



M Ű E G Y E T E M 1 7 8 2

**Budapesti Műszaki és Gazdaságtudományi Egyetem**  
Villamosmérnöki és Informatikai Kar  
Automatizálási és Alkalmazott Informatikai Tanszék

Szabó Máté

**KÉPESSÉGSZINT BECSLÉSÉT  
VÉGZŐ PSZICHOMETRIAI  
MODELL KITERJESZTÉSE  
FIZIOLÓGIAI TÉNYEK ALAPJÁN**

KONZULENS

**Dr. Forstner Bertalan**

BUDAPEST, 2017

# Tartalomjegyzék

<b>Összefoglaló .....</b>	<b>4</b>
<b>Abstract.....</b>	<b>5</b>
<b>1 Bevezetés .....</b>	<b>6</b>
<b>2 IRT típusú pszichometriai modellek .....</b>	<b>8</b>
2.1 Motiváció .....	8
2.2 Item response theory .....	8
2.3 Alapfogalmak.....	9
2.4 A 3PL és az itemválasz függvény.....	11
2.5 Egyéb IRT modellek.....	13
2.6 IRT modellek paramétereinek becslése .....	14
2.7 Az IRT alkalmazása.....	17
2.8 A hagyományos IRT modellek korlátai.....	18
<b>3 Mentális állapot mérése fiziológiai tények segítségével .....</b>	<b>19</b>
3.1 Biofeedback .....	19
3.2 EEG.....	20
3.2.1 Elméleti alapok .....	20
3.2.2 Felhasznált eszközök .....	21
3.3 Szívritmus .....	24
3.3.1 Elméleti alapok .....	24
3.3.2 A felhasznált eszköz .....	25
3.4 Összefoglalás .....	26
<b>4 Mérési kódkönyvtár.....</b>	<b>27</b>
4.1 Motiváció .....	27
4.2 Előzmények.....	27
4.3 Architektúra .....	28
4.3.1 Rendszerarchitektúra .....	28
4.3.2 Kliens oldali architektúra.....	29
4.3.3 Felhasznált technológiák.....	30
4.4 Fő funkciók .....	30
4.4.1 Alkalmazások csatlakoztatása.....	30
4.4.2 Események küldése.....	31

4.4.3 Lokális adatmentés .....	31
4.4.4 Együttműködés a szenzorokkal .....	31
4.5 Összefoglalás .....	31
<b>5 Az IRT modellek kiterjesztése a mentális állapot figyelembevételével.....</b>	<b>33</b>
5.1 Bevezetés .....	33
5.2 A modell bemutatása.....	33
5.3 Mérési paraméterek.....	35
5.3.1 A mérés menete .....	35
5.3.2 Megfeleltetés az IRT modellnek.....	36
5.3.3 Bemeneti paraméterek .....	37
5.3.4 Mért értékek.....	37
5.4 Szimuláció.....	39
5.4.1 Algoritmus és paraméterek .....	39
5.4.2 Szimulációs eredmények .....	40
5.5 Mérési eredmények.....	43
5.5.1 Körülmények .....	43
5.5.2 Eredmények, értékelés .....	43
<b>6 Összegzés.....</b>	<b>52</b>
<b>Irodalomjegyzék.....</b>	<b>53</b>

# Összefoglaló

Az általunk elvégzett feladatok minőségét alapvetően meghatározzák képességeink. Ezek mérése az élet számos területén hangsúlyos. Egy állás betöltésekor például fontos szerepet kap, hogy az adott pozíciót a képességei alapján a lehető legmegfelelőbb személy kapja meg. Munkahelyen, az egyes munkatársak által elvégzendő teendők meghatározásakor célszerű, hogy mindenki a képességeihez igazítva, testreszabottan kapjon feladatot, elősegítve ezzel a hatékony munkavégzést.

Különböző kutatócsoportok általában 4-10 különböző területre osztják az emberi alapképességeket, ezek számszerűsítése azonban továbbra is nyitott kérdés. A közelmúltig különféle pszichometriai modelleket használtak arra, hogy matematikailag is megfoghatóvá tegyék az emberek képességszintje és egy adott nehézségű feladat megoldásakor általuk mutatott teljesítmény közötti kapcsolatot. A kutatások azonban arra az eredményre jutottak, hogy ezek a modellek nem veszik figyelembe, hogy egy személy az adott állapotában mennyire képes a képességei teljes kihasználására. A területen kutatást végző csoportok között egyetértés alakult ki, hogy további tudományos vizsgálat, interdiszciplináris kutatás szükséges a használt pszichometriai modellek további finomításához.

Dolgozatomban bemutatom vizsgálatokra építő eredményeimet, amelyben az ismert pszichometriai modellekhez képest, de azokra építve egy olyan új módszert fejlesztettem ki, amely a vizsgált személy képességei mellett annak releváns fiziológiai állapot mutatóit is képes figyelembe venni. A módszerem alkalmazhatóságához mérnöki tevékenység, hardveres és szoftveres integráció, valamint gondosan kialakított kódkönyvtár megtervezése és megvalósítása is szükségessé vált. A dolgozatban egy általam bevezetett, új paramétereket is tartalmazó zárt modellt adok az Item response típusú feladatok megoldásának valószínűségére, amit szimulációk és a kódkönyvtár segítségével lebonyolított mérések segítségével gondosan megalapozok, valamint vizsgálatokat végzek arra vonatkozóan is, hogy különböző körülmények között milyen teljesítményre képes a kialakított modell.

## **Abstract**

Our abilities fundamentally determine the quality we can achieve while performing tasks. The measurement of ability levels is therefore highly relevant in numerous areas of life. For instance, when filling a particular job opening, it is important that the position is given to the most able and skilled person available. In workplaces, it is advisable that each colleague work on a task which is personalized for their own abilities and needs, which enhances their efficiency.

Different research groups define 4 to 10 different areas of basic ability. However, quantifying these ability levels is still an open question. Until recently, various psychometric models had been used to mathematically describe the connection between people's ability and their performance when solving a task with a given difficulty. Nevertheless, research concluded that these models do not take into account that how much a person can fully utilize their ability in their actual, given state. The research groups agree that more scientific, interdisciplinary research is needed to refine the currently used psychometric models.

In my paper, I present my results based on original research, in which I developed a new method that is able to consider a person's relevant physiological state indicators as well as abilities compared to and built on existing psychometric models. I introduce a new, closed model containing new parameters for the solving probability of Item response tasks, thoroughly established by measurements and simulations. I also conduct research about how well the new model performs under different conditions.

# 1 Bevezetés

Dolgozatom fő motivációja az emberi tanulás témaköre. A tanulás minden életkorban releváns, fontossága fokozatosan nő idősebb életkorban is, hiszen a körülöttünk lévő világ dinamikusan változik a technológiai fejlődéssel összhangban. Napjainkban egyre inkább középpontba kerülnek a hatékony, számítógéppel is támogatott tanulási módszerek: az információs rendszerek felhasználásának egy lehetséges iránya a tanulás támogatása a tanulni kívánó személyről való minél több tény vagy ezek alapján történő következtetés segítségével.

Kutatások szerint a tanulás hatékonysága erős összefüggésben áll a tanuló képességeivel. Ez a felismerés vezetett a képesség, mint tényező mérésének szükségletéhez, azonban hamar belátható, hogy a képesség a fizikai paraméterektől eltérően közvetlenül nem mérhető. Gondoljunk csak bele, hogy ugyanaz az ember különböző teljesítményre lehet képes más-más nehézségű tesztek kitöltésekor, valamint a teszt megírásakor jelen lévő körülmények is döntően befolyásolhatják annak eredményét. Innentől kezdve az első logikus lépés a nehézség, mint a feladatokat alapvetően jellemző tényező belefoglalása a képességekről alkotott modellbe.

A szakirodalomban *item response theory* (a magyar fordításokban gyakran *valószínűségi tesztelmélet* vagy *modern tesztelmélet – sugallva az összehasonlítást a klasszikus tesztelmélettel [8]* – rövidítve gyakran *IRT*) néven emlegetik a személyek képességei és az általuk megoldott feladatok nehézségei között kapcsolatot teremtő, matematikai modelleket felvonultató paradigma-rendszert. Ezek az ún. *pszichometriai modellek* valószínűségi függvényt adnak egy adott képességű személy által adott paraméterekkel rendelkező feladat sikeres megoldására vonatkozóan, ahol a feladat paraméterei közül gyakran csak az egyik a nehézség.

Az IRT típusú modellek matematikai robusztusságát az adja, hogy az embereket egyetlen személyi paraméterrel, a képességükkel jellemzi. Azonban ez a megközelítés azt is jelenti, hogy az ilyen modellekkel alapesetben nem vehető figyelembe az a tapasztalatilag is megerősíthető tény, hogy egy ember teljesítménye annak mentális állapotától is nagyban függ: a kialvatlanság, fáradtság, vagy a felfokozott érzelmi állapot számos esetben ronthatja az alany egy teszten mutatott teljesítményét, míg kipihent, nyugodt, kiegyensúlyozott állapotban jobb esélyekkel indulunk ugyanazon megméréstetesen.

Míg egy teszten a lehető legjobb teljesítmény nyújtása a cél, tanulás esetén inkább a minél nagyobb hatékonyság, valamint az, hogy a tanultak lehetőleg minél jobban rögzüljenek, a tanuló meglévő képességszintjét használja ki annak minél nagyobb mértékű fejlesztésére. A mentális állapot azonban itt is döntő szerephez jut. Ennek mérése nem egyszerű feladat, azonban ma már léteznek olyan, akár olcsó, mindenki számára elérhető, hordozható szenzorok, amelyek képesek bizonyos fiziológiai folyamatok vizsgálata alapján fiziológiai tények előállítására. Ezen tényekből következtethetünk a mért alany mentális állapotára, kellően sok mérést elvégezve pedig megfogalmazható a teljesítmény és a mentális állapot közötti összefüggés.

Kutatásom távlati célja egy olyan rendszer kifejlesztése, ami a hatékony tanulást mind a tanuló képességei, mind annak releváns fiziológiai állapot mutatói alapján képes adaptívan, a megfelelő feladattípusok és feladatnehézség megválasztásával elősegíteni. Jelen dolgozat arra a nyitott kérdésre koncentrálni, hogy hogyan lehetséges az IRT alapú pszichometriai modellek kiterjesztése olyan módon, hogy abban mind az alanyok képessége, mint az aktuális mentális állapotuk is befolyásolja egy adott feladat megoldásának valószínűségét.

Dolgozatomban ennek megfelelően először bemutatom az IRT típusú pszichometriai modellek alapjait, külön kitérve a modellek által használt alapfogalmakra, valamint az egyes IRT változatok által figyelembe vett paraméterekre, többek közt megemlítve a gyakorlatban kevésbé használt négyparaméteres IRT modellt, ami a kutatásom egyik alapját képezi. Ezután szólok a mentális állapot mérési lehetőségeiről, megemlítve a kutatás során használt eszközöket, azok jellemzőit, valamint a kutatásban betöltött szerepüket. Mivel a kutatás során ezen eszközökkel mérések elvégzése is szükséges volt, a dolgozat az ezek elvégzéséhez szükséges, nagy részben saját fejlesztésű kódkönyvtár rövid bemutatásával folytatódik.

