében a felvételeket egy készülékkel, míg a férfiak esetében több módszerrel és több helyszínen készítették. A digitalizálás után a háttérzajokat is kiszűrték, minden jelet 16 biten kvantáltak, a mintavételi frekvenciát 10 kHz-nek választották, továbbá egy 5 kHz-es aluláteresztő szűrőn is átengedték a jelet. Ebben a kutatásban a fél másodpercnél hosszabb szüneteket eltávolították a folyamatos beszéd hatásának eléréseért.

A felvételeken a következő paramétereket vizsgálták:

- Alapfrekvencia: Kepsztrális hangmagasság határoló algoritmust használtak a beszélt és néma részek megkülönböztetéséért.
- Intenzitás: A négyzetes átlagoló algoritmust MATLAB-ban [8] implementálták.
- Formánsfrekvencia és sávszélesség: MATLAB-ban megírt lineáris predikciós kódolást (LPC) használtak az első három formáns átlagos frekvenciájának és sávszélességének meghatározásához.
- Spektrumsűrűség: Az energia-eloszlást 0 Hz és 2000 Hz közötti frekvenciatartományban számították Welch-módszerrel és nem átlapolódó Hamming ablakkal MATLAB segítségével.

A női beszélők alapfrekvencia és amplitúdó vizsgálata hatástalan volt, míg a formánsok és a spektrumsűrűség alkalmasnak bizonyult a kontroll csoport és a depressziós csoport megkülönböztetésére. Súlyosan depressziós és öngyilkos hajlamú csoportokban a férfiaknál

csökkent alapfrekvencia tartomány, megnövekedett formánsfrekvenciák és harmadik formáns sávszélesség volt megfigyelhető. A férfi kontroll csoporttól a betegek hatékonyan megkülönböztethetőek bizonyultak a formánsfrekvenciák és a spektrumsűrűségi jellemzők segítségével. A depressziós és szuicid beszédnél általános mintaként jelent meg a formánsfrekvenciák és az első formáns sávszélességének emelkedése, és a simult spektrum..

3 Adatbázis

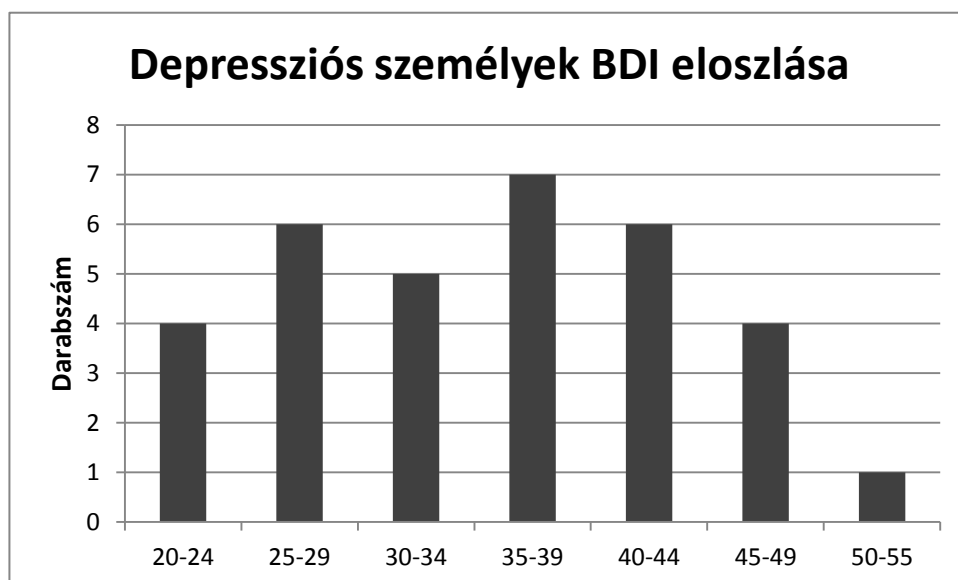
3.1 Depressziós csoport adatbázisa

A kísérletben résztvevő személyeket a Semmelweis Pszichiátriai és Pszichoterápiás Klinikáról választottuk ki Dr. Simon Lajos, a Pszichiátriai Klinika docensének segítségével. A kiválasztott depressziós betegek tizennyolcadik életévüket betöltötték, és írásban is hozzájárultak a felvételek elkészítéséhez és elemzéséhez. Az utóbbi években összesen 46 felvétel készült, ám ezekből csupán a fent említett 33 bizonyult alkalmasnak – 18 nő és 15 férfi. A kihagyott felvételek során egyedi problémák léptek fel, úgymint olvasási nehézség, látáskorlátozottság illetve volt olyan páciens is, aki többször elsírta magát a felvétel alatt.

A betegek osztályozására a BDI (Beck Depression Inventory) [9] skálát használtam, mely 0-63 érték között, a beteg által szubjektíven értékeli a depresszió súlyát.

A BDI index szerint az alábbi kategóriákba sorolhatóak a betegek [9]:

- 0-13: minimális depresszió,
- 14-19: enyhe depresszió,
- 20-28: mérsékelt depresszió,
- 29-63: súlyos depresszió.



3.1. ábra: BDI értékek eloszlása

A depressziós páciensek BDI indexük szerinti eloszlása látható a 3.1. ábrán. A felvételek számának növelésével akár a depresszió súlya és a vizsgált akusztikai paraméterek közti korreláció is vizsgálható. A résztvevő betegek a BDI skálán a mérsékelt és a súlyosan depressziós kategóriákba tartoznak. A depressziós csoportot a későbbiekben minimális illetve enyhe depresszióban szenvedő betegekkel is szükséges bővíteni, hisz a depresszió felismerése ezekben a szakaszokban a legnehezebb.

A felvételeket különálló helyiségben rögzítettem. A kötetlen szakaszban sokszor a kezelőorvosnak is közbe kellett szólni, mert megakadtak a páciensek, illetve többször előfordult, hogy elfelejtették, hogy miről beszéltek. A kötetlen beszélgetés után egy kötött szöveg felolvasása következett. Ebben a szakaszban *Az Északi Szél és a Nap* című fonetikailag kiegyensúlyozott mesét olvasták fel a betegek. A felvételek Audio-Technica ATR335 típusú mikrofonnal és Cool Edit 2000 programmal lettek rögzítve. A páciens és a kérdező orvos is külön mikrofonba beszélt, így külön, tiszta hangsávokat nyertem. A felvételek 44,1 kHz-en, 16 bites kvantálással kerültek rögzítésre.

A depressziós felvételek közül csak a kötött szöveges mese felolvasásokat vizsgáltam, mellyel számos prozódiai paraméter kerülhetett be a vizsgálandó jellemzők közé, melyek tanulmányozására kötetlen szöveg elemzése esetén nem lett volna lehetőség.

3.2 Kontroll csoport adatbázisa

Az egészséges személyek kiválasztásakor a depressziós csoport koreloszlását vettem figyelembe, így a két csoport közti korból adódó lehetséges eltéréseket kiküszöböltem. A kontroll csoportban 18 női és 18 férfi hangfelvétel található. A felvételek csupán kötött szöveges részt tartalmaznak, melynek során a depressziósoknál ismertetett *Északi Szél és a Nap* című mesét olvasták fel. Az így készült felvételeket fonémaszinten szegmentáltam. A felvételek kvantálása, mintavételezése és mikrofonozása a depressziós csoportéval megegyezik.

3.3 Fonémaszintű szegmentálás

A hangminták szegmentálásakor a tanszéken fejlesztett automataszegmentáló program volt segítségemre. Ehhez a felvételeket hangzás alapján, fonémák szintjén lejegyeztem, majd a hangfájl és az így nyert szövegfájl segítségével az automataszegmentáló egy címkefájlt hozott létre. A *.TextGrid* kiterjesztésű annotációs fájl használja a Praat program a szegmentáláshoz és címkézéshez. Egy címkefájlban az időintervallumok vannak felcímkézve SAMPA [10] karakterekkel.

4 Jellemző kinyerés

4.1 Vizsgálandó akusztikai paraméterek

A beszéd szegmentális szerkezetének vizsgálatakor a beszédhang vagy a beszédhangnál kisebb egységek akusztikai jellemzőinek elemzésére kerül sor, míg a szupraszegmentális, vagy prozódiai szerkezet tárgyalásakor a beszéd több fonémán átnyúló akusztikai jellemzőinek elemzése történik [11]. A korábbi kutatásokban ismertetett eredmények, az előző félévek során nyert tapasztalataim és a vizsgálandó adatbázisok tulajdonságait figyelembe véve, a következő pontokban bemutatott akusztikai jellemzőket választottam ki elemzésre.

4.1.1 Szegmentális paraméterek

Az 'E' beszédhangoknál mért alaphang és formánsfrekvencia értékeket a hang középpontjában mértem, a jitter és shimmer értékeket 60 ms-nál hosszabb 'E' fonémáknál a középponttól ± 20 ms-os tartományban. A felolvasott mesében legtöbbször előforduló beszédhang az 'E', emiatt esett rá a választás. Az alaphangfrekvencia, első és második formánsfrekvencia meghatározásakor a mérési hibák elkerüléséért a felső és alsó egy százalékát elhagytam a sorba rendezett értékeknek. A mérési hibák kiszűrése miatt a jitter maximális értékét 10 százalékának, a shimmer maximálisan megengedett értékét 20 százalékának választottam.

- Alaphangfrekvencia [Hz]

A hangmagasságra jellemző fizikai mennyiség a beszéd alaphangfrekvenciája, amely a folyamatos beszéd alaphangját jelenti [12]. Az alaphang személyenként változó egyedi fiziológiai jellemző, melynek magassága a személy magasságától, a hangszalag feszességétől függ, emiatt nem gondoltam, hogy a két csoport alaphangfrekvenciája különbözne. A beszédhang periodicitása egy kereszt-korrelációs algoritmussal lett kiszámolva [13].

- Első és második formánsfrekvencia [Hz]

A magánhangzókat egyértelműen meghatározzák az első két formánsuk [14]. Elviekben akár négy, öt formánsfrekvenciát is ki lehetne mérni ám a gyakorlat azt mutatja, hogy a második formáns utáni formánsfrekvenciák mérésekor már túl nagy a hibaszázalék. A formánsok meghatározásához 10 ms-os időléptéket és 25 ms-os ablakméretet választottam.

Gauss ablakozást követően Burg algoritmussal [15] számoltam az LPC együtthatókat mindenegybes ablakkeretre.