Az ezt követő szakaszban mutatom be az általam kifejlesztett, újszerű modellt. Ennek során kitérek a mérési eredmények ismertetésére és az ezekből való következtetések levonására, hangsúlyt fektetve arra, hogy az egyes mérési eredmények hogyan illeszthetők ebbe a modellbe, amit a modell szimulációjával, a szimulációs eredményekkel megalapozva is bemutatok. A dolgozat zárásaként teljeskörűen értékelem a kialakított modellt, valamint javaslatot adok annak továbbfejlesztési lehetőségeire, illetve röviden írok arról is, hogy az hogyan alkalmazható a gyakorlatban, egy valós rendszer komponenseként.

## 2 IRT típusú pszichometriai modellek

### 2.1 Motiváció

A pszichometriai modellek legfőbb alkalmazási területe a többkérdéses tesztek, kérdőívek kifejlesztése és kiértékelése. Az ilyen tesztek célja a képesség, mint emberi tényező mérése valamilyen módon. Egy teszt tipikusan több kérdést, ún. teszt *itemet* (magyar fordításban néha: *tételt*) tartalmaz, amelyek különböző nehézségűek lehetnek. A cél a mért alany képességszintjének meghatározása annak ismeretében, hogy az mely és milyen nehézségű itemekre milyen minőségű választ adott.

Kutatásomban a tanulás vizsgálata, annak jobbá tétele, hatékonyságának növelése az elérendő cél. Ebben az esetben a képességre, mint fejlesztendő tényezőre tekintünk, viszont ilyenkor is fontos annak mérése, hogy a tanulónak a legmegfelelőbb, a képességeit leginkább fejleszteni képes feladatot tudjunk adni. Emiatt a pszichometriai modellek a tanulás során, többek között az előrehaladás mérésére is jól alkalmazhatók.

### 2.2 Item response theory

A pszichometriai modellek egy gyakran használt fajtája az *item response theory* (IRT) modellek családja [3][4][5][6]. A következő szakaszban az IRT modellek általános tulajdonságairól, fajtáiról, csoportosításáról, valamint azok egyéb jellemzőiről lesz szó.

Az item response theory név első fele abból ered, hogy az általa vizsgált többkérdéses tesztek elemeit általában *itemeknek* nevezik, a dolgozat további részében is ezt a kifejezést fogom használni. Az item kifejezés általános értelmű: egy teszt item sokféle dolgot takarhat: lehet egy feleletválasztós teszt egy kérdése, ahol a kitöltő személynek csak ki kell választania néhány válasz közül a helyeset, de lehet akár egy teljesen nyílt végű, szabad szöveggel megválaszolható kérdés is. Azt, hogy pontosan miről van szó egy konkrét esetben, az adott teszt jellege határozza meg.

A névben a *response*, azaz *válasz* szó arra utal, hogy az IRT modellek a vizsgált képességet az itemekre a személy által adott válaszok helyessége alapján mérik, mégpedig valószínűségi, matematikai megközelítéssel. Az IRT kulcsfogalma a *képesség* (*ability*), ami egy ún. *látens vonása* a tesztet kitöltő személyeknek. Ez azt jelenti, hogy a képesség nem mérhető közvetlenül, hanem a tesztre adott válaszok alapján számszerűsíthető. Az IRT különféle matematikai statisztikai eszközöket alkalmaz a



modellek becslésére és jellemzésére, ezért is nyújt jó elméleti háttérrel különböző mérőeszközök megtervezéséhez és elemzéséhez.

## 2.3 Alapfogalmak

Az IRT modellek alapvetően két dolgot vizsgálnak: magát a tesztet, illetve azt ezt felépítő egyes teszt itemeket, valamint a tesztet kitöltő személyeket. A modell tulajdonságai közül kiemelt fontosságú az ún. *személyi paraméterek* száma, amit a modell dimenziójának is szokás nevezni. A leggyakrabban használt IRT modellek csupán egyetlen személyi paramétert tartalmaznak, ez pedig a már fentebb is emlegetett látens vonás, a képesség. Ennek megfelelően ezeket a modelleket *egydimenziós* IRT modelleknek nevezzük. Több, egymástól független személyi paraméter esetén *többdimenziós* IRT modellekről beszélünk, azonban ezek a gyakorlatban kevésbé terjedtek el, rendszerint tehát az személyi vonás az adott személy képessége, ez határozza meg az adott személy a teszt itemekre adott válaszainak helyességét. A többdimenziós modellekben a választ egynél több látens vonás magyarázza, ezen modellek elemzése általában matematikailag bonyolultabb módszereket kíván, mint az egydimenziós modelleké.

Alapvető fontosságú a vizsgált képesség *homogenitása*. Ez azt jelenti, hogy az IRT modellek feltételezik, hogy a képesség nem bontható szét több részre. Ez úgy fogható meg, hogy az IRT alapú tesztekkel egyetlen, részekre nem osztható képességet vizsgálunk. A homogenitás vizsgálata rendszerint faktoranalízissel történik, ami egy matematikai statisztikai módszer. Megemlítendő az is, hogy a képesség megnevezés néhány esetben megtévesztő lehet, mivel bizonyos alkalmazásokban például inkább attitűd, vélemény mérése a cél. Ilyenkor gyakran használnak a személyi paraméter elnevezésére az alkalmazástól függő egyéb neveket. A dolgozat további részében a képesség megnevezést fogom használni.

Az IRT azzal az alapfeltevéssel él, hogy a képesség egy skálán mérhető. Tipikusan standard normális, azaz 0 várható értékű és 1 szórású skálát szokás használni, ami azt jelenti, hogy ebben az esetben a képességértékek döntő többsége -3 és 3 között lesz, bár elméletben bármilyen valós értéket felvehet.

Egy IRT modell másik nagyon fontos tulajdonsága az, hogy a modell hogyan jellemzi a teszt itemeket. Egy teszt esetén az item a legkisebb, önállóan is értékelhető, tovább már nem bontható feladategység. Az első, itemeket jellemző tényező az, hogy

hogyan osztályozzuk az arra adott válaszokat. Amennyiben az itemekre csak kétféle, helyes és helytelen válasz lehetséges, akkor az itemeket dichotómnak, az ilyen itemeket tartalmazó modellt pedig dichotóm modellnek nevezzük. Több, külön pontszámmal értékelt válasz esetén politóm itemekről, illetve modellről beszélünk [7]. A továbbiakban a dichotóm modelleket tekintjük át.

Az IRT a teszt itemekre vonatkozóan azzal a kikötéssel él, hogy azoknak lokálisan függetleneknek kell lenniük. Ez két dolgot jelent: elsőként azt, hogy egy adott item felhasználásának esélye nem függ attól, hogy azon kívül milyen más itemeket használunk fel. Ezen kívül a lokális függetlenséghez tartozik még az a követelmény is, hogy egy személy itemekre adott válaszait ne befolyásolják más személyek, azaz minden személy függetlenül döntsön az egyes itemek esetében, kizárva ezzel a párban történő-, vagy csoportmunkát.

Az itemek jellemzőit *item paramétereknek* nevezzük. Míg a személyi paraméterből tipikusan az egyetlen a képesség, item paraméterből általában több is van, a pontos szám a modell fajtájától függ, de általában – dichotóm modellek esetén – az 1-4 paraméteres modellek a leginkább használatosak a gyakorlatban. A legfontosabb, minden IRT modellben jelenlévő item paraméter az item *nehézsége* (*difficulty*). Ezt a paramétert mindig a vizsgált képességgel közös skálán mérjük, ez adja az IRT modellek egy különlegességét, nevezetesen azt, hogy egy személyhez kiválasztható a képességeivel „azonos” nehézségű item. Ez teszi lehetővé az IRT modellek adaptív környezetben [2] való használatát, ahol is a személynek legmegfelelőbbnek ítélt item éppen a képességeihez nehézségben legközelebb álló item lesz.

A két- vagy több paraméteres dichotóm IRT modellekben jelenlévő második item paraméter a *diszkrimináció*. Ez a paraméter szemléletesen az adott item megkülönböztető képességét írja le, és pozitív értékű. Minél magasabb ez az érték, annál inkább „szétválasztja” az item az alacsony és magas képességű egyéneket, míg alacsony érték esetén fokozatosan nő a helyes megoldás valószínűsége a képesség növekedésével.

A harmadik, három- vagy több paraméteres dichotóm modellekhez tartozó item paraméter a pseudo-találgatás (pseudoguessing). Ezzel a paraméterrel fogható meg a véletlen tippelés válaszdásra vonatkozó hatása. Ez egy 0 és 1 közötti szám lehet, viszont 0,5-nél magasabb érték esetén már bármilyen képességszint esetén 1/2-nél nagyobb valószínűséggel lehetne helyesen válaszolni, így a gyakorlatban ilyen nem fordul elő.

A negyedik item paraméter a négyparaméteres modellben használatos. Míg a pszeudo-találgatással egy alsó határt állítunk a helyes megoldás valószínűségének, a negyedik paraméter egy felső határt határoz meg. Eszerint nagyon magas (akár „végtelen”) képességszint esetén sem tart 1-hez az item helyes megoldásának valószínűsége: ezzel a véletlen hibázást kívánják modellezni. Érdeemes elgondolkodni azon, hogy mitől függ a véletlen hiba valószínűsége. Több független kutatócsoport is arra az eredményre jutott, hogy ebben nagy szerepet játszik a mért alany aktuális mentális állapota is, azaz, hogy valójában ez a tényező nem is az itemtől, hanem az adott megoldó személytől függ, nagyban befolyásolja annak figyelme, fáradtsága, pillanatnyi koncentrációja a feladatra. Ebben a témakörben mindenképp további tudományos vizsgálat szükséges, ezért kutatásom is többek között ennek vizsgálatára épül, a későbbi fejezetekben található ennek részletes kibontása, valamint a mérésekre alapuló részletes kutatási eredményeim a témában.

## 2.4 A 3PL és az itemválasz függvény

Minden IRT modellben közös, hogy definiál egy ún. *itemválasz függvényt* (*item response function*, IRF). Ez a matematikai függvény az adott teszt item és az itemet megoldó személy paraméterei függvényében adja meg az egyes válaszok (dichotóm modellek esetében a helyes válasz) adásának valószínűségét. Egydimenziós modellek esetén egy adott, rögzített itemre nézve a függvény egyetlen változója a személy képessége, az item paraméterek konstansként jelennek meg.

A legtöbb IRT modell esetén az IRF grafikus ábrázolása jellegzetes, szigmoid S-alakot vesz fel. A gyakorlatban legtöbbször használt ún. *logisztikus* IRT modellek nevüket onnan kapták, hogy a rájuk jellemző IRF egy módosított logisztikus függvény. Ezen kívül használják még a normális eloszlás eloszlásfüggvényét is, ami azonban nem adható meg zárt képlettel, emiatt történelmileg jobb teljesítményre voltak képesek a logisztikus függvény használatával, ma azonban a kettő között nincs számottevő különbség, sőt a két modellfajta egy konstans tényezőt leszámítva a gyakorlatban megfeleltethető egymásnak. A továbbiakban a gyakrabban használt logisztikus modellek tárgyalására szorítkozom.

Az egydimenziós, dichotóm, logisztikus IRT modellek közül az egyik legkifinomultabb, legbővebb a *háromparaméteres logisztikus modell* (*3-parameter logistic model*, 3PL). A 3PL modell a fent tárgyalt item paraméterek közül az első hármat,

azaz a nehézséget, a diszkriminációt és a pszeudo-találgatást használja. Az itemválasz függvény képlete ekkor a következő egy rögzített  $i$  item esetén:

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Az egyes betűk jelentése a következő:

- $\theta$  – a személy képessége,
- $a_i$  – az  $i$  itemre jellemző diszkrimináció,
- $b_i$  – az  $i$  item nehézsége,
- $c_i$  – az  $i$  itemre jellemző pszeudo-találgatás érték.

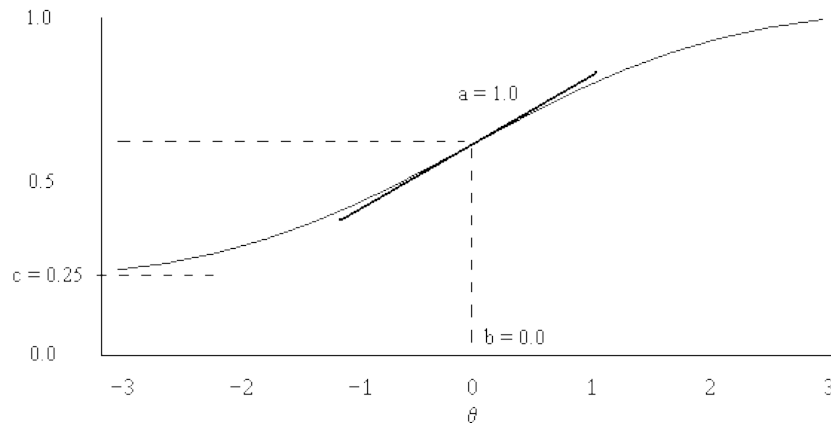
Az IRF grafikus megjelenítését *item karakterisztikus görbének* (*item characteristic curve*, ICC) nevezzük. Mint már említettem, logisztikus modellek esetén ennek az alakja szigmoid, ami a használt módosított logisztikus függvény képletéből következik. Az egyes item paraméterek határozzák meg a görbe jellegét. Az item nehézsége az item helyét adja meg a vizsgált képesség kontinuumán. Eszerint egy  $b_i$  nehézségű itemnél pontosan a  $b_i$  helyen lesz az ICC középpontja, azaz az elméletileg lehetséges és minimális és maximális érték számtani közepének megfelelő értéket ezen a helyen fogja felvenni a függvény.

Emlékeztetőül, a második paraméter, a diszkrimináció azt a mértéket jelenti, ahogyan az adott item megkülönbözteti az alacsony képességgel rendelkező személyeket a magas képességgel rendelkezőktől. Matematikailag ez az ICC érintőjének meredekségét jelenti abban a pontban, ahol ez a meredekség maximális, azaz pontosan az előbb tárgyalt, az item nehézsége által megadott középpontban. Egy magas diszkrimináció értékkel rendelkező item esetén egy alacsony képességű személynek jóval kisebb valószínűsége van a helyes válasz adására, mint egy magas képességűnek, míg, ha ez az érték alacsony, akkor a görbe laposabb lesz, tehát ekkor a helyes megoldás valószínűsége egyenletesebben növekszik a képességszint növekedésével.

A 3PL harmadik paramétere, a pszeudo-találgatás adja meg az ICC alsó aszimptotáját, azaz ennél kisebb értéket a függvény nem képes felvenni. Ennek a paraméternek legtöbbször feleletválasztós tesztek esetén van nagy jelentősége. Példaként egy olyan kérdést véve, ahol négy lehetséges válasz van, 25% eséllyel olyan valaki is eltalálja a helyes megoldást, akinek fogalma sincs a kérdésben feltettekről. Amennyiben

egy ilyen kérdésnél egy ilyen esetben minden válasz adásának valószínűsége közel azonos, a  $c_i$  paraméter értéke megközelítőleg 0,25 lesz, azonban, ha ez nem így van, például az egyik válasz nagyon nyilvánvalóan helytelen, akkor a paraméter értéke ettől szignifikánsan eltérhet, ezért használatos a pszeudo- előtag a paraméter nevében.

Az alábbi ábrán egy 3PL egy itemjének ICC-je látható:



**1. Ábra: Példa egy háromparaméteres logisztikus IRT modell itemválasz függvényének grafikonjára. Az ábrán jól látszik, hogy az egyes paraméterek a görbe mely jellemzőit befolyásolják.**

## 2.5 Egyéb IRT modellek

Amint arra fentebbi szakaszban is utaltam, a 3PL-en kívül léteznek egyéb hasonló, egydimenziós, dichotóm itemeket használó, logisztikus modellek is. A *kétparaméteres logisztikus modell* (*2-parameter logistic model*, 2PL) nem használja (matematikailag minden item esetén 0-nak tekinti) a pszeudo-találgatás paramétert [9], míg az *egyparaméteres logisztikus modell* (*1-parameter logistic model*, 1PL) a diszkriminációt sem, azaz az itemeket kizárólag a nehézségükkel jellemzi, a diszkrimináció értéke minden item esetén egyenlő. Ezt a modellt megalkotója, Georg Rasch után gyakran nevezik *Rasch-modellnek* is. [10]

A *négyparaméteres logisztikus modellben* (*4-parameter logistic model*, 4PL) jelenik meg a fentebb már említett negyedik item paraméter. Matematikailag ez a pszeudo-találgatáshoz nagyon hasonlóan jelenik meg, viszont nem alsó, hanem felső aszimptotát ad a 4PL IRF-jének. Ebben az esetben az IRF képlete a következő:

$$p_i(\theta) = c_i + \frac{d_i - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

A képletben  $d_i$ -vel jelöltem az új, negyedik paramétert. Látszik, hogy ahogyan a két- illetve egyparaméteres logisztikus modellek tekinthetők a 3PL speciális eseteinek, úgy a 3PL is valójában megfelel egy olyan 4PL modellnek, ahol a negyedik paraméter értéke minden item esetén 1.

Nyilvánvaló, hogy minél több paramétert használunk, annál jobban tudjuk közelíteni a valóságot a modellünkkel, azonban a megnövekedett hűség ára a magasabb matematikai komplexitás lesz. A minél több paraméteres modellt használók mellett érvelnek, hogy egy teszt fejlesztésekor jobb, ha a modellt igazítjuk a valósághoz, azaz figyelembe kell venni a valós teszt itemek minél több sajátosságát. Ebben az esetben a cél, hogy a modellünk minél több teszt itemhez illeszkedjen, azaz minél kevesebb itemet kelljen a nem megfelelő illeszkedés miatt „kidobni”. A Rasch-féle paradigma kedvelői viszont az egyszerűsége helyezik a hangsúlyt, és inkább az adott teszt itemeit igazítják a modellhez. Ez azt is jelenti, hogy egy egyparaméteres esetben az itemek modellhez illesztése jelentős kompromisszumokat kíván, egyes itemeket egyszerűen el kell vetni a nem megfelelő illeszkedés miatt, itt tehát a hangsúly a modellhez megfelelően illeszkedő itemek keresésén van. Emiatt a Rasch-modellt gyakran megkülönböztetik az egyéb IRT modellektől, egyes szerzők a fenti szemléletbeli különbségek miatt nem is tekintik valódi IRT modellnek.

## 2.6 IRT modellek paramétereinek becslése

Az IRT modellek hagyományos alkalmazásait tekintve, a gyakorlatban az item- és személyi paraméterek is ismeretlenek a modell specifikációjakor. Emiatt a specifikáció szerves részét képezi az item paraméterek megbecslése, kalibrálása. A tipikus esetben egy véletlen, ún. kalibrációs mintát választanak ki a célpopuláció egyedeiből, majd egyszerre, szimultán módon végzik az item és személyi paraméterek becslését.

A becsléshez, mivel a modell matematikai, természetesen matematikai eljárások használhatóak, több különféle módszer is született erre a célra. Minden esetben igaz azonban, hogy sem az item, sem pedig a személyi paraméterek nem megfigyelhetőek. Emiatt egyfajta nondeterminizmus lép fel, amit formálisan *identifikációs problémának* szokás nevezni. Leegyszerűsítve, ez azt jelenti, hogy az itemválasz függvény invariáns marad bizonyos függvénytranszformációkra. Ebben az esetben a megoldás az, hogy vagy a személyek képességét, vagy az itemek nehézségét kalibrálnunk kell. Egyparaméteres

modell esetén elég, ha a minta átlagát 0-ra igazítjuk, két- vagy többparaméteres modell esetén fontos a minta szórásának 1-re igazítása is.

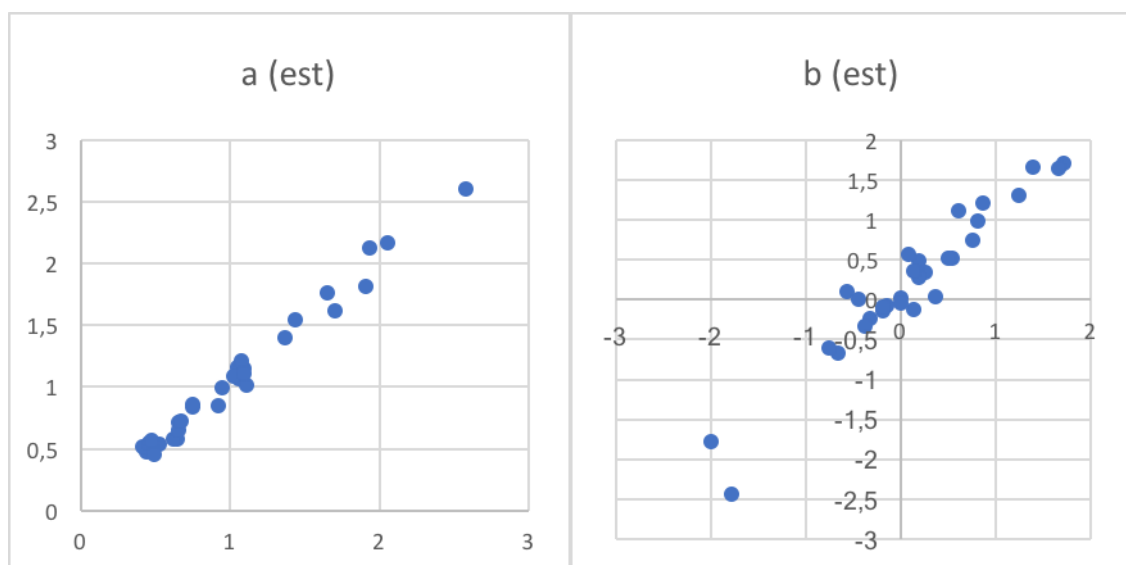
Tegyük fel, hogy  $N$  darab személy old meg egy  $n$  darab itemből álló tesztet. Ebben az esetben a megbecslendő paraméterek száma egyparaméteres modell esetén  $N + n$ , kétparaméteres modell esetén  $N + 2n$ , háromparaméteres modell esetén pedig  $N + 3n$ . A fenti megfontolásokat is figyelembe véve a becslendő paraméterek száma az egyparaméteres modell esetén 1-gyel, a két- és háromparaméteres modell esetén 2-vel csökken. Látható, hogy ellentétben a hagyományos statisztikai modellekkel, ahol a paraméterek száma független a megfigyelések számától, itt a paraméterek száma a tesztet megoldók számával arányosan nő.