- Jitter [%]

A jitterrel a fonémán belüli alapfrekvencia bizonytalanságát határozhatjuk meg. A következő algoritmust használtam a paraméter meghatározásához:

$$Jitter(másodperc) = \frac{\sum_{i=2}^N |T_i - T_{i-1}|}{N - 1}$$

$$\text{Átlagos periódus}(másodperc) = \frac{\sum_{i=1}^N T_i}{N}$$

$$Jitter = \frac{Jitter(másodperc)}{\text{Átlagos periódus}(másodperc)}$$

- Shimmer [%]

A jitterhez hasonlóan a shimmer is a hangképzés bizonytalanságát mutatja meg, ám a jitter az alapfrekvenciában, míg a shimmer az amplitúdóban történő változásokat méri. A használt algoritmus a jitter mérésekor alkalmazottal megegyező, csupán a periódusok változása helyett az egymást követő amplitúdók közti átlagos eltérés és az átlagos amplitúdó hányadosa adja a shimmer értéket.

- Mel-sávós energia

Az mel-sávós energia kiszámítása egy lényegkiemelési eljárás, melynek alapja az emberi hallásmechanizmusra épít. A mel-sávós energia értékeket úgy nyertem, hogy a beszédjelet Fourier-transzformáltam majd a szűrősoros elemzést elvégeztem - tehát az összetevőket mel sávok szerint összegeztem. Ennek eredményeként a 24 sávszűrő kimenetén egy-egy számszerű érték jelent meg, amely az adott kritikussávba eső intenzitások összege. Az értékeket átszámoltam W/m^2 -be, majd a sávok intenzitásértékeit normalizáltam, úgy hogy a 24 intenzitásérték összege egy legyen, majd visszaszámoltam a mel-sávós energiákat dB mértékegységbe. Így a felvételek közti eljárásbeli különbségek, mint mondjuk a mikrofon távolsága a beszélőtől elhanyagolható. Végül a sávok egymáshoz viszonyított arányát vizsgáltam.

4.1.2 Prozódiai paraméterek

- Intenzitás szórás és tartomány [dB]

Az intenzitás ingadozása fontos prozódiai paraméter, mellyel a beszéd színesebbé és dinamikusabbá tehető. Feltettem, hogy a pszichológusok által, és általam is sokszor monotonnak, élettelennek tűnő depressziós beszéd eltérhet a normál beszédétől e paraméter szerint. Az intenzitás kiszámításához a program megméri a hangnyomást és ebből RMS-t (Root Mean Square) számol meghatározott időablakra [5]. Az intenzitás ingadozását kétféleképpen mértem, egyrészt vizsgáltam a felvétel alatti intenzitások szórását, másrészt a legalacsonyabb és legmagasabb intenzitás közti különbséget, tehát az intenzitás tartományt.

- Alaphang szórás és tartomány [Hz]

Az előző paraméterrel megegyezően szórást és tartományt vizsgáltam. Minél nagyobb intervallumban vesz fel értéket a beszélő alaphangfrekvenciája, annál színesebbnek tűnik a hallgatónak. Ezt a paramétert is úgy ítélem meg, hogy magas osztályozóerővel bírhat a depressziósok és kontroll csoport tagjainak elkülönítésére.

- Szünethossz [másodperc]

A korábban bemutatott kutatások, egyik fontos szupraszegmentális paramétere volt. Előzetes feltevésem szerint a depressziós páciensek több szünetet tartanak olvasáskor, a felvételek alatt sokszor a koncentráció is nehezebbre esett, és sokszor zavarodtak bele a felolvasásba.

- Felolvasás időtartam [másodperc]

A mese felolvasásához szükséges időben is eltérésnek kell mutatkoznia, amennyiben a szünetek hossza megnövekszik, és a betegek beszéde is lassabb az egészséges személyekénél.

- Szünetarány [%]

A szünetarány megmutatja, hogy a teljes felolvasás során hány százalékban tartott szünetet az olvasó. Feltevésem szerint ez a jellemző is eltérést fog mutatni két megfigyelt csoport között. A teljes szünet arányát a felvétel időtartamához képest a következő képlettel számoltam:

$$\text{Szünetarány} = \frac{\text{Szünethossz}}{\text{Felolvasás időtartam}}$$

- Beszédtempó [fonéma/másodperc]

A beszédtempó, az időegységre jutó nyelvi jelek száma, függetlenül attól, hogy a közlésben volt-e szünet vagy más megakadásjelenség. A beszéd tempóját szokásosan fonéma/másodpercben adjuk meg. A mai magyar beszéd átlagos tempóindexe köznyelvben 12 fonéma/másodperc, amelybe beletartozik minden, nem a beszédképzésre fordított időt is [16]. A fent ismertetett tanulmányok fontos paramétere, mely a szünetek hosszával is összefüggésben áll, így a depressziós betegeknél alacsonyabb beszédtempót figyeltek meg.

A beszédtempót a következő képlettel számoltam:

$$\text{Beszédtempó} = \frac{\text{Fonémaszám}}{\text{Felolvasás időtartam}}$$

- Artikulációs tempó [fonéma/másodperc]

Az artikulációs tempó az artikuláció tiszta idejére eső nyelvi jelek számát jelenti, vagyis ekkor nem vesszük figyelembe a szüneteket és megakadásjelenségeket [16]. A beszédtempóhoz képest fontos az az eltérés, hogy a szünetek várhatóan megnövekedett időtartama nem játszik szerepet az artikulációs tempó kiszámításánál. Az artikulációs tempót a következő képlettel számoltam:

$$\text{Artikulációs tempó} = \frac{\text{Fonémaszám}}{\text{Felolvasás időtartam} - \text{Szünehossz}}$$

4.2 Akusztikai paraméterek mérése

A Praat programban saját szkriptnyelvén is, PSL-ben (Praat Script Language) fejlesztettem. A PSL az objektumorientált nyelvekhez képest számos eltérést mutat, ami sokszor megnehezíti a fejlesztést, ám az ismert ciklusok (for, while), és feltételek (if, else) használhatóak.

A szükséges adatok kinyeréséhez a program beolvassa a hangfelvételt és az annotációs fájlokat, majd vagy megadott időkeretenként, vagy fonémánként végiglépdel a teljes felvételen, és elvégzi a szükséges műveletet.

A következő kódrészlet az 'E' hang első és második formánsfrekvenciáit számolja ki, és menti el a kimeneti szövegfájlba:


```
if label$="E"  
  start = Get starting point... tier interval  
  end = Get end point... tier interval  
  midpoint = (start + end) / 2  
  
  select Formant 'soundname$'  
  f1 = Get value at time... 1 midpoint Hertz Linear  
  f2 = Get value at time... 2 midpoint Hertz Linear  
  
  resultline$ = "'soundname$'      'f1' 'f2' 'newline$'"  
  fileappend "'resultfile$'" 'resultline$'  
  select TextGrid 'soundname$'  
endif
```

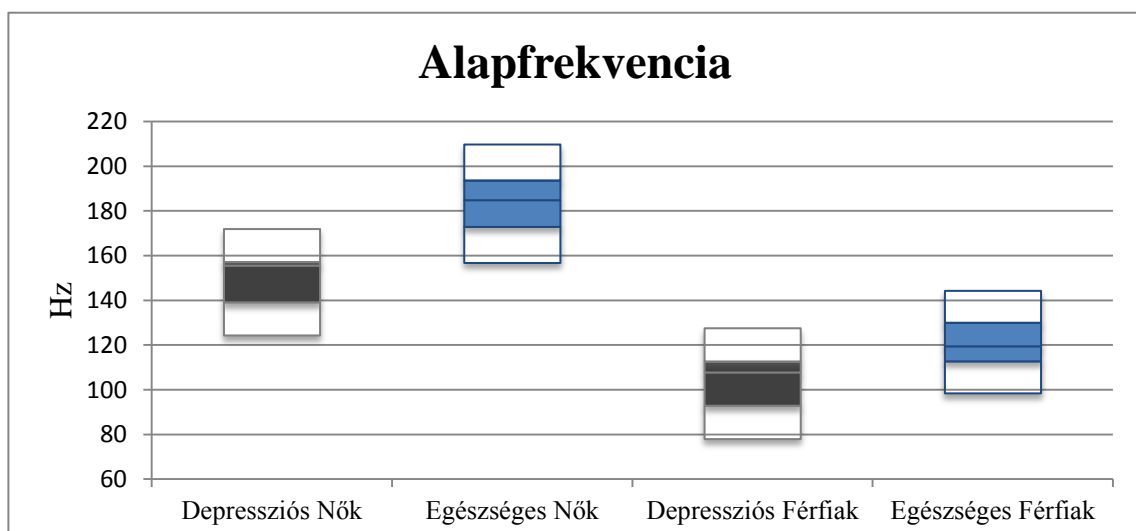
5 Eredmények

Az előzetesen kijelölt akusztikai paraméterek jelentős része szignifikáns eltérést mutatott a két tanulmányozott csoport között. A jellemzők közül az alapfrekvencia, az első formánsfrekvencia, a beszéd és az artikulációs tempó alacsonyabbnak, a szünetarány, felvétel időtartama és a szünethossz magasabbnak bizonyult a depressziós csoportban, mindkét nem esetén. A mel-sávós energiaértékek is eltérőek a csoportok között, ám az intenzitás dinamika a várakozással ellentétben nem tűnik hatékony paraméternek a csoportok elkülönítésére ezekkel a hangadatbázisokkal.

5.1 Szegmentális paraméterek

5.1.1 Alapfrekvencia

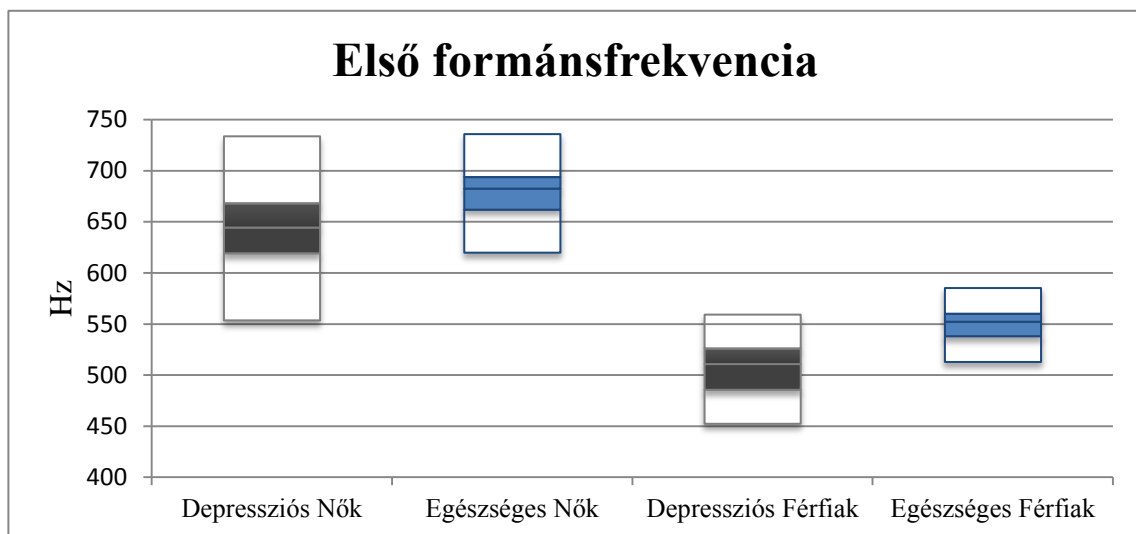
Az átlagos alapfrekvencia férfiaknál 100-200 Hz, nőknél 150-300 Hz [12]. A kontroll, és depressziós csoport értékei is a fenti tartományokba esnek, ám a két csoport átlagos alapfrekvenciája eltérő. Annak ellenére, hogy az előfeltevés szerint az alaphang frekvenciája egyes személyekre jellemző biológiai adottságoktól függ, nem elhanyagolható az osztályok közötti különbség – 5.1. ábra. Az alaphang csökkenését depressziós személyeknél az életkorbeli eltérés nem eredményezhette, mivel a csoportok koreloszlása megegyező.



5.1. ábra Alapfrekvencia

5.1.2 Első formánsfrekvencia

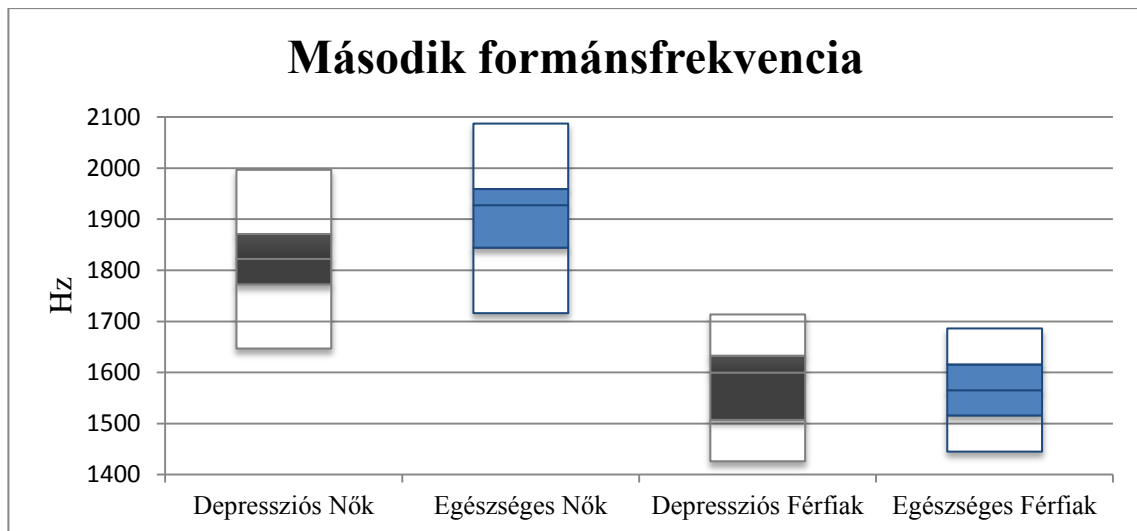
Magyar nyelvben az 'E' beszédhang első formánsfrekvenciája férfiaknál 450-650 Hz, nőknél 550-800 Hz [17] között vesz fel értékeket átlagosan. A két vizsgált csoport értékei is ebbe a tartományba esnek, ám a depressziós személyek első formánsfrekvenciái alacsonyabbak, mint az egészséges felolvasók formánsfrekvenciái, mindkét nem esetén – 5.2. ábra. A férfiaknál nagyobb az eltérés a két osztály között, és a depressziós betegek szórása is nagyobb, mint az egészséges társaiké.



5.2. ábra Első formánsfrekvencia

5.1.3 Második formánsfrekvencia

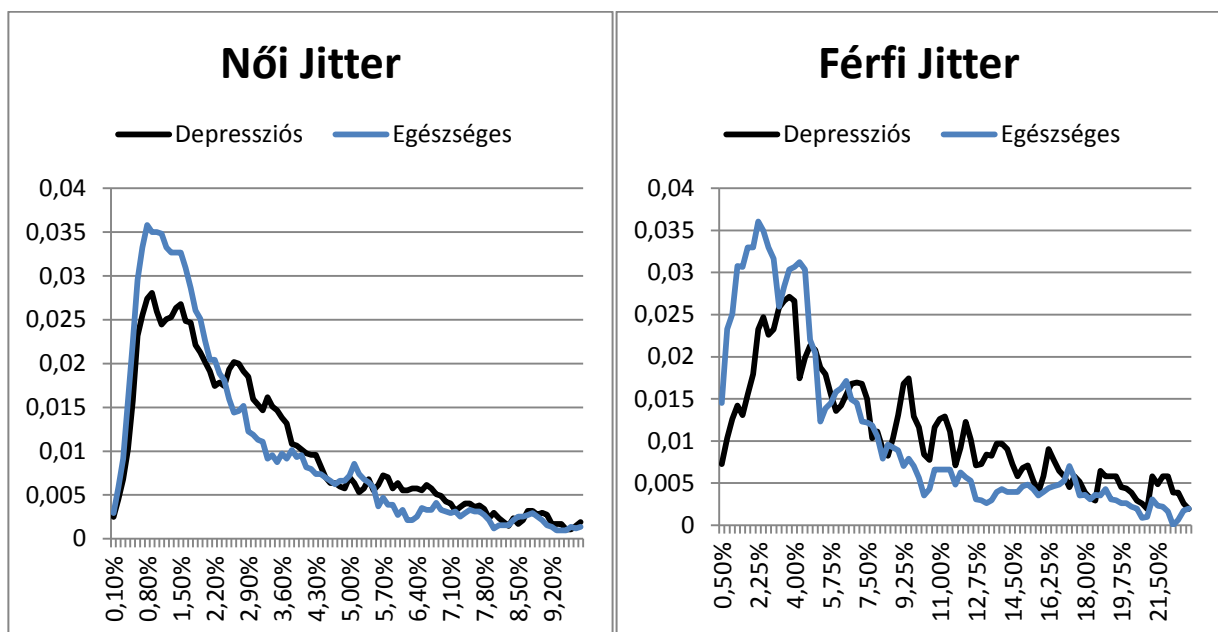
A magyar beszédben az 'E' beszédhang második formánsfrekvenciája nőknél átlagosan 1700-2300 Hz, férfiaknál 1500-2000 Hz tartományokba esnek [17]. A két csoport értékei ezekbe a tartományokba esnek, a férfiak és a depressziós nők formánsfrekvenciái kicsit alacsonyabbak az elvártnál. Az alaphangnál és az első formánsfrekvenciánál látott eltérés itt csupán a nőknél tapasztalható, férfiaknál az értékek egy tartományba esnek, sőt a depressziós férfiaknál a medián alacsonyabb is, mint az egészséges férfiaké.



5.3. ábra Második formánsfrekvencia

5.1.4 Jitter

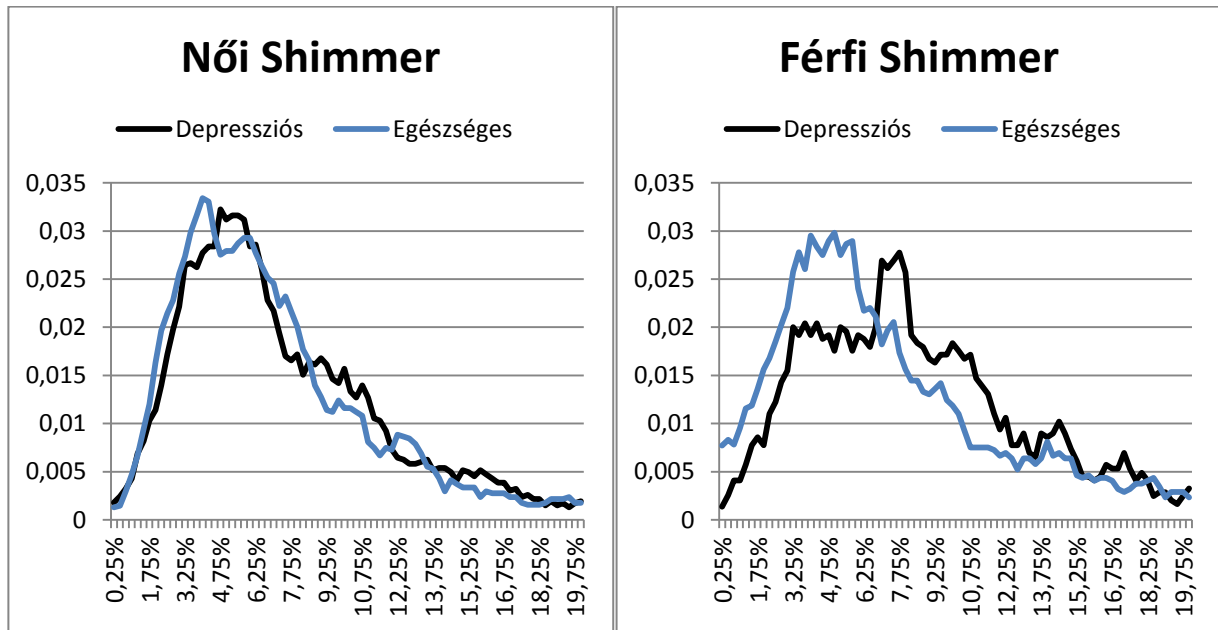
Az alapfrekvencia bizonytalanságát mutató paraméter. Férfiaknál és nőknél is nagyobb arányban vesznek fel alacsony értéket az egészséges személyek – 5.4. ábra –, tehát a depressziósok 'E' beszédhangján belül több az alapfrekvencia ingadozás. Azonban az ábrán az is jól látszik, hogy mindkét csoporthoz tartoznak magasabb jitter értékek is.



5.4. ábra Jitter sűrűségfüggvények

5.1.5 Shimmer

Az 'E' hangon belüli amplitúdó bizonytalanságát mutatja meg. A női shimmer sűrűségfüggvények szinte teljesen egybeesnek, míg férfiaknál az egészséges csoport alacsonyabb értékekkel bír – 5.5. ábra. A női shimmer szórása alacsonyabb, a férfiak görbéje kisebb mértékben változik a magasabb shimmer értékek felé. A jitterhez hasonlóan a shimmer sem tűnik alkalmasnak a két csoport elkülönítésére.