A szimultán paraméterbecsléshez használt első módszer a *joint maximum likelihood becslés* (joint maximum likelihood estimation, JMLE) viseli. Ez egy matematikai statisztikai módszer, amely az item és személyi paramétereke együttesen maximalizálja az ún. likelihood függvényt, amit  $L(\mathbf{u} | \boldsymbol{\theta}, \mathbf{b}, \mathbf{a}, \mathbf{c})$ -vel jelölnék. Az  $\mathbf{u}$  egy  $N$ -szer  $n$  dimenziós vektor, ami az  $N$  kitöltő személy  $n$  darab itemre adott válaszait tartalmazza, 0-val jelölve a helytelen, 1-gyel pedig a helyes választ. A megbecslendő vektorok közül a  $\boldsymbol{\theta}$  egy  $N$  dimenziós vektor, ami a személyek képességparamétereit tartalmazza, míg a  $\mathbf{b}$ ,  $\mathbf{a}$ , és  $\mathbf{c}$   $n$  dimenziós vektorok az itemek nehézség, diszkrimináció és pszeudo-találgatás paramétereit tartalmazzák. A likelihood függvény helyett, ami (az egyes itemek illetve személyek paramétereinek lokális függetlensége miatti) szorzatot tartalmaz, szokásosan az összeget tartalmazó ún. log-likelihood függvényt szokták maximalizálni, ami a likelihood függvény természetes alapú logaritmus, ám a logaritmusfüggvény szigorú monotonitása miatt maximumhelyei megegyeznek a likelihood függvény maximumhelyeivel. A maximumhelyeket rendszerint a log-likelihood függvény deriválásával, majd pedig a derivált függvény gyökeinek megkeresésével oldják meg. Mivel ebben az esetben ez egy nemlineáris, ezáltal analitikusan nem megoldható egyenletrendszerhez vezet, a megoldásban a numerikus módszerek elterjedtek, ezek közül gyakran használt az iteratív Newton-Raphson módszer többváltozós verziója.

Egyparaméteres modell esetén használható módszer a *feltételes maximum likelihood becslés* (conditional maximum likelihood estimation, CMLE), mivel ekkor elérhető megfelelő előzetes statisztika a tesztet kitöltő alanyok személyi paraméteréről, hiszen ekkor a teszten elért eredmény (pontszám) ilyen. Többparaméteres modellek

esetén alternatíva a *marginális maximum likelihood becslés* (*marginal maximum likelihood estimation*, MMLE) módszere. [11] Ebben az esetben a becslés előtt a likelihood függvényt a személyi paraméterekre vonatkozóan kiátlagoljuk, emiatt az az  $L(\mathbf{u} \mid \mathbf{a}, \mathbf{b}, \mathbf{c})$  formát veszi fel. Amennyiben van előzetes (a priori) informáciánk, bayesi módszerek is használhatóak, máskülönben azok megoldása tulajdonképpen a fenti maximum likelihood módszerek által adott egyenletrendszer megoldásává fajulnak.

A következő ábrán látható egy szimulációs kísérlet eredménye, amit a marginális maximum likelihood becslés módszere egy implementációjának kipróbálásával végeztem. Összesen 30 véletlenszerűen generált itemet és 100 véletlenszerűen generált szimulált „személyt” használtam egy kétparaméteres logisztikus modellben, ahol az ágensek képességének eloszlása standard normális, a teszt itemek nehézségének eloszlása szintén standard normális, diszkriminációjuk eloszlása pedig log-normális volt. Az ágensek az itemeket a szimulációs algoritmus szerint a modell képletének megfelelő valószínűséggel oldották meg, majd az így keletkezett adatra használtam a marginális maximum likelihood módszert megoldó algoritmust. Az ábrán látható, hogy a becült értékek jól konvergálnak a valódi értékekhez.



**2. Ábra:** A grafikonokon a pontok az egyes teszt itemeket jelzik, a grafikonokon az algoritmus által becült érték látható a valódi érték függvényében. A baloldali ábra a diszkriminációs (a) paramétert, a jobboldali ábra a nehézségi (b) paramétert ábrázolja.

Amennyiben a kalibráció már megtörtént, és az itemek paramétereinek értéke már rendelkezésre áll, azokat a további alkalmazás szempontjából már ismertként tartjuk számon. A kalibrált itemekből item bankok készíthetők, amikből megfelelő



megfontolásokkal tesztek állíthatók össze. Ezután a tesztet megoldó személyek képességeinek megállapítása a cél. Itt már az egyszerű *maximum likelihood módszer* (*maximum likelihood estimation*, MLE) a leginkább alkalmazott a gyakorlatban. Az egyszerű maximum likelihood becslés esetén már könnyebb dolgunk van, mert fennállnak olyan kedvező tulajdonságok, mint az aszimptotikus normalitás, vagy a konzisztencia. A számított képesség hibájára is becslést kaphatunk az ún. információs függvény segítségével. Amennyiben egy személy zérus vagy teljes pontszámot ér el a teszten, a maximum likelihood becslés nem képes helyes megállapításokra jutni, ilyenkor a bayesi módszerek segíthetnek, viszont ezek használatához előzetes információkra is szükségünk van az alanyok képességeit illetően. A bayesi módszer további előnye, hogy a standard hiba nagysága itt alacsonyabb, mint a maximum likelihood módszer használata esetében.

## **2.7 Az IRT alkalmazása**

Az IRT legfontosabb alkalmazási területe természetesen a különféle tesztek, kérdőívek összeállítása, kalibrációja és elemzése, mivel kitűnően használható mérőeszközök fejlesztéséhez. Emellett az IRT-nek fontos szerep jut az oktatásban az általános teszteken túl is. A *számítógépes adaptív tesztelés* (*computerized adaptive testing*, CAT) módszere pszichometriai modellekre, nagyrészt az IRT-re támaszkodva éri el, hogy egy számítógépen végrehajtott teszt esetén a kérdések adaptívan illeszkedjenek a tesztet kitöltő személy képességeihez. [12] Ehhez egy már kalibrált item bankból indul ki, és a folyamatosan beérkező válaszokból következtet egyre pontosabban a válaszadó képességszintjére. Mivel az elején nincs információnk a személyről, az első kérdés rendszerint átlagos nehézségű, a további kérdések pedig az előző válaszok alapján kerülnek kiválasztásra, mindig az éppen becsült képességnek megfelelően, tehát rossz válasz esetén a válaszadó valószínűleg könnyebb, jó válasz esetén valószínűleg nehezebb kérdést kap. Egy ilyen tesztnek természetesen azonnal kiértékelhetőnek kell lennie, valamint a közben adaptívan változó nehézség miatt nem megengedhető, hogy a válaszadó megváltoztasson egy korábbi választ, máskülönben a módszert „kijátszva” sorozatos rossz válaszokkal elérheti, hogy a legkönnyebb kérdéseket tegye fel neki a rendszer, majd visszamehet és megváltoztathatja az adott rossz válaszait. A CAT alkalmazása mérések szerint akár 50%-kal megrövidítheti a teszt idejét, valamint pontosabb eredményt is képes elérni, mint a fix kérdésekkel operáló tesztek.

Kutatásom távlati céljaként elsődlegesen nem mérni, hanem inkább fejleszteni szeretném a képességet, viszont ehhez a pontos mérés elvégzése is fontos lépés. Ezen különbségtől eltekintve azonban feltevésem szerint ekkor is alkalmazható a számítógépes adaptív teszteléshez hasonló módszer. Ebben az esetben az adaptivitás megjelenése azt az értelmet hordozza magával, hogy a tanulás során a tanulók minél jobb, pozitívabb élményben, sikerélményekben részesüljenek, hiszen így érhetik el a Csíkszentmihályi Mihály által leírt ún. flow állapotot, amelyben kutatások szerint a teljesítményük, és ezáltal a tanulásuk határfoka is a legjobb. [13]

## **2.8 A hagyományos IRT modellek korlátai**

Mint már a dolgozat során többször felmerült, az IRT modellek általában a személyek képességét tekintik a teszt itemek megoldását magyarázó tényezőnek. Ezt a képességet az IRT-féle megközelítés a teszt során állandónak tekinti, hiszen a cél pontosan az erről való tájékozódás. Azonban egy teszt item sikeres vagy sikertelen megoldása mögött pillanatnyilag fennálló tényezők is lehetnek. A fentebb említett CAT esetén különösen releváns ez, főleg abban az esetben, amikor a tesztelt személynek adott ideje van csak egy-egy kérdés megválaszolására, valamint nincs lehetősége a korábban megadott válaszában módosítani.

A mögöttes pillanatnyi tényezők közül kiemelkedik a személy mentális, idegrendszeri állapota, amibe akár a külső zavaró tényezők is belefoglalhatók, hiszen ezek is jól meghatározható agyi aktivitást váltanak ki, ami a megfelelő eszközök segítségével jól mérhető. Amint láthattuk, az IRT modellek közül egyedül a négyparaméteres modell próbálta figyelembe venni annak lehetőségét, hogy egy adott feladat megoldásához elvileg több mint elégséges képességszinttel rendelkező személy is hibázhat bizonyos valószínűséggel, viszont ezt a valószínűséget feladatonként állandónak tekintette, holott ez inkább függ a vizsgált személytől, mintsem a megoldott feladattól. Erre a gondolatra építve először érdemes megvizsgálni, hogyan mérhető egyszerűen az ember mentális állapota az IRT modelleknek megfelelő tesztek, feladatok megoldása közben.

## 3 Mentális állapot mérése fiziológiai tények segítségével

### 3.1 Biofeedback

Az előző fejezet tárgyalta az IRT modellek témakörét, egyben rámutatott azok fő hiányosságára is, nevezetesen a mentális állapot figyelembevételének mellőzésére. Ebben a fejezetben bemutatom, hogy néhány egyszerű, költséghatékonyan beszerezhető, könnyen hordozható mérőeszköz segítségével hogyan következtethetünk a használó mentális állapotára. A fiziológiai tények mérésére képes eszközöket összefoglaló néven biofeedback szenzoroknak fogom nevezni a dolgozatban. Ahogyan az elnevezés is sugallja, ezek a műszerek visszacsatolást adnak az általuk szolgáltatott adatok segítségével.