5.5. ábra Shimmer sűrűségfüggvények

5.1.6 Mel-sávos energia

A sávok dB-ben nyert intenzitásértékeit, az alábbi módon számoltam át W/m^2 mértékegységbe:

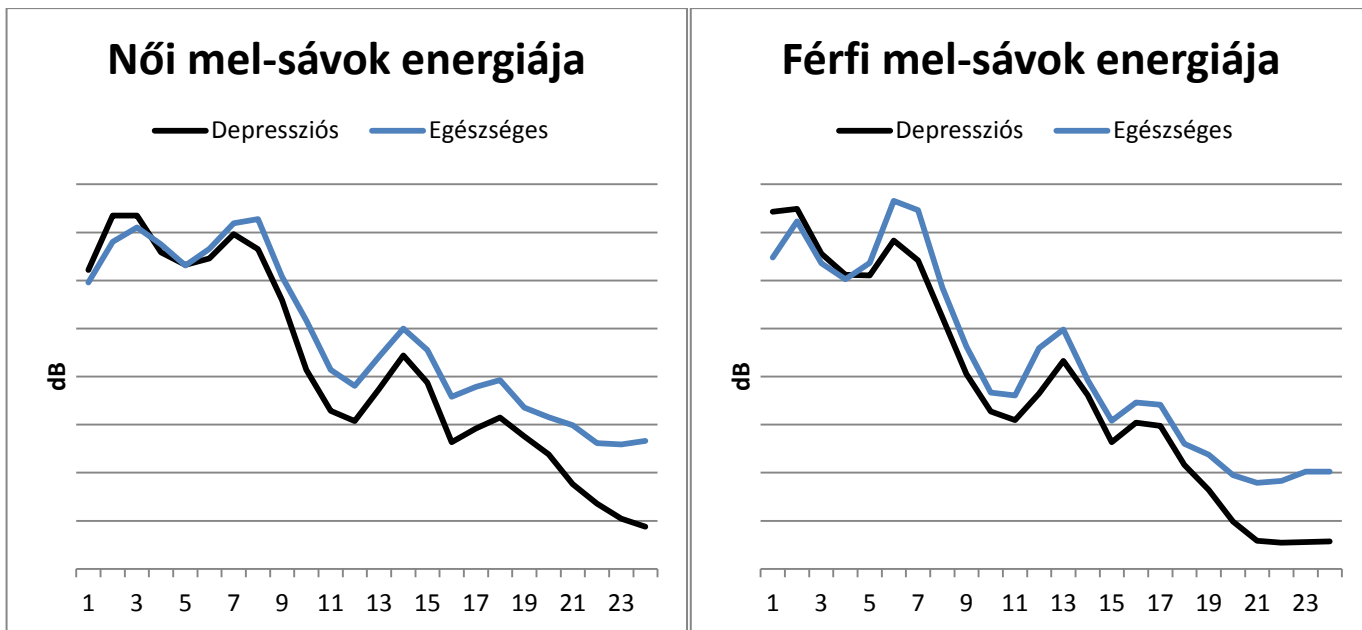
$$I[dB] = 10 \log_{10} \frac{I}{I_0} = 10 \log_{10} I - 10 \log_{10} I_0$$

,ahol a viszonyítási alap a hallásküszöb intenzitás értéke vagyis: $I_0 = 10^{-12} W/m^2$, tehát

$$10 \log_{10} I = mért_{dB} + 10 \log_{10} I_0 = I[dB] - 120$$

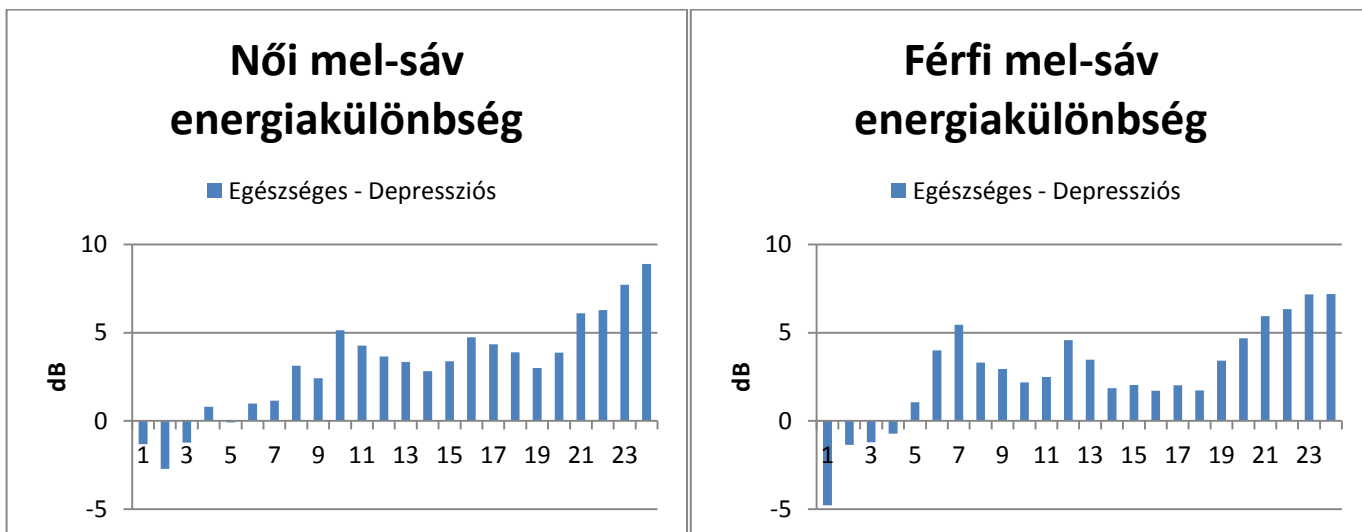
$$I[W/m^2] = 10^{(I[dB]-120)/10}$$

Az így számított intenzitás értékeket elosztottam személyenként a 24 sáv összegzett intenzitásával, majd visszszámoltam az értékeket dB-be. Az intenzitás értékek sávonkénti eltéréseit vizsgáltam a depressziós és egészséges csoportokon – 5.6. ábra.



5.6. ábra Mel-sávok energia aránya

Az alacsonyabb sávokban a depressziós csoport emelkedett intenzitás értéket mutat az egészséges beszélőkhöz képest, míg a magasabb sávok felé haladva az egészséges intenzitások növekednek. A női beszélők között a legnagyobb eltérés a 10., 16., a férfi csoportban a 7. és 12., és mindkét nem esetén a magasabb sávokban látható az 5.7.ábrán.

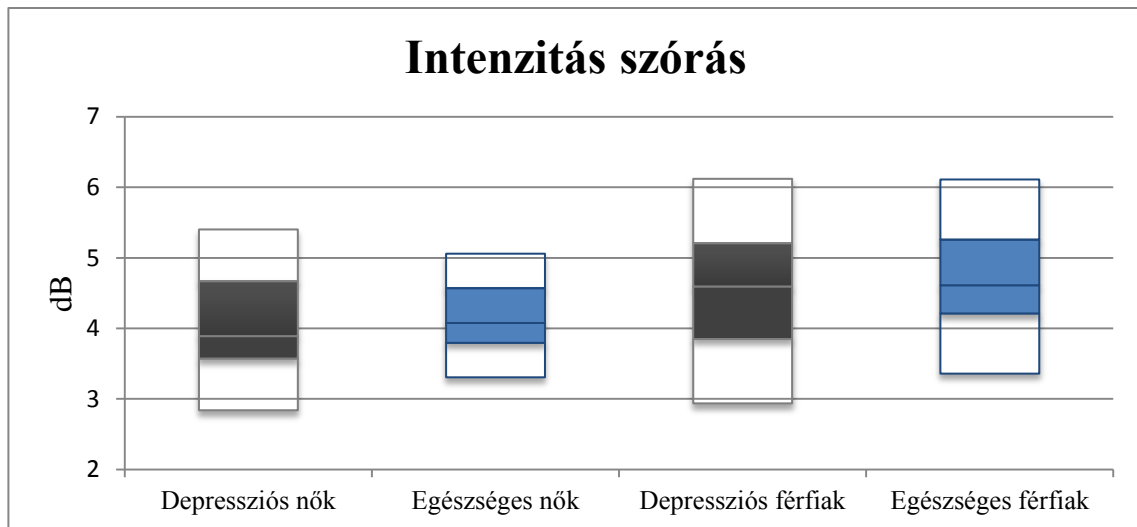


5.7. ábra Egészséges és depressziós mel-sáv energiakülönbség

5.2 Szupraszegmentális paraméterek

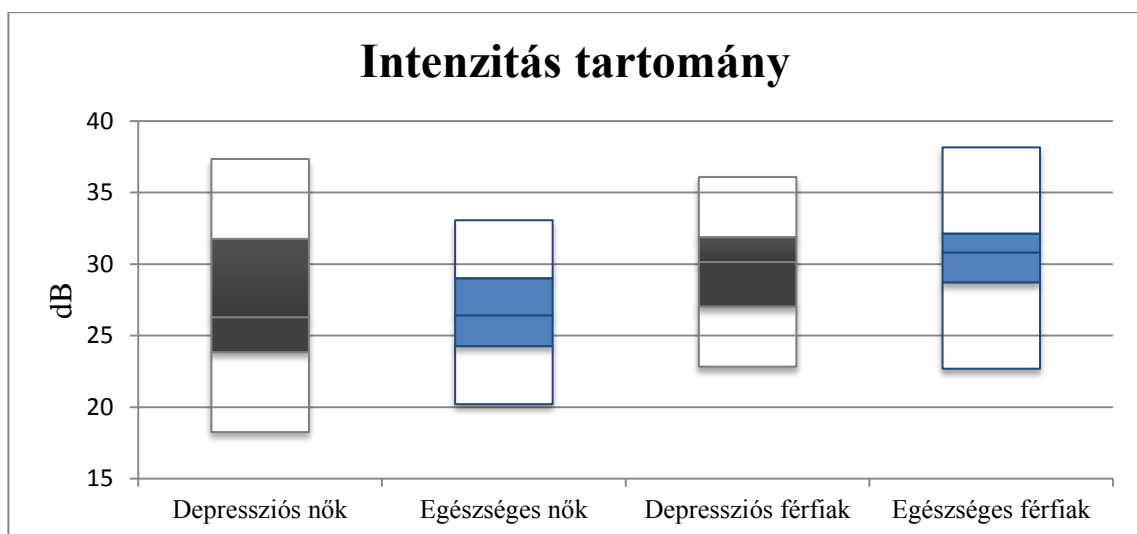
5.2.1 Intenzitás szórás és tartomány

Az előfeltevés, miszerint a depressziós betegek intenzitás szórása és tartománya is alacsonyabb, nem bizonyult igaznak. Egy felolvasáson belül az intenzitás változtatásával színesebbé válik a beszéd, emiatt úgy gondoltam, hogy a monotonnak hangzó depressziós beszédben eltérések lesznek megfigyelhetőek.



5.8. ábra Intenzitás szórás

Az 5.8. ábrán jól látszik, hogy a két tanulmányozott csoport nem tér el egymástól. Az egészséges beszélők mediánja enyhén magasabb, mint a depressziósoké, ám az értékeket szinte azonos tartományban veszik fel.