Napjainkban számos kutatás folyik az *ember-számítógép interakció* (*human-computer interaction*, HCI) területén. [23] Ezen kutatások célja, hogy minél könnyebbé, direkttebbé tegyék a számítógépeket (ide értve a hagyományos PC-ken és a modern mobil eszközöket is) használó ember kognitív állapotának mérését, géppel megkülönböztethetővé tegyenek többek között olyan alapvető mentális állapotokat, mint a figyelem, a nyugalom, az éberség, a frusztráció vagy a koncentráció. Ehhez a kutatók számos technikát és algoritmust fejlesztettek ki, amelyek képesek a szenzorok adatainak értelmezésére és belőlük a mentális állapotra vonatkozó információ kinyerésére. Gyakran már az egyes szenzorok szoftverrendszerbe való illesztéséhez szükséges programkönyvtárak tartalmazznak ilyen algoritmusokat, jelentősen megkönnyítve az ezekkel való munkát.

A biofeedback szenzorok által mért jelek közül vizsgálataim az EEG (*elektorenkefalográfia*) [1][21] és EKG (*elektrokardiográfia*) [22] jelekre terjedtek ki. Megjegyzendő még, hogy a mentális állapotra következtethetünk még számos egyéb adatból is, ilyenek például a szemmozgás követése és a pupilla tágulatának mérése [16], azonban ezekhez speciális mérési körülmények szükségesek, míg az EEG és EKG jelek mérésére ma már rendelkezésre állnak olyan viselhető eszközök, amikkel a megfelelő szabályok betartása esetén bárki képes akár önállóan méréseket végezni. A kutatásom során végzett eszközök mind ilyenek voltak, ezáltal hordozhatóvá és könnyen újra összeállíthatóvá téve a mérési környezetet. Ez azt is lehetővé tette, hogy a mérések a személyeknek leginkább megfelelő, természetes környezetben történtek, így a mérés

során folytatott tevékenységük is szokásaik jól megfeleltek a normál környezetben tapasztalhatónak.

A következő szakaszokban áttekintem a mérések során használt egyes eszköztípusok által szolgáltatott jelek mögötti elméleti hátteret, valamint az egyes használt konkrét eszközök funkcióit, lehetőségeit.

## 3.2 EEG

### 3.2.1 Elméleti alapok

Az EEG rövidítéssel illetjük mind az agyi elektromos aktivitás mérésére szolgáló pszichofiziológiai mérőeljárást, mind az ehhez szükséges mérőeszközt. A 10-100  $\mu\text{V}$  közötti feszültségű EEG jelek segítségével vizsgálható a központi idegrendszeri aktivitás élettani háttere. Ezt a módszert a neuronok által kiváltott elektromos változások valós idejű monitorozásával éri el. Az EEG által szolgáltatott jel egy komplex, több komponensű periodikus görbe, az ún. *elektroencefalogram*. Az EEG eljárás tipikusan nem invazív, azaz az EEG eszköz elektródái egyszerűen, behatolásmentesen a fejre helyezhetők.

Az EEG használata az utóbbi években rendkívül elterjedt az ember-számítógép interakciót kutatók körében. Az EEG jelek frekvenciája legtöbbször az 1-20 Hz közötti tartományba esik, a legtöbb kutatásban bizonyos frekvenciatartományok (delta, theta, alfa, béta, gamma) vizsgálata a meghatározó. Az egyes tartományok a következő táblázatban láthatók.

Név	Frekvenciatartomány
Delta hullámok	< 4 Hz
Théta hullámok	4–7 Hz
Alfa hullámok	7,5–12,5 Hz
Béta hullámok	12,5–30 Hz
alacsony béta	12,5–16 Hz
közép béta	16,5–20 Hz

magas béta	20,5–28 Hz
Gamma hullámok	> 30 Hz

**1. Táblázat: Az EEG jelek besorolása frekvencia alapján.**

A delta hullámok felnőtteknél mély alvás közben, a REM fázist kivéve fordulnak elő, míg a théta hullámok félálomban jelentkeznek, de éber állapotban is jelezhetik az álmoságot, illetve meditáció során is előfordulnak. Az alfa hullámok éber állapotban a szem becsukásakor, illetve relaxáció alatt erősödnek fel, a szem kinyitásával, valamint mentális erőfeszítés közben pedig csillapodnak, így ezek a hullámok már a kognitív teljesítményről is szolgáltatnak információt.

A béta hullámok ébrenléti állapotban, nyitott szem mellett jelennek meg, amplitúdójukból az aktuális mentális terhelés mértékére következtethetünk. A gamma hullámok szerepe máig nem tisztázott, de számos kutató véleménye szerint jelenlétük a tudatos észlelésre, valamint az információrészek aktív összekötésére utal. Összességében elmondható, hogy az ébrenléti mentális állapotra való következtetésre leginkább az alfa, béta és gamma hullámok alkalmasak.

### **3.2.2 Felhasznált eszközök**

#### **3.2.2.1 NeuroSky MindWave Mobile**

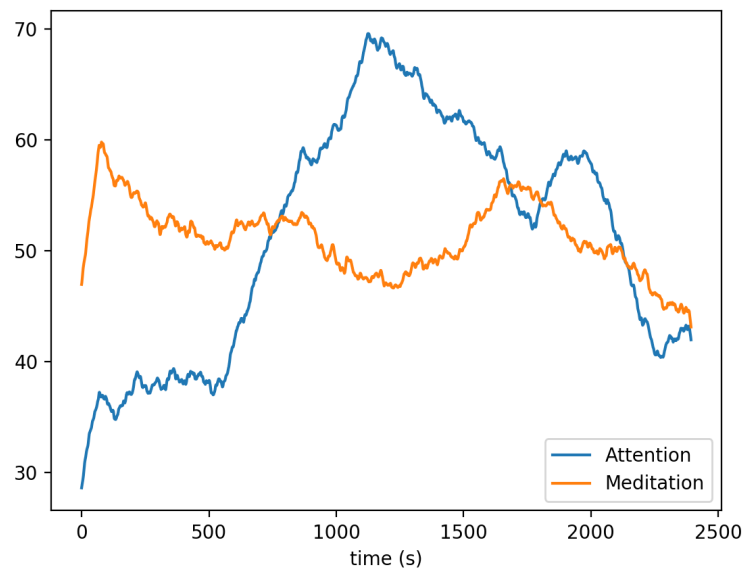
A piacon található mobil EEG headsetek közül a *NeuroSky* cég által gyártott *MindWave Mobile* az egyik legolcsóbb, mérésre és fejlesztésre használható készülék. [34] Mivel a kutatás során fontos volt a könnyen hordozható mérőkörnyezet kialakítása, ezért nagy pozitívum, hogy az eszköz nagyon könnyű, mindössze 90 gramm tömegű, illetve képes mobil operációs rendszereket futtató telefonokkal, tablettel való kommunikációra is. A kutatás során a gyártó által Android operációs rendszerre készített SDK-t használtam, aminek segítségével tetszőleges Android alkalmazásban lekérdezhetők, feldolgozhatók és megjeleníthetők a headset által küldött adatok.

A MindWave eszköz egy headsetből, egy szenzorfejből és egy fülre rögzíthető csipeszből áll, és egyszerű AAA elemekkel üzemeltethető akár 8 órán át. Az eszköz 512 Hz-es mintavételi frekvenciával képes a 3–100 Hz frekvenciájú agyhullámok mérésére, 12 biten való megjelenítésére, valamint a jel spektrumának felbontására és kiadására, emellett a gyártó által adott speciális, *eSense* névre keresztelt algoritmussal származtatott

értékeket is nyújt a használó tevékenységéről. Az eszköz által számolt, és a mérések során általam is használt értékek a *figyelem (attention)* és a *nyugalom (meditation)*. Mindkettő hasznos információval szolgálhat az alany kognitív teljesítményével kapcsolatban. A számolt adatok küldésén kívül további funkciók a pislogás automatikus felismerése, valamint a küldött jelek minőségének elemzése, aminek segítségével könnyedén észrevehető és korrigálható a fejre való nem megfelelő illeszkedés. A készülék az adatokat hagyományos Bluetooth és a kisebb energiafogyasztású Bluetooth Low Energy kapcsolaton keresztül is képes továbbítani a fogadó eszköz felé.



**3. Ábra: A NeuroSky MindWave Mobile EEG headset.**



**4. Ábra: Példa a NeuroSky MindWave Mobile eszköz által küldött figyelem (attention) és nyugalom (meditation) értékekre egy mérés során.**

### 3.2.2.2 Emotiv Insight

A mérések során használt másik EEG eszköz az *Emotiv* cég *Insight* nevű terméke. [35] Ez a készülék is szintén kedvező árú, kereskedelmi forgalomban is elérhető megoldást nyújt az agy-számítógép kutatók számára. A kis tömeg és könnyű hordozhatóság erre az eszközre is igaz, csakúgy, mint a magas szintű, sok eszközzel és interfésszel való kompatibilitás.

Az Emotiv Insight egy 5+2 csatornás, szintén fejre helyezhető készülék beépített, minimum 4 órán át folyamatosan üzemeltethető tölthető akkumulátorral. Öt darab EEG és két darab referencia szenzorfejjel rendelkezik, amik a zajcsökkentésben segítenek. Másodpercenként és csatornánként 128 darab minta vételére képes megfelelő körülmények között az 1–43 Hz közötti frekvenciájú agyhullámokból. Vezetékkel és Bluetooth Low Energy típusú vezeték nélküli kapcsolaton keresztül is csatlakoztatható a küldött adatokat fogadó eszközhöz, ami lehet akár Windows, Linux vagy macOS rendszert futtató személyi számítógép, akár pedig iOS vagy Android rendszerű mobil készülék. A mérések során ezen eszköz esetében is a gyártó által Android rendszerre készített programkönyvtárat használtam.

Az eszköz használata mellett komoly érv, hogy a NeuroSky MindWave Mobile által támogatott két származtatott érték mellett négy további számolt teljesítménymetrika küldését támogatja. Ezekkel az értékekkel még pontosabb kép kapható az eszközt használó személy mentális teljesítményéről. A támogatott metrikák:

- *érdeklődés (interest)*,
- *izgatottság (excitement)*,
- *elmélyültség–unalom (engagement–boredom)*,
- *összpontosítás (focus)*,
- *feszültség (stress)*,
- *kikapcsolódás (relaxation)*.

Ezek közül az összpontosítás nevű metrika felel meg a NeuroSky MindWave Mobile által mért figyelem szintjének, míg a kikapcsolódás nevű metrika a MindWave által továbbított nyugalom szinttel állítható párba. A másik eszközhöz hasonlóan ez az eszköz is képes a megfelelő érintkezés automatikus, valós idejű érzékelésére, emellett pedig tartalmaz egy kilenc tengelyes, multifunkciós szenzort, ami magában foglal egy

giroszkópot, egy gyorsulásérzékelőt és egy, a mágneses tér érzékelésére képes érzékelőt is, amelyek akár külön-külön is ki-be kapcsolhatók. A további funkciók közé tartozik a különböző arc kifejezések (bal-, jobb, illetve mindkét szem lehunyása, szemöldök ráncolása és felemelése, mosoly, fogak összeszorítása) érzékelése, valamint sokféle „mentális parancsra” való betaníthatóság, amely funkció segítségével akár pusztán agyhullámokkal irányítható szoftveres alkalmazások fejlesztése is lehetségessé válik.



5. Ábra: Az Emotiv Insight EEG eszköz.

## 3.3 Szívritmus

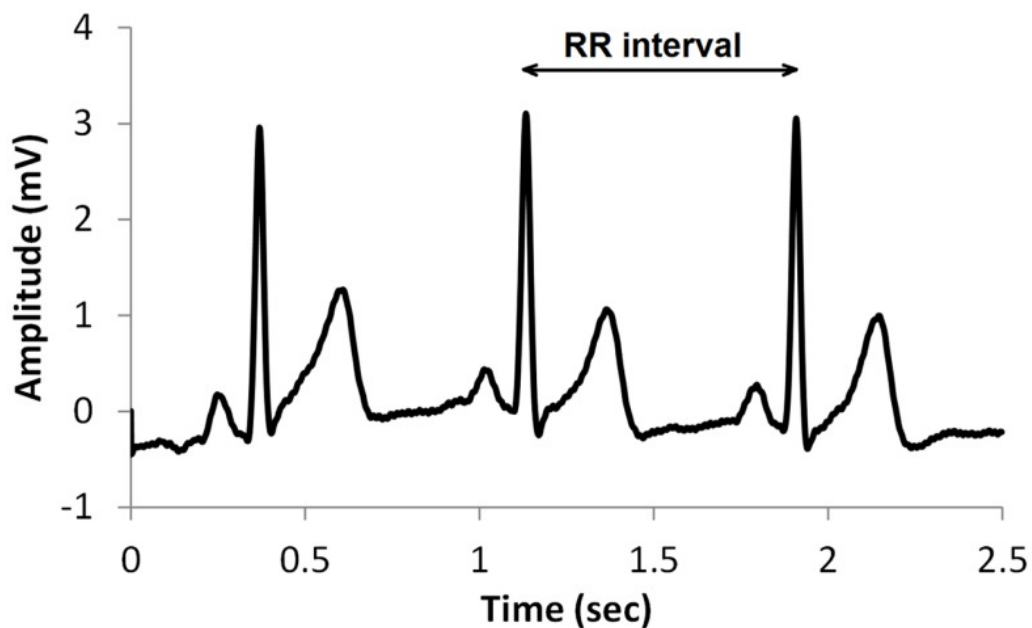
### 3.3.1 Elméleti alapok

A piacon sok készülék érhető el, amelyek segítségével az ember szívritmusa könnyedén mérhető. Az ilyen készülékek által rögzített jelek az EKG hullámok. A mérés metodikája azon alapszik, hogy a szív összehúzódása a szinuszcsomó által keltett elektromos impulzus következtében jön létre, ezen impulzus pedig jól detektálható a testre helyezett elektródák segítségével.

A nyers pulzusszámnak, azaz a percenkénti szívverések számának rögzítése mellett az EKG hullámokból kinyerhető a *szívfrekvencia-variabilitás* (*heart rate variability*, HRV), ebből pedig a *szívdobbanások között eltelt idő variabilitása* (*heart period variability*) is. [15] A HRV és a HPV az egyszerű szívritmus értékénél jobban mutatja az emberben keletkező idegességet, feszültséget, mivel ezek a metrikák alkalmasak a szív különböző külső ingerekre történő, ún. R–R távolságok



(szívdobbanástól szívdobbanásig eltelt idő) változása által mutatott reakciójának leírására.



6. Ábra: Egy tipikus EKG jel, jelölve az R–R távolságot is.

### 3.3.2 A felhasznált eszköz

#### 3.3.2.1 Zephyr HxM BT Heart Rate Monitor

A kutatás során a mérésekhez a *Zephyr* cég *HxM BT Heart Rate Monitor* nevű eszközét használtam. [36] Mint ahogyan az EEG eszközök esetében, itt is fontos volt a könnyű használhatóság, azaz, hogy a használót a lehető legkevésbé zavarja a mérőeszköz jelenléte. Az eszközhöz egy egyszerű, mellkasra felhelyezhető pánt tartozik, aminek segítségével könnyen és akadálymentesen rögzíthető a készülék, anélkül, hogy az lecsúszna, ezáltal a mérés során végig biztosítható a megfelelő mértékű illeszkedés.

Az eszközhöz a gyártó mellékel egy Java alapú SDK-t is, amely segítségével gyakorlatilag minden platformon, így PC-n és mobileszközökön, kiemelten a méréskor használt Androidon is véghez vihető a készülék illesztése és az általa küldött adatok kinyerése.

A mérés során az eszköz által küldött nyers szívritmus adatokból az R–R intervallumok meghatározása, majd ebből a HRV, majd pedig a HPV értékek kiszámítása történt, ezután ezekből az értékekből következtettünk a felhasználó mentális állapotára, pl. idegességének/nyugodtságának mértékére.



7. Ábra: A Zephyr HxM BT Heart Rate Monitor eszköz.

### 3.4 Összefoglalás

Napjainkra sokféle biofeedback eszköz érhető el kereskedelmi forgalomban is, könnyítve ezzel a kutatók hozzáférését és bevonását az ember-számítógép interfész témájú projektekbe. A fejezetben az általam a kutatás során használt eszközök fajtáinak, illetve konkrét típusainak bemutatására szorítkoztam, csak említve az esetleges egyes egyéb lehetőségeket. Természetesen ezen eszközök pontossága nem éri el az orvosi alkalmazásokban használt típusokét, de egyrészt a biofeedback kutatásokhoz már most is megfelelnek, másrészt pedig az a tény, hogy olcsón beszerezhetőek, alkalmassá teszi őket akár csoportos, tantermi használatra, ahol is a tanulók által használt szoftverek egy biofeedback alapú rendszerrel vannak összekötve, ami képes manipulálni pl. egy oktatójáték nehézségét mind a játékos konkrét teljesítményének, mind pedig a fiziológiai tényekből következtetett mentális állapotának megfelelően. A témában tanszékünkön jelenleg is intenzív, interdiszciplináris kutatás folyik, melynek része az én kutatásom is.

A fent részletezett eszközökkel való mérések elvégzéséhez egy kódkönyvtár összeállítása is szükséges volt, amelynek funkciója a mérési eredmények rögzítése és vizualizációja későbbi elemzés céljából. A következő fejezet ezt mutatja be.

## 4 Mérési kódkönyvtár

### 4.1 Motiváció

Az előző fejezetben szereplő szenzorok mind rendelkeznek saját szoftvercsomaggal, aminek segítségével kiolvashatók az általuk küldött adatok, viszont minden eszköz más és más interfészt kínál ehhez. Mivel egy-egy mérés során a gyakorlatban egyszerre több eszközt is használtam, célszerű volt egy közös rendszerbe összefoglalni és abban tárolni a mérési eredményeket. Mint már említettem, a hordozható, mobil mérőkörnyezet kialakítása a kezdetektől fő szempont volt, emiatt kézenfekvő volt a szegmensben piacvezető Android rendszerre építeni a kialakítandó környezetet. Az Android további előnye, hogy széleskörű támogatást nyújt a *folyamatok közötti kommunikációhoz* (*inter-process communication*, IPC), ezáltal egy lazán csatolt, jól skálázható keretrendszer hozható létre a platformon, amihez az alkalmazások egy egyszerű interfészen keresztül csatlakozni tudnak.

### 4.2 Előzmények

Tanszékemen, a Budapesti Műszaki és Gazdaságtudományi Egyetem Automatizálási és Alkalmazott Informatikai Tanszékén fejlesztés alatt áll az AdaptEd keretrendszer, aminek célja, hogy mobil oktatójátékokat biofeedback támogatás segítségével adaptívvá, a felhasználók képességeinek és állapotától függő nehézségűvé tegyen. [17][18][19] Mivel a keretrendszert már viszonylag rég óta fejlesztik hallgatók, a kód az évek során rendkívül nehezen módosíthatóvá, törékennyé vált. Emiatt néhány társammal (Pomázi Krisztián, Gazdi László, Bodolai Dorottya, ma a tanszék hallgatói vagy doktoranduszai) úgy döntöttünk, az alapoktól kezdve gondoljuk, tervezzük és implementáljuk újra ezt a rendszert a clean coding elveinek megfelelően. [24] Mindenek előtt a cél könnyű tesztelhetőség és olvashatóság, valamint a rugalmas kialakítás, a jó karbantarthatóság volt, így adta magát, hogy az én kutatásomhoz kapcsolódó mérések esetén is ezt a keretrendszert használjam, mint kódkönyvtárt. Az újragondolt változat elkészítésekor prioritizáltuk, hogy azok a funkciók készüljenek el először, amelyek lehetővé teszik a mérések gördülékeny elvégzését, valamint a mérési eredmények összegyűjtését egyszerűen feldolgozható formába. A következő szakaszokban röviden

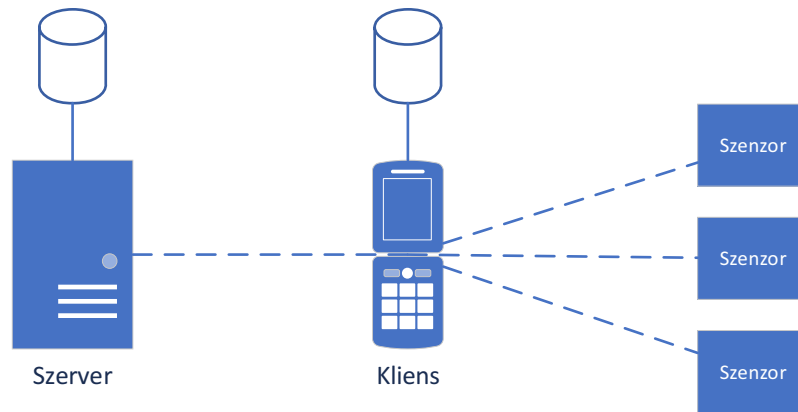
bemutatom a keretrendszer architektúráját, valamint kitérek a számomra fontos funkciók leírására.

## 4.3 Architektúra

### 4.3.1 Rendszerarchitektúra

Az AdaptEd keretrendszer kliens-szerver architektúrájú. Az Android rendszerű mobilkliensen egy háttérben futó Android service felel a keretrendszer fő funkcióiért. A szerver egy ASP.NET alapú backend, amit nem implementáltunk újra, hanem felhasználtuk a már meglévő verziót, és a mobil szolgáltatást alakítottuk úgy ki, hogy a már létező REST alapú WCF interfészt használja. A kliens a szerverrel HTTP kapcsolaton keresztül kommunikál. A mobilklienshez csatlakoznak a különböző mérőszensorok, amelyek az előző fejezetben említetteknek megfelelően Bluetooth vagy Bluetooth Low Energy kapcsolatot használnak.