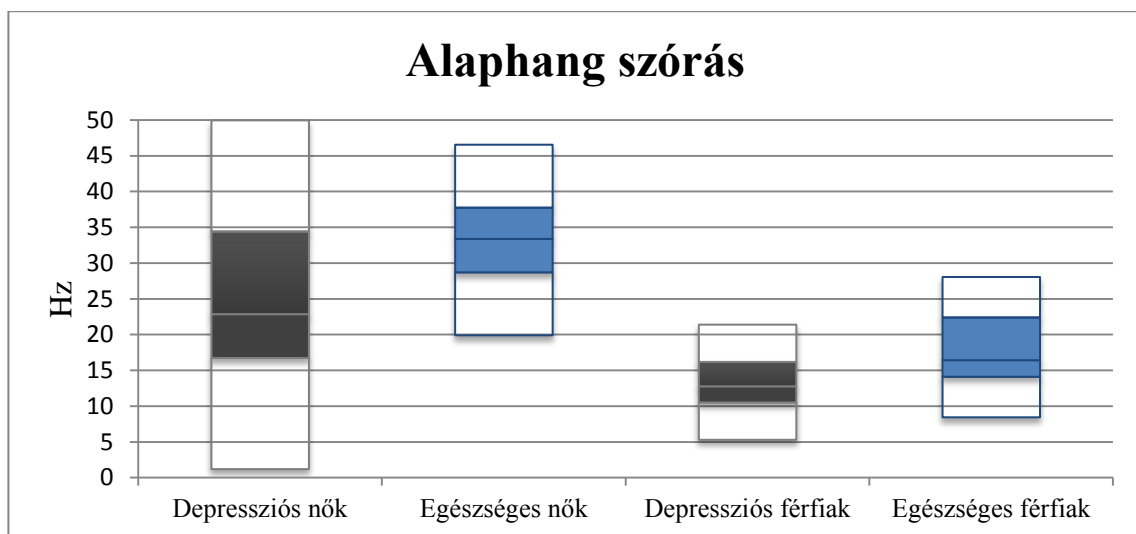


5.9. ábra Intenzitás tartomány

A felvétel alatti legmagasabb és legalacsonyabb intenzitás értékek közti különbség sem bizonyult megfelelőnek a két csoport elkülönítésére, ahogy az 5.9. ábrán látható. A férfiak szélesebb tartományban változtatták a hangerejüket felolvasás közben.

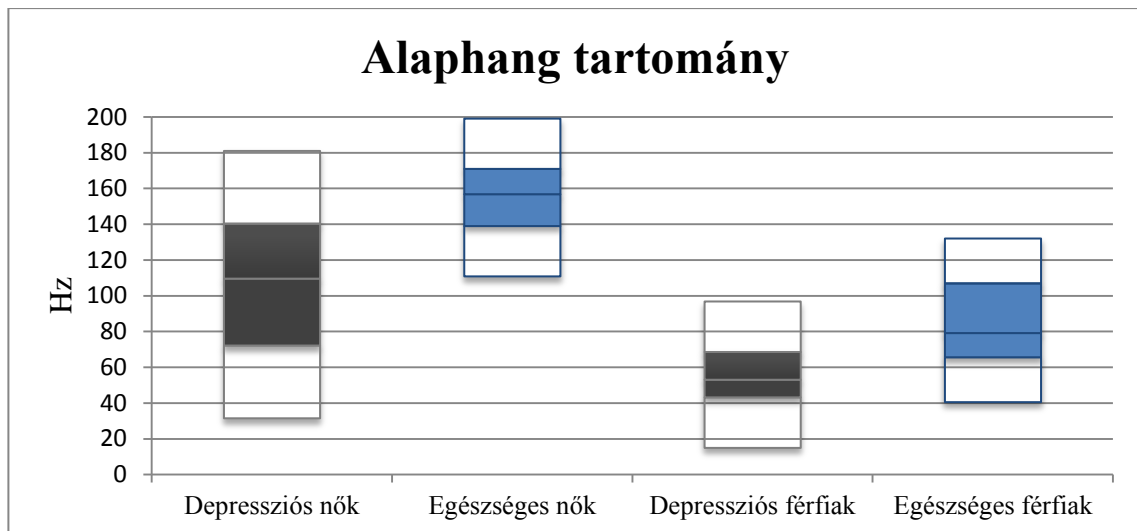
5.2.2 Alaphang szórás és tartomány

Az alaphang szórása és tartománya is jelentős különbségeket mutatott a depressziós és egészséges csoport között. A két osztály közti eltérések a szórás mentén jól láthatóak az 5.10. ábrán. A depressziós nők közül sokan voltak, akik az egészséges társaikkal megegyező szórásértéket produkáltak a felolvasások során, ám számos olyan beszélő is olvasott, akiknek az alaphangjuk szinte változatlan maradt a felolvasás során.



5.10. ábra Alaphang szórás

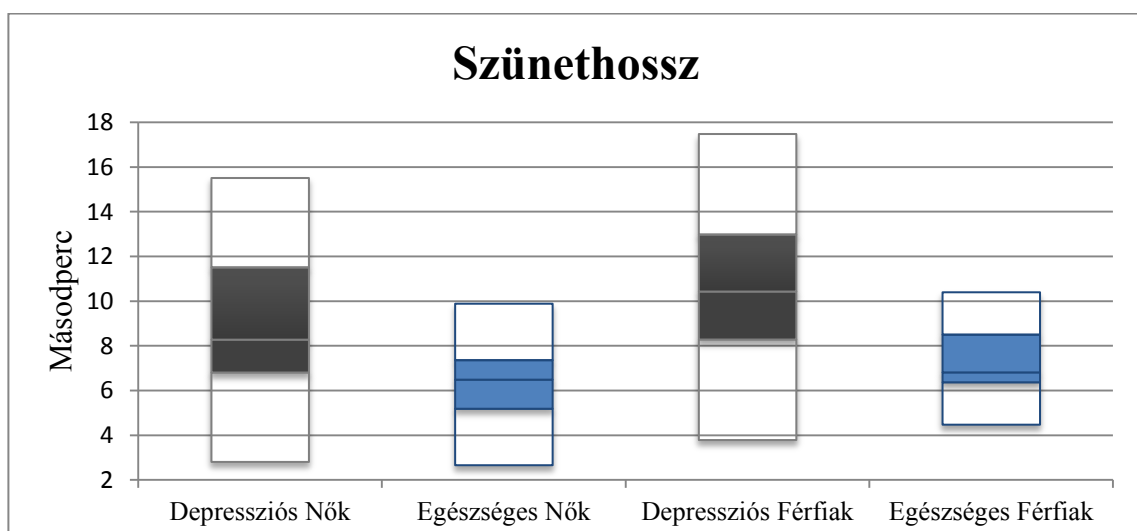
A tartományok közti eltérések is szignifikánsnak bizonyultak, ahogy az 5.11. ábrán látható. A nőknél és a férfiaknál is hasonló eltérés látható az alaphang tartományok tekintetében, mint amilyen differencia a szórásoknál megfigyelhető volt a tanulmányozott csoportok között.



5.11. ábra Alaphang tartomány

5.2.3 Szünethossz

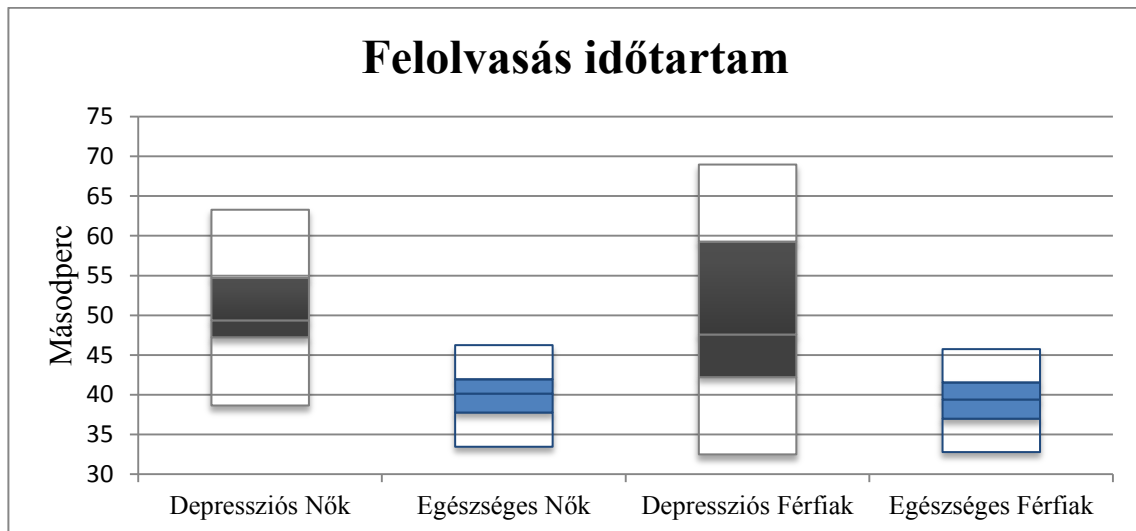
Az említett publikációkban a szünethossz magas osztályozóerejű paraméternek bizonyult. A felvétel alatt tartott szünetek hossza valóban eltéréseket mutat a két általam vizsgált csoport között is – 5.12. ábra. A depressziós személyek szünetei közti szórás nagyobb, mint az egészséges személyeknél, ám a tendencia így is kivehető. A depressziós személyek több szünetet tartottak a felolvasások alatt, ami a megakadásokból és az apró hibákból tevődik össze. A tapasztalt eltérés a kapott nyugtatók mellékhatása is lehet. A férfiak a nőkhez képest több szünetet tartottak.



5.12. ábra Szünethossz

5.2.4 Felolvasás időtartam

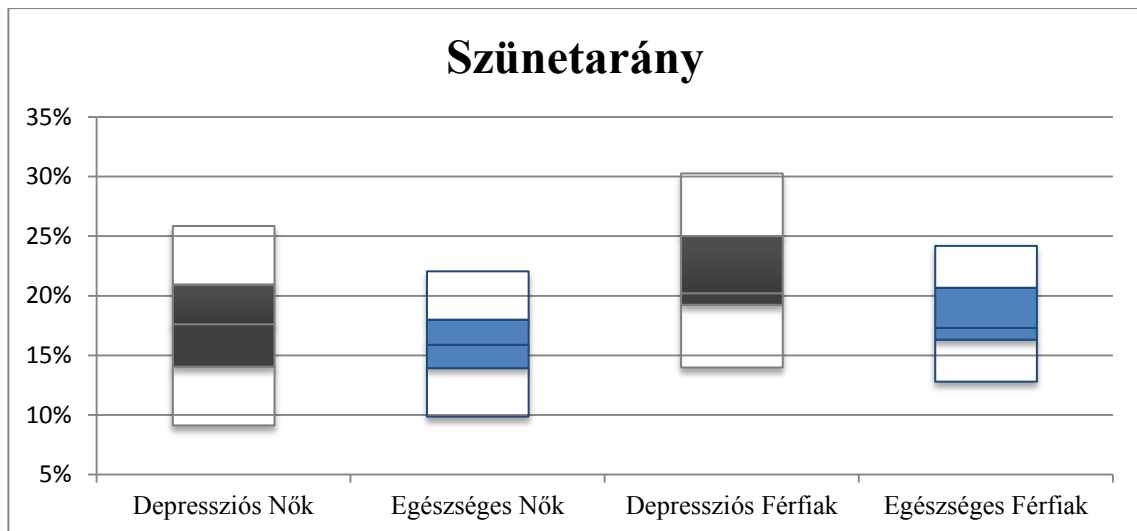
A felolvasáshoz szükséges idő közti eltérések láthatóak az 5.13. ábrán. A nőknél nagyobb, a férfiaknál kisebb eltérések voltak megfigyelhetők, ám a depressziós férfiak felolvasásához szükséges időnek lényegesen nagyobb a szórása is. A teljes felolvasásra fordított idő közti különbségek a szünetek hosszától és az artikulációs tempótól is függ. Mivel az artikulációs tempó alacsonyabb, a szünetek hossza magasabb, ezért a felolvasás időtartama is magasabb a depressziós csoportoknál.