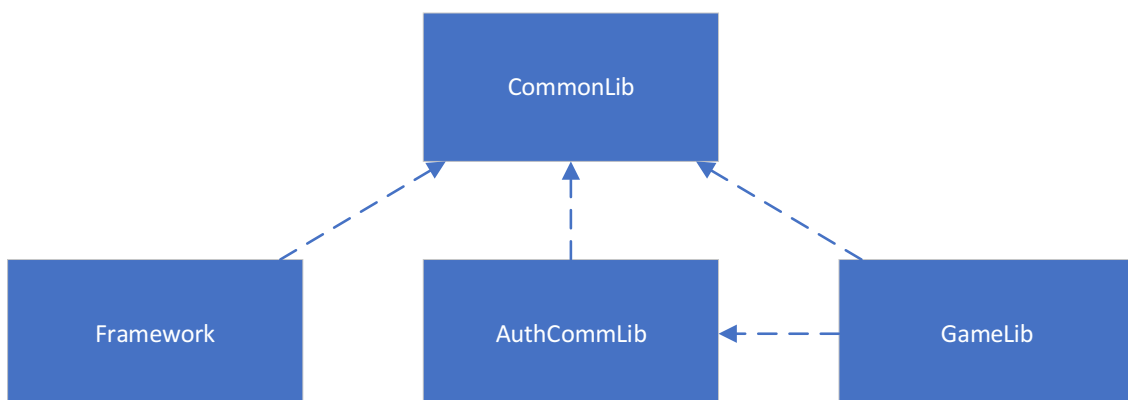
Mivel a végcél az oktatójátékok keretrendszerbe foglalása [20], ezért a kliens szolgáltatáshoz csatlakozó alkalmazások definiálhatnak ún. *játékmeneteket*, ezeken belül pedig tetszőleges tartalmú, időbélyeggel ellátott *eseményeket* küldhetnek a szervernek. Ezen felül maga a szolgáltatás is küld beépített eseményeket, pl. időközönként képernyőmentéseket, illetve az egyes szensorok által kiolvasott adatokat. A jó tesztelhetőség érdekében a valódi szensorok a teszteléskor helyettesíthetők mock szensorokkal is. Az események lokálisan egy SQLite adatbázisban tárolódnak, amit ORM keretrendszer segítségével érünk el, a szerverre pedig periodikusan feltöltődnek, ha van internetkapcsolat. Kapcsolat hiánya esetén a következő kapcsolódáskor történik a feltöltés. A szerver felületen több nézet is rendelkezésre áll egy adott játékmenet eseményeinek megtekintésére, vagy akár CSV vagy JSON formátumba exportálására is. Az exportált adatokon további adatfeldolgozás, tisztítás végezhető.



8. Ábra: Az AdaptEd rendszer architektúrája.

### 4.3.2 Kliens oldali architektúra

Az Android rendszeren futó komponenseket több modulra bontottuk. A *CommonLib* nevű modulba kerültek azon kódrészletek, amelyet minden másik modul hasznosít. Az *AuthCommLib* nevű modul tartalmazza az autentikációval és autorizációval foglalkozó, biztonságkritikus kódot. Erre azért is szükség van, mert a szerverrel való kommunikációhoz azonosítás is szükséges. A *GameLib* modul foglalja magába azon részeket, amelyek egy alkalmazás számára szükségesek a keretrendszerrel való kommunikációhoz. Minden a keretrendszer szolgáltatásait használni kívánó alkalmazásnak függenie kell ettől a komponenstől. A fő komponens *Framework* névre hallgat. Ez tartalmazza a háttérben futó Android service-t, valamint ez kommunikál a szerverrel és a csatlakoztatott szenzorokkal. A service-t egy *AIDL (Android Interface Definition Language)* interfészen keresztül lehet elérni, ezzel lényegében távoli metódushívás-szerű a működés. [25]



9. Ábra: A keretrendszer kliens oldali moduljai és függőségeik.

### 4.3.3 Felhasznált technológiák

A tesztelés során az egyes függőségek könnyű helyettesítéséhez egy függőséginjektálást (dependency injection) lehetővé tévő könyvtárat, a *Dagger 2*-t [33] alkalmaztuk, ami annotáció alapú kódgenerálással működik.

Az események helyi mentéséhez egy objektum-relációs leképző (object-relational mapping, ORM) könyvtárat, a *Requery*-t [28] használtuk, ami megfelelően annotált Java osztálydefiníciók alapján képes SQL-alapú adatbázis generálására, majd az abból való lekérésre, az általunk definiált osztályok példányait visszaadó, egyszerű Java metódusok segítségével.

A szerverrel való kommunikációért a *Retrofit* [26] és *OkHttp* [27] könyvtárak felelnek, amelyek segítségével a HTTP kérések is egyszerű metódushívásokká tehetők kód szinten. Természetesen ügyeltünk rá, hogy mind a diszk, mind a szerver elérése háttérszálon történjen, ehhez az Android SDK által tartalmazott *HandlerThread*, *Looper* és *Handler* osztályokat vettük igénybe.

A csatlakoztatott alkalmazások és a keretrendszer service közötti AIDL interfészen átvitt objektumok esetén az Android beépített, *Parcelable* interfészt használó sorosítási megoldása működik a háttérben.

A teszteléshez a *JUnit* [29] és *Mockito* [30] könyvtárakat, valamint a kódminőség biztosítása érdekében a *JaCoCo* [31] és a *SonarQube* [32] eszközöket használtuk.

## 4.4 Fő funkciók

### 4.4.1 Alkalmazások csatlakoztatása

A keretrendszerhez tetszőleges Android alkalmazás csatlakoztatható, ehhez csupán függőségként kell tartalmaznia a *GameLib* modult vagy az ebből fordított Android osztálykönyvtárat (Android Library). A csatlakoztatás egy inicializációs metódus meghívásával történik, majd minden Android Activity-nek, ami kommunikálni kíván a keretrendszerrel, le kell származnia a *GameLib GameBase* osztályából, és a megfelelő metódusok meghívásával eseményeket küldhet annak, illetve lekérdezheti annak állapotát.

#### 4.4.2 Események küldése

A csatlakoztatott alkalmazásoknak lehetősége van tetszőleges események definiálására, amik egyszerű Java osztályok (angol terminológiával *plain old Java object*-ek, POJO-k) lehetnek. Az osztályok felküldendő mezőit egy speciális annotációval kell ellátni, majd ebből egy reflexió (reflection) alapú megoldás egy csomagoló (wrapper) osztály példányát készíti el. Ez az osztály már implementálja az Android SDK *Parcelable* interfészét. Ez azért fontos, mert ez a követelménye, hogy át lehessen küldeni azt két folyamat, esetünkben az alkalmazás és a service folyamatai között az AIDL interfészen keresztül.

#### 4.4.3 Lokális adatmentés

Az alkalmazástól beérkezett és a szenzorok által generált események egy helyi SQLite adatbázisban is tárolódnak. A fentebb említettek szerint, itt egy ORM megoldást alkalmaztunk, amely fölött egy *repository* tervezési minta van megvalósítva. Ennek az az előnye, hogy készíthető belőle tesztelésre alkalmas (mock) implementáció is, amely segítségével az ORM keretrendszer használata helyett memóriabeli objektumok is használhatók, így gyorsabbá, gördülékenyebbé válik a tesztelés, ráadásul az ORM keretrendszertől függetlenül tesztelhetővé válnak a komponensek.

#### 4.4.4 Együtműködés a szenzorokkal

Az előző fejezetben bemutatott szenzorok mind képesek kommunikálni az Android rendszerrel, ennek megfelelően mindegyik illesztve lett a keretrendszerhez a megfelelő, gyártó által biztosított osztálykönyvtár segítségével. Annak érdekében, hogy a szenzorok által küldött adatok egységesen legyenek kezelve, minden egyes illesztést végző osztály implementál egy speciális interfészt, amelyen keresztül lekérdezhető a szenzor állapota, valamint a csatlakoztatott alkalmazások értesítést kaphatnak az állapotváltozásokról és a szenzor által küldött eseményekről. A háttérben futó service természetesen a szerverre is felküldi a szenzoreseményeket.

### 4.5 Összefoglalás

Ebben a szakaszban bemutattam a mérésekhez szükséges szoftveres környezetet. A tervezéskor lefektetett alapelveknek megfelelően egy jól bővíthető, tesztelhető, lazán csatolt, skálázható, robosztus rendszer jött létre több ember munkájának eredményeként. Személyes kontribúcióm a tervezés és megvalósítás során kiemelkedő volt, többek között

részt vettem az architektúrális tervezésben, a komponensek interfészeinek definiálásában, az események küldését, valamint lokális mentését megvalósító logika és a szerverrel való kommunikáció implementálásában, valamint elvégeztem az Emotiv Insight eszköz keretrendszerrel való illesztését.



## **5 Az IRT modellek kiterjesztése a mentális állapot figyelembevételével**

### **5.1 Bevezetés**

Az előző fejezetekben bemutatásra kerültek az IRT alapú pszichometriai modellek alapjai, valamint a fiziológiai tények alapján a mentális állapotra való következtetés lehetőségei. Az egyszerű IRT modellek azonban nem foglalnak magukba semmilyen, a mért alany pillanatnyi állapotával kapcsolatos változót, így a mentális állapotot sem. A témával foglalkozó kutatók álláspontja alapján további tudományos vizsgálat szükséges a pillanatnyi tényezők figyelembevételéhez, ezért kutatásom ezt az irányt célozza meg.

Ebben a fejezetben bemutatok egy újszerű megoldást a fenti problémára vonatkozóan. A megoldás alapja a már korábban bemutatásra került, négyparaméteres logisztikus IRT modell (4PL). A fejezet során az új modellt szimulációs, illetve mérésekkel történő vizsgálatnak vetem alá és teljeskörűen értékelem.

### **5.2 A modell bemutatása**

A szokásos négyparaméteres, logisztikus IRT modell egy egydimenziós modell, azaz az egyetlen személyi paraméter a képesség. Ezáltal egy ilyen modellben kizárólag a képesség áll a háttérben az adott személy által bármilyen itemre adott válasznak.

A valóságban természetesen a teszteken mutatott teljesítménynek két komponense van: a tudás maga nem elég, az optimum eléréséhez megfelelő mentális és érzelmi állapot is szükséges. Éppen ezért célszerű egy olyan személyi paraméter bevezetése, amely arányos a személy mentális megterhelésével. Az IRT modellek alapja egy adott itemre a helyes megoldás valószínűségét a képesség függvényében megadó itemválasz függvény, tehát az új paraméternek ebben kell megjelennie. A cél, hogy rosszabb, kevésbé koncentrált, fáradtabb állapotban a modell szerint kevesebb legyen a helyes megoldás valószínűsége.