5.13. ábra Felolvasás időtartam

5.2.5 Szünetarány

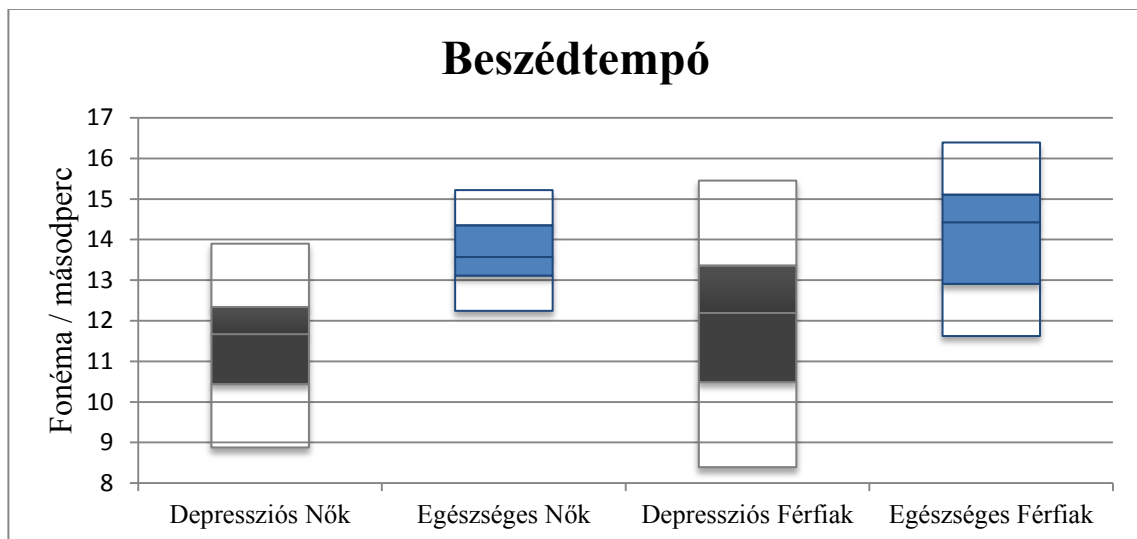
A férfiak felolvasáskor arányaiban több szünetet tartottak, mint a nők. A depressziós csoport nemenként magasabb szünetarányal bír az egészséges személyekhez képest – 5.14. ábra -, azonban az eltérés nem olyan mértékű, mint amit a szünetek hosszánál, vagy a felolvasáshoz szükséges időnél tapasztaltam.



5.14. ábra Szünetarány

5.2.6 Beszédtempó

A kapcsolódó tanulmányokban jelentős különbségeket véltek felfedezni a depressziós személyek beszédtempójában az egészséges személyekhez képest. A paraméter az általam vizsgált csoportok között is komoly eltérést mutat – 5.15. ábra. A depressziós betegek jóval lassabban beszélnek az egészséges csoporthoz képest.

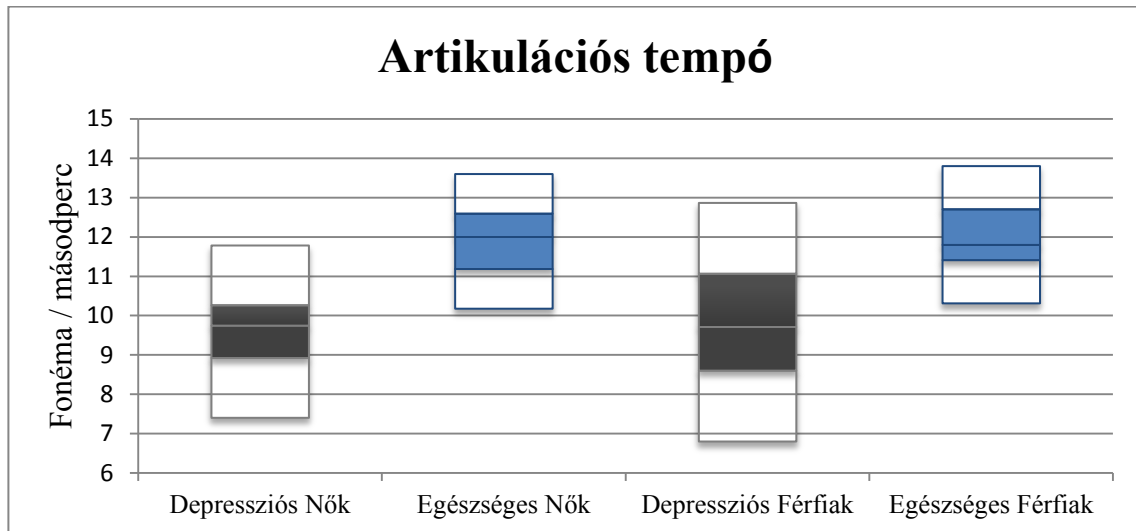


5.15. ábra Beszédtempó

5.2.7 Artikulációs tempó

A korábbi tanulmányok és az általam mért értékek is a feltevésemet igazolták, miszerint a depressziós artikulációs tempó alacsonyabb. Fontos különbség a beszédtempóhoz képest,

hogy az artikulációs tempó számításánál a megnövekedett szünethosszok nem játszanak szerepet, és az eltérések ennek ellenére is jól kivehetőek az 5.16. ábrán.



5.16. ábra Artikulációs tempó

5.3 Statisztikai analízis

5.3.1 A statisztikai próbák általános leírása

A depressziós és egészséges csoportok közti különbségek szignifikancia vizsgálatára a kétmintás t-próba két esetét használtam. A t-próba megköveteli a valószínűségi változók normális eloszlását, azonban egyik esetben a szórások egyezése is szükséges, míg a Scheffé módszernek nem feltétele a szórás egyezés. A normális eloszlás vizsgálatához Kolmogorov-Szmirnov próbát, a szórások egyezésének vizsgálatához pedig F-próbát használtam.

5.3.1.1 Kolmogorov-Szmirnov-féle egymintás próba [19]

A normális eloszlás ellenőrzéséhez használtam a statisztikát. Legyen a standard normális eloszlás eloszlásfüggvénye F , és $F_n^*(x)$ az X -re vett X_1, \dots, X_n mintákból számolt empirikus eloszlásfüggvény. A nullhipotézis:

$$H_0 : X \text{ eloszlásfüggvénye } F.$$

Az alábbi statisztika használható a H_0 hipotézis eldöntésére:

$$D_n = \sup_{-\infty < x < \infty} |F_n^*(x) - F(x)|.$$

5.3.1.2 F-próba [20]

A nemenkénti depressziós és egészséges csoport szórásainak vizsgálatához F-próbát használtam, mely két független, normális eloszlású valószínűségi változó szórásainak egyenlőségét ellenőrzi. A nullhipotézis a szórások egyezése:

$$H_0 : \sigma_1^2 = \sigma_2^2.$$

A H_0 hipotézis igazolására az alábbi statisztikát használtam, ahol s^* a korrigált empirikus szórásnégyzet.

$$F^* = \max \left\{ \left(\frac{s_1^{*2}}{s_2^{*2}}, \frac{s_2^{*2}}{s_1^{*2}} \right) \right\}$$

A H_0 hipotézist akkor fogadhatjuk el, ha a nemenként kapott F^* nem haladja meg az adott szinten meghatározott kritikus értéket.

5.3.1.3 Kétmintás t-próba [20]

X és Y két független, normális eloszlású valószínűségi változó. A szórások egyezésének függvényében két típusa használható a kétmintás t-próbának.

- I. típus

Amennyiben F-próba segítségével bebizonyosodott a valószínűségi változók szórásának egyezése, úgy a hipotézis vizsgálatára az alábbi statisztikát használtam:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{(n_1 - 1) * s_{n_1}^{*2} + (n_2 - 1) * s_{n_2}^{*2}}} * \sqrt{\frac{n_1 * n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

A nullhipotézis a következő: $H_0 : \mathbb{E}X = \mathbb{E}Y$. A szabadsági fok: $Df = n_1 + n_2 - 2$

- II. típus – Scheffé módszere

A módszer akkor használható, ha a szórások nem azonosak. Legyen $n_1 \leq n_2$, és definiáljuk a $Z_1 \dots Z_n$ valószínűségi változókat a következőképpen:

$$Z_i = X_i - \left(\frac{n_1}{n_1} \right)^{\frac{1}{2}} Y_i + \frac{1}{\sqrt{n_1 n_2}} \sum_{j=1}^{n_1} Y_j - \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j$$

Belátható, hogy a Z_i normális eloszlású, $\mathbb{E}Z = \mathbb{E}X - \mathbb{E}Y$ és $\mathbb{D}^2Z = \sigma_X^2 + \frac{n_1}{n_2}\sigma_Y^2$. Az így kapott mintaelemekre egymintás t-próbát alkalmazhatunk, ahol a t értékét az alábbi képlettel számoljuk:

$$t = \frac{\bar{Z} - \mathbb{E}Z}{s_n^*} \sqrt{n}$$

a következő nullhipotézissel: $H_0 : \mathbb{E}Z = 0$. Az I. típustól eltérően a szabadsági fok: $Df = n_1 - 1$

5.3.2 Normalitás vizsgálat

A Kolmogorov-Szmirnov-féle próba szignifikancia szintjét 95 százaléknak választottam, így a H_0 hipotézist, miszerint a minták adott realizációja normális eloszlást vesz fel, akkor fogadom el, ha a számított D_n érték kisebb, mint a 95 százalékos szinthez tartozó, mintaszámnak megfelelő kritikus érték. A depressziós férfiak mintaszáma 15, a többi csoporté 18. A hozzájuk tartozó kritikus értékek 0,338 és 0,309. A statisztikai vizsgálat eredményét mutatja be a melléklet 1. táblázata, miszerint mindegyik vizsgált paraméterre elfogadtam a nullhipotézist, tehát a vizsgált jellemzők eloszlása normális eloszlást követ.

5.3.3 Szignifikancia vizsgálat

A kétmintás t-próba az összes paraméterre alkalmazható, mivel az eloszlásuk normális eloszlást követ. A melléklet 2. táblázata mutatja be az F-próba 90 százalékos szinten számolt értékeit. Abban az esetben, ha a szórások egyezése bebizonyosodott az egészséges és depressziós csoportok között, úgy a kétmintás t-próba I. típusát, öt esetben pedig, mikor nem sikerült kimutatni az egyezést, akkor a Scheffé módszert használtam a szignifikancia vizsgálatra. A nullhipotézist, miszerint a két eloszlás várható értéke megegyezik, akkor kell elutasítani, ha a számított t érték meghaladja a megfelelő szignifikancia szinthez tartozó Student-t eloszlás kritikus értékét. 90 százalékos szint alatt nem vizsgáltam a hipotézist, viszont a 90, 95 és 99 százalékos szintek közül mindig a legmagasabbat választottam ki, melyre még elvethető a várható értékek egyezése. A kétmintás t-próbával kapott eredményeket az 5.1. táblázat foglalja össze. Azok a paraméterek, melyeknél legalább 90 százalékos szinten a férfi és női csoportnál is elvettem a nullhipotézist félkövéren vannak kiemelve.

Akusztikai paraméter	Nem	t érték	Kritikus érték	Szignifikancia szint	Kétmintás t-próba típusa
Alapfrekvencia	Nők	4,842	2,728	99%	I. típus
	Férfiak	2,127	2,040	95%	I. típus
Első formánsfrekvencia	Nők	2,086	2,032	95%	I. típus
	Férfiak	2,946	2,744	99%	I. típus
Második formánsfrekvencia	Nők	1,775	1,691	90%	I. típus
	Férfiak	0,833	1,696	<90%	I. típus
Jitter	Nők	2,281	2,110	95%	Scheffé módszer
	Férfiak	1,064	1,696	<90%	I. típus
Shimmer	Nők	1,112	1,691	<90%	I. típus
	Férfiak	0,697	1,761	<90%	Scheffé módszer
Felolvasás időtartam	Nők	3,782	2,898	99%	Scheffé módszer
	Férfiak	3,337	2,744	99%	I. típus
Szünethossz	Nők	1,918	1,691	90%	I. típus
	Férfiak	2,997	2,977	99%	Scheffé módszer
Szünetarány	Nők	0,730	1,691	<90%	I. típus
	Férfiak	2,160	2,040	95%	I. típus
Beszédtempó	Nők	4,131	2,728	99%	I. típus
	Férfiak	3,319	2,744	99%	I. típus
Artikulációs tempó	Nők	9,724	2,898	99%	Scheffé módszer
	Férfiak	2,826	2,744	99%	I. típus
Alaphang szórás	Nők	1,137	1,691	<90%	I. típus
	Férfiak	1,745	1,696	90%	I. típus
Alaphang tartomány	Nők	3,474	2,728	99%	I. típus
	Férfiak	1,758	1,696	90%	I. típus
Intenzitás szórás	Nők	0,101	1,691	<90%	I. típus
	Férfiak	0,600	1,696	<90%	I. típus
Intenzitás tartomány	Nők	0,838	1,691	<90%	I. típus
	Férfiak	1,598	1,696	<90%	I. típus

5.1. táblázat Szignifikancia vizsgálat eredményei

5.4 Eredmények értékelése

- Alapfrekvencia

Az 5.1. ábrán látható, hogy a depressziós férfiak és nők is alacsonyabb alaphangon beszélnek, mint az egészséges társaik. A szignifikancia vizsgálat során a nőknél 99, a férfiaknál 95 százalékos szinten bizonyosodott be, hogy a várható értékük különbözik. Az ismertett irodalmakban hasonló eltolódásról nem számoltak be.

- Első formánsfrekvencia

Az első formánsfrekvencia eltolódását depressziós esetben már korábban is felismerték [4], [7]. Az eltérések az általam vizsgált csoportokban is láthatóak az 5.2. ábrán, és a szignifikancia vizsgálat során is bebizonyosodott, hogy a nők és a férfiak első formánsfrekvenciájának várható értéke is különbözik 95 és 99 százalékos szinten.

- Második formánsfrekvencia

Az 5.3. ábrán látható, hogy a csoportok közti differencia nőknél kismértékben észrevehető, míg férfiaknál jelentéktelen, annak ellenére, hogy az előzetesen ismertett kutatások egyik fontos differenciáló paramétere volt [4], [6], [7]. A szignifikancia vizsgálat során csupán a nők esetén lehetett elutasítani a nullhipotézist - várható értékek egyezését - és azt is csak 90 százalékos szinten.

- Jitter

Az 5.4. ábra szerinti sűrűségfüggvény szerint a két csoport hasonló jitter értékekkel bír, azonban az egészséges beszélők az alacsonyabb tartományban jóval több értéket vesznek fel, mint a depressziós személyek, miközben a depressziós csoport a magasabb tartományban is számos értékkel bír. A két vizsgált osztály női jitter értékeinek várható értéke a kétmintás t-próba alapján statisztikailag nem egyezik 95 százalékos szignifikancia szinten, és a minták szórása is eltér az F-próba alapján 90 százalékos szint mellett. Férfiaknál ilyen eltérések nem tapasztalhatóak.

- Shimmer

A női sűrűségfüggvények szinte teljesen megegyeznek, míg a férfiak esetén a depressziós sűrűségfüggvény maximuma el van tolva a magasabb értékek felé az egészségeshez képest, az 5.5. ábra alapján. A férfi shimmer értékek szórása 90 százalékos szinten eltérőnek bizonyult az F-próba alapján. A kétmintás t-próba segítségével azonban nem sikerült egyik nem esetén sem szignifikáns eltérést kimutatni.

- Mel-sávok energia

A korábbi kutatások során alkalmasnak találták az egészséges és a depressziós csoport elkülönítésére. A két csoport közti eltérések nálam is jól kivehetőek az 5.7. ábrán. Az alacsonyabb mel sávokban magasabb a depressziós beszélők intenzitása, a felsőbb sávok felé pedig az egészséges személyeknek magasabb az intenzitása.

- Felolvasás időtartam

A felolvasáshoz a depressziós csoportoknak több időre volt szüksége, mint az egészséges személyeknek, ahogy az 5.13. ábrán látható. A férfi csoportok szórásai eltérőnek bizonyultak az F-próbával, a nők megegyeztek. Kétmintás t-próbával 99 százalékos szinten szignifikánsan eltérőnek bizonyultak a várható értékei a depressziós és egészséges csoportnak, mindkét nem esetén.

- Szünethossz

Az 5.12. ábrán jól látható, hogy a depressziós személyek több szünetet tartottak az egészséges beszélőkhöz képest. A férfiak szórásai eltérőnek bizonyultak, a nők szórásai nem tértek el 90 százalékos szinten az F-próba szerint. A szignifikancia vizsgálat során a várható értékek egyezését a nőknél 90, míg a férfiaknál 99 százalékos szinten is el kellett vetni.

- Szünetarány

A vizsgált csoportok szünetarányai közti eltérés látható az 5.14. ábrán. Az egészséges csoportok kevesebb szünetet tartottak arányaiban a felolvasás alatt, mint a depressziós beszélők. Férfiaknál, nőknél egyaránt egyenlőnek bizonyultak a szórások az F-próba alapján, azonban a kétmintás t-próbával csupán a férfiak szünetarányának a várható értékei különböztek 95 százalékos szinten.

- Beszédtempó

A beszédtempó csoportok közti eltérései az 5.15. ábrán láthatóak. A depressziós személyek alacsonyabb beszédtempóval beszéltek, ami természetesen a megnövekedett szünethosszoknak is köszönhető. A kétmintás t-próba 99 százalékos szintjén férfiaknál és nőknél egyaránt el kellett vetni a hipotézist, miszerint a várható értékeik megegyeznének.

- Artikulációs tempó

A beszédtempóhoz hasonló eltérés látható a két csoport között az 5.16. ábrán, tehát a depressziós betegek nem csak annak köszönhetően beszélnek lassabban, mert több szünetet tartanak, hanem az artikulációs tempójuk is alacsonyabb. A szignifikancia vizsgálat során 99 százalékos szinten is el kell utasítani a várható értékek egyezését nemtől függetlenül.

- Alaphang szórás és tartomány

Férfiaknál és nőknél is jól látható az 5.10. és 5.11. ábrán az alaphang szórása és tartománya közti eltérés a csoportok között. A kétmintás t-próba eredményei alapján az

alaphang tartomány várható értéke nagyobb mértékben különbözik a csoportok között, mint a szórása. Az alaphang szórása nőknél nem mutatott szignifikáns eltérést, míg férfiaknál 90 százalékos szinten kellett elvetni a kétmintás t-próba nullhipotézisét. Az alaphang tartomány esetében azonban nőknél 99 százalékos szignifikancia szinten el kell vetni a hipotézist, férfiaknál 90 százalékos szinten vethetjük csak el a várható értékek egyezését.

- Intenzitás szórás és tartomány

Az előzetes feltevés, miszerint a csoportok intenzitás dinamikája alkalmas lesz a két osztály elkülönítésére nem bizonyult igaznak. Az 5.8. és 5.9. ábrán nem tapasztalható számottevő eltérés a depressziós és egészséges csoport intenzitás szórás és tartomány értékei között sem. A kétmintás t-próba hipotézise, miszerint a várható értékek egyenlők, 90 százalékos szinten sem volt elvethető.

6 Konklúzió

A WHO aggasztó előrejelzései szerint a jövőben a depressziós betegek száma a jelenlegi 350 millióról tovább fog emelkedni. A jelenleg alkalmazott diagnosztizáló és monitorozó módszerek helyett szükséges egy objektív mérő rendszer bevezetése, ami nagymértékben segítené a szakismerettel nem rendelkező orvosok és ápolók munkáját. A mobil eszközök és az internet széleskörű elterjedésével számos új lehetőség nyílt a távoroslás előtt, így a pszichiátriai betegségek hang alapú diagnosztizálása is lehetségessé vált technológiailag. A biztató kutatási eredmények ellenére, legjobb tudomásom szerint a mai napig nem sikerült a gyakorlatban megvalósítani egy hang alapú pszichiátriai diagnosztizáló eszközt.

Dolgozatomban számos akusztikai paramétert azonosítottam, melyek szignifikáns különbséget mutattak a depressziós és az egészséges csoportok között. Az irodalom feldolgozás során bemutatott tanulmányokban ismertetett paraméterek, úgymint az alaphang tartomány, az első formánsfrekvencia, mel-sávós energia és beszédtempó az általam végzett mérésekben is eltéréseket mutattak, ám a második formánsfrekvencia, jitter és shimmer értékekben tapasztalt különbségek nem voltak megfigyelhetők. Az alaphangfrekvencia, az alaphangfrekvencia ingadozása, az első formánsfrekvencia, a felolvasás időtartama, a szünethossz, a beszéd és az artikulációs tempó mindkét nem esetén szignifikáns különbségeket mutatott a két vizsgált csoport között a szignifikancia vizsgálat során. Ezek a paraméterek feltehetőleg alkalmasak lehetnek egy automatikus osztályozást megvalósító gépi eljárás bemenetének. Az előzetes feltevés, miszerint az intenzitás ingadozás tekintetében is eltérés várható nem bizonyult igaznak.

Az általam kapott eredmények sok tekintetben bizakodásra adnak okot, ám a vizsgált adatbázis jelenlegi mérete és a vizsgált depressziós személyek magas BDI értéke miatt ezek még csak részeredmények. A vizsgált személyek száma a témában publikált tanulmányokban sem magasabb, mivel körülményes az adatgyűjtés, továbbá a depresszió ritkán jelentkezik egyedül, legtöbbször más neurológiai betegségekkel társul. Az alacsonyabb BDI értékű, enyhe depresszióban szenvedő betegek hangmintáival a betegség felismerésének legfontosabb szakaszában is vizsgálható lenne a depresszió, továbbá a kötetlen szövegek elemzése is segíthetne az elkülönítésben. Számos vizsgált paraméter is szignifikáns különbséget mutatott az egészséges és depressziós csoportok között, az automatikus osztályozás megvalósítható. Feltehetőleg gépi osztályozással a depresszió automatikusan detektálható, akár súlyossága is megállapítható lenne.

7 Köszönetnyilvánítás

Köszönetet szeretnék mondani Dr. Vicsi Klára Tanárnőnek, Kiss Gábornak és Sztahó Dávidnak a rengeteg segítségért, amit tőlük kaptam.

8 Irodalomjegyzék

- [1] Marina Marcus, M. Taghi Yasamy, Mark van Ommeren, and Dan Chisholm, Shekhar Saxena, *Depression – A Global Public Health Concern*, WHO Department of Mental Health and Substance Abuse, 2012
- [2] Colin D. Mathers, Dejan Loncar, *Projections of Global Mortality and Burden of Disease from 2002 to 2030*, PLoS Med 3(11): e442, 2006
- [3] Sharifa Alghowinem, Roland Goecke, Michael Wagner, Julien Epps, Michael Breakspear, Gordon Parker, *Detecting Depression – A Comparison Between Spontaneous and Read Speech*, 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013
- [4] Brian S. Helfer, Thomas F. Quatieri, James R. Williamson, Daryush D. Mehta, Rachelle Horwitz, Bea Yu, *Classification of depression state based on articulatory precision*, 14th Annual Conference of the International Speech Communication Association, 2013
- [5] Paul Boersma, *Praat, a system for doing phonetics by computer*. Glot International 5:9/10, 341-345., 2001
- [6] James C. Mundt, Peter J. Snyder, Michael S. Cannizzaro, Kara Chappie, Dayna S. Geralts, *Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology*, J Neurolinguistics, 2007
- [7] Daniel J. France, Richard G. Shiavi, Senior Member, IEEE, Stephen Silverman, Marilyn Silverman, and D. Mitchell Wilkes, Member, IEEE, *Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk*, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 47, NO. 7, 2000
- [8] MATLAB version 7.10.0. Natick, Massachusetts: The MathWorks Inc., 2010
- [9] Abela, J. R. Z., & D'Allesandro, D. U. *Beck's cognitive theory of depression: The diathesis-stress and causal mediation components*. British Journal of Clinical Psychology, 41, 111-128., 2002
- [10] Wells, J.C., *SAMPA computer readable phonetic alphabet*. In Gibbon, D., Moore, R. and Winski, R. (eds.), *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter. Part IV, section B., 1997

- [11] Vicsi Klára, *Beszédkommunikáció*, BME TMIT – Laboratory of Speech Acoustics <http://alpha.tmit.bme.hu/speech/docs/education/beszedkomm.pdf>, 2003
- [12] Vicsi Klára, *A magyar beszéd 3.3 fejezet – A beszéd fizikai jellemzése*, Akadémia Kiadó, ISBN: 9789630589666, szerkesztette Németh Géza, Olasz Gábor, 2010
- [13] Paul Boersma, *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound*, Proceedings of the Institute of Phonetic Sciences 17: 97–110. University of Amsterdam, 1993
- [14] Olasz Gábor, *A magyar beszéd 5.1 fejezet – A magyar beszédhangok*, Akadémia Kiadó, ISBN: 9789630589666, szerkesztette Németh Géza, Olasz Gábor, 2010
- [15] Donald G. Childers, *Modern spectrum analysis: 252-255*. IEEE press, New York, 1978
- [16] Olasz Gábor, *A magyar beszéd 6.4 fejezet – Időszerkezeti tényezők*, Akadémia Kiadó, ISBN: 9789630589666, szerkesztette Németh Géza, Olasz Gábor, 2010
- [17] Abari Kálmán, Rácz Zsuzsanna Zsófia, Olasz Gábor, *Formant maps in Hungarian vowels - online data inventory for research, and education*, 12th Annual Conference of the International Speech Communication Association 1262-1265, 2011
- [18] Olasz Gábor, *A magyar beszéd 6.1 fejezet – A beszéddallam*, Akadémia Kiadó, ISBN: 9789630589666, szerkesztette Németh Géza, Olasz Gábor, 2010
- [19] Fazekas István, *Bevezetés a matematikai statisztikába V. fejezet – Nemparaméteres próbák*, Kossuth Egyetemi Kiadó, szerkesztette Fazekas István, 2005
- [20] Nagy Márta, *Bevezetés a matematikai statisztikába IV. fejezet – Hipotézisek vizsgálata. Paraméteres próbák*, Kossuth Egyetemi Kiadó, szerkesztette Fazekas István, 2005

9 Melléklet

Akusztkai paraméter	Csoport	Nem	D_n	Kritikus érték	Szignifikancia szint
Alapfrekvencia	Depressziós	Nők	0,210	0,309	95%
		Férfiak	0,245	0,338	95%
	Egészséges	Nők	0,155	0,309	95%
		Férfiak	0,200	0,309	95%
Első formánsfrekvencia	Depressziós	Nők	0,223	0,309	95%
		Férfiak	0,227	0,338	95%
	Egészséges	Nők	0,217	0,309	95%
		Férfiak	0,193	0,309	95%
Második formánsfrekvencia	Depressziós	Nők	0,243	0,309	95%
		Férfiak	0,198	0,338	95%
	Egészséges	Nők	0,190	0,309	95%
		Férfiak	0,228	0,309	95%
Jitter	Depressziós	Nők	0,275	0,309	95%
		Férfiak	0,332	0,338	95%
	Egészséges	Nők	0,189	0,309	95%
		Férfiak	0,267	0,309	95%
Shimmer	Depressziós	Nők	0,158	0,309	95%
		Férfiak	0,278	0,338	95%
	Egészséges	Nők	0,131	0,309	95%
		Férfiak	0,307	0,309	95%
Felolvasás időtartam	Depressziós	Nők	0,211	0,309	95%
		Férfiak	0,227	0,338	95%
	Egészséges	Nők	0,234	0,309	95%
		Férfiak	0,278	0,309	95%
Szünethossz	Depressziós	Nők	0,236	0,309	95%
		Férfiak	0,216	0,338	95%
	Egészséges	Nők	0,261	0,309	95%
		Férfiak	0,286	0,309	95%
Szünetarány	Depressziós	Nők	0,161	0,309	95%
		Férfiak	0,240	0,338	95%
	Egészséges	Nők	0,212	0,309	95%
		Férfiak	0,266	0,309	95%
Beszédtempó	Depressziós	Nők	0,235	0,309	95%
		Férfiak	0,215	0,338	95%
	Egészséges	Nők	0,160	0,309	95%
		Férfiak	0,175	0,309	95%
Artikulációs tempó	Depressziós	Nők	0,165	0,309	95%

		Férfiak	0,185	0,338	95%
	Egészséges	Nők	0,183	0,309	95%
		Férfiak	0,170	0,309	95%
Alaphang szórás	Depressziós	Nők	0,279	0,309	95%
		Férfiak	0,214	0,338	95%
	Egészséges	Nők	0,225	0,309	95%
		Férfiak	0,303	0,309	95%
Alaphang tartomány	Depressziós	Nők	0,206	0,309	95%
		Férfiak	0,252	0,338	95%
	Egészséges	Nők	0,146	0,309	95%
		Férfiak	0,232	0,309	95%
Intenzitás szórás	Depressziós	Nők	0,213	0,309	95%
		Férfiak	0,219	0,338	95%
	Egészséges	Nők	0,221	0,309	95%
		Férfiak	0,230	0,309	95%
Intenzitás tartomány	Depressziós	Nők	0,243	0,309	95%
		Férfiak	0,166	0,338	95%
	Egészséges	Nők	0,182	0,309	95%
		Férfiak	0,302	0,309	95%

1. táblázat: Normalitás vizsgálat – Kolmogorov-Szmirnov próba eredmények

Akusztikai paraméter	Nem	F^*	Kritikus érték	Szignifikancia szint
Alapfrekvencia	Nők	1,397	3,128	90%
	Férfiak	1,557	3,422	90%
Első formánsfrekvencia	Nők	2,251	3,128	90%
	Férfiak	2,635	3,422	90%
Második formánsfrekvencia	Nők	1,040	3,128	90%
	Férfiak	1,047	3,422	90%
Jitter	Nők	3,598	3,128	<90%
	Férfiak	1,774	3,422	90%
Shimmer	Nők	1,043	3,128	90%
	Férfiak	3,659	3,422	<90%
Felolvasás időtartam	Nők	4,134	3,128	<90%
	Férfiak	2,786	3,422	90%
Szünethossz	Nők	2,627	3,128	90%
	Férfiak	3,794	3,422	<90%
Szünetarány	Nők	1,557	3,128	90%
	Férfiak	2,207	3,422	90%
Beszédtempó	Nők	2,489	3,128	90%
	Férfiak	1,715	3,422	90%

Artikulációs tempó	Nők	3,282	3,128	<90%
	Férfiak	1,570	3,422	90%
Alaphang szórás	Nők	3,081	3,128	90%
	Férfiak	1,417	3,422	90%
Alaphang tartomány	Nők	2,086	3,128	90%
	Férfiak	1,054	3,422	90%
Intenzitás szórás	Nők	2,195	3,128	90%
	Férfiak	1,050	3,422	90%
Intenzitás tartomány	Nők	1,870	3,128	90%
	Férfiak	2,471	3,422	90%

2. táblázat: Szórás vizsgálat – F-próba eredmények