A négyparaméteres modellben a negyedik paraméter egy felső aszimptotát ad az item karakterisztikus görbének. A paraméter értékét 1-nek választva pontosan a háromparaméteres modellhez jutunk, azt csökkentve pedig a függvény grafikonja az  $y$

tengely mentén fokozatosan „összenyomódik”, azaz a helyes megoldás valószínűsége minden képességszint esetén csökken.

Az általam javasolt modell matematikailag ekvivalens a négyparaméteres modellel. A lényegi különbség abban rejlik, hogy míg a felső aszimptotát megadó paraméter a négyparaméteres modell esetén item paraméter, azaz a teszt adott feladatának tulajdonsága, az új modellben viszont egy folyamatosan változó személyi paraméter lesz. Az új modell tehát kétdimenziós, a modell itemválasz függvénye kétváltozós, és a következő alakú:

$$p_i(\theta, \delta) = c_i + \frac{\delta - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

A képletben  $\delta$ -val jelöltem a mentális állapotból származó paramétert, a többi jelölés jelentése megegyezik az IRT modellekről szóló fejezetben írtakkal.

A változtatás alapját az adja, hogy az eredeti négyparaméteres modellben is a figyelmetlenségből adódó véletlen hibázás hatásának modellezésére vezették be a negyedik paramétert. A négyparaméteres modellt azonban a gyakorlatban kevésbé használták, mivel ennek a faktornak a mérésére nem állt rendelkezésre megfelelő módszertan. Mivel a figyelmetlenség a fáradtság és az ideálistól eltérő mentális állapot eredményeképpen jön létre, biofeedback eszközök segítségével pedig ez mérhető, így a negyedik paramétert személyi paraméterként figyelembe véve az eredeti IRT modell kiegészíthető a személy pillanatnyi állapotát is magába foglaló modellé.

Természetesen felmerülhet a kérdés, hogy miért nem a képességet befolyásoló módosítóként tekintünk a mentális terhelésből származó paraméterre. A válasz egyszerű: maga a képesség valójában nem változik pl. a fáradtság növekedésével, hanem valójában csak a teljesítmény csökken, ami az IRT modellben a helyes megoldás valószínűségéként, azaz az itemválasz függvény értékeként fogható meg. Másrészt amennyiben a képesség módosítójaként tekintenénk az új paraméterre, magas képességszintek esetén a megoldási valószínűség a háromparaméteres modellhez hasonlóan 1-hez tartana, ami kevésbé felel meg a valóságos helyzeteknek.

## 5.3 Mérési paraméterek

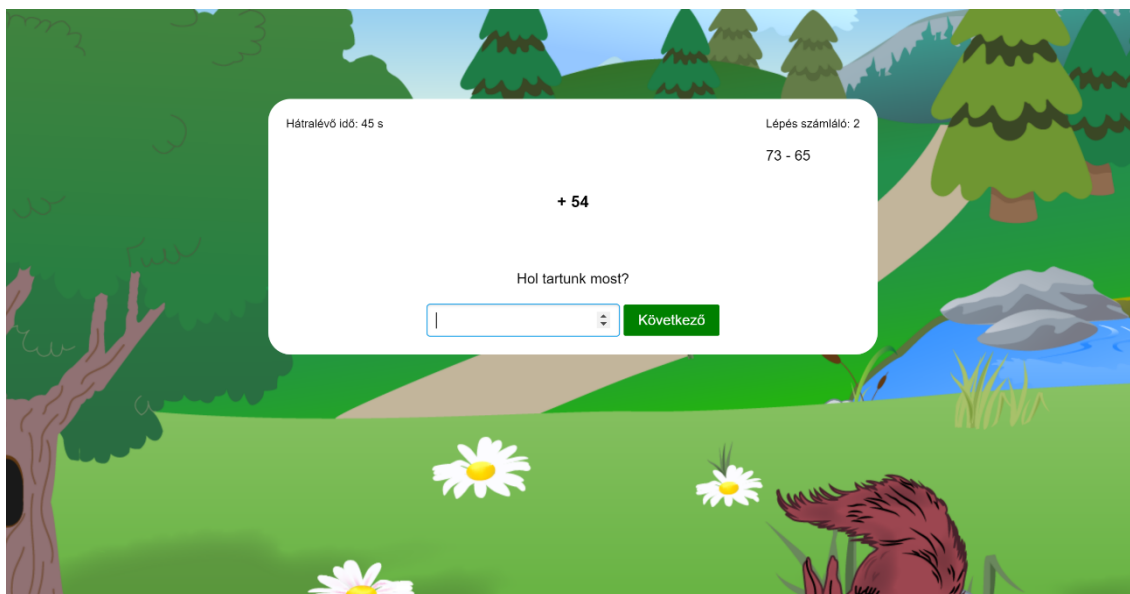
### 5.3.1 A mérés menete

A mérésekhez a tanszéken fejlesztett kognitív játékok alkalmazást használtam. Ez egy webes és mobil környezetben is elérhető alkalmazás, amely nevének megfelelően az egyes kognitív képességterületeket mérő játékok együttese.

Az egyes mért képességterületek a Howard Gardner amerikai pszichológia professzor által jegyzett többszörös intelligencia elméletet alapul véve kerültek kiválasztásra. [14] Gardner szerint az intelligencia ugyanis nem egy egységes értelmi képesség, hanem több különböző területen beszélhetünk intelligenciáról, ilyenek például a nyelvi, matematikai-logikai, térbeli, zenei, testi-kinesztéziás, intraperszonális és interperszonális területek. A kutatás során ezeket képességterületnek tekintettem, így gyakorlatilag is megfoghatóvá az IRT modell által mért képesség fogalma, hiszen egy-egy kognitív területen jelentkező képesség megfeleltethető azon teszt által mért képességnek, amely az adott kognitív terület mérésére lett kifejlesztve.

A kognitív játékok alkalmazásban pszichológusok segítségével összeállított feladatokkal találkozhat a felhasználó, minden játék egy-egy jól meghatározott képességterület mérésére lett kifejlesztve. Példák az egyszerű számolós játék, vagy a különböző n-back tesztek, ahol a játékosnak mindig az n-nel megelőző ingerre (ami lehet szöveg, szám, kép, hang stb.) kell emlékeznie.

A mérések a matematikai-logikai képesség mérésére szolgáló számolós játék segítségével történtek. Ennek a lényege nagyon egyszerű. A felhasználónak mindig egy összeadás vagy kivonás műveletet kell elvégezni, az eredményt fejben tartva. Ezen kívül néha a számok helyett absztrakt formák szerepelnek. Ebben az esetben először meg van adva, hogy az adott forma melyik számnak felel meg, majd később a játékosnak kell emlékeznie erre. Amennyiben a játékos elrontja a műveletet, a következő műveletnek újból mindkét operandusa meg lesz adva, csakúgy, mint a legelső művelet esetén. A játék mindig fix, előre meghatározott időre megy: a cél az adott idő alatt minél több művelet elvégzése.



10. Ábra: Példa a számolós játék egy játékmenetére.

### 5.3.2 Megfeleltetés az IRT modellnek

A fentebb részletezett számolós játék egy játékmenete jól megfeleltethető egy speciális, mentális állapottal kibővített IRT modellt használó tesztnek. Az IRT által mért képesség ebben az esetben megfelel a matematikai-logikai képességnek, a mentális állapotot pedig a korábbi fejezetben bemutatott mérőeszközök (EEG és szívritmusmérő) által szolgáltatott adatokból származtatjuk.

A teszt itemek az egyes elvégzendő műveletek (összeadás vagy kivonás) lesznek. Itt számos egyszerűsítéssel éltem: először is minden teszt itemet azonos nehézségűnek vettem, holott a valóságban elképzelhető, hogy bizonyos számok összeadása vagy kivonása nehezebb, mint másoké. Másodszor a teszt itemek diszkriminációs paraméterét minden teszt item esetén azonosan 1-nek választottam, ez lényegében azonos a kétparaméteres IRT modell egyszerűsítéséhez a háromparamétereshez képest. Harmadszor pedig a pseudo-találgatás paraméterét egyszerűen a lehetséges válaszok száma alapján becsültem. Ez úgy volt lehetséges, hogy a játékban az eredmény mindig csak egy előre meghatározott, a felhasználó által is ismert intervallumból kikerülő egész szám lehet, így egyszerűen az intervallumba tartozó egész számok számának reciproka lett a paraméter értéke.

A fenti egyszerűsítéseknek több oka van: először is, első körben gyakran célravezető lehet egy ún. relaxált probléma megoldása, ahol az eredeti megkötések egy részét elengedjük. Másodszor pedig az item paraméterek becsléséhez a IRT modelltől

szóló fejezetben említett becslő algoritmusok sok, tipikusan több száz alanytól igényelnek mérési adatokat, ami sajnos a számolós játék esetén nem állt rendelkezésre.

képesség	matematikai-logikai
mentális állapot	EEG + szívritmus alapján számolt
teszt itemek	összeadás, kivonás
nehézség	azonos
diszkrimináció	azonos (1)
pszeudo-találgatás	1 / intervallum hossza

2. Táblázat: A számolós játék megfeleltetése a kiterjesztett IRT modellnek.

### 5.3.3 Bemeneti paraméterek

A mérések során minden alany esetén a mérés bemeneti paramétereinek értéke azonos volt. Minden mérés egységesen 40 percig tartott, megszakítás nélkül, az alanyoknak a számolós játékkal kellett játszaniuk. A számolások a 100-as számkörben történtek, ami azt jelenti, hogy a műveletek eredménye minimum 0, maximum 99 lehetett. Az egyszerre megjegyzendő szimbólumok száma maximum 2, a szimbólumot is tartalmazó kérdés adásának valószínűsége pedig 25% volt.

minimális eredmény	0
maximális eredmény	99
szimbólumok száma	2
szimbólum valószínűség	25%

3. Táblázat: A mérés bemeneti paramétere.

### 5.3.4 Mért értékek

Minden alany esetén a következő értékek mérése történt meg:

- összes válasz mennyisége,

- helyes válaszok mennyisége,
- az egyes válaszok időbélyege,
- az egyes válaszok helyessége,
- az egyes válaszok időtartama,
- az alany EEG (MindWave vagy Insight) által mért értékeinek változása.

Néhány alany esetén a szívritmust is mértük a Zephyr HxM Heart Rate Monitor eszközzel. Ilyenkor a nyers értékekből R–R intervallumok számolása, majd ezekből kétféle algoritmussal szívritmus-variancia számolása is történt. Az EEG értékek esetén felváltva kétféle eszközt használtunk, ezzel az egyes eszközöket is jobban össze tudtuk hasonlítani.

Az egycsatornás NeuroSky MindWave Mobile eszköz által mért értékek:

- az alany figyelme (attention),
- az alany nyugalma (meditation).

Mindkét érték 0 és 100 közötti egész szám.

Az ötcsatornás Emotiv Insight eszköz által mért értékek:

- az alany érdeklődése (interest),
- az alany feszültsége (stress),
- az alany összpontosítása (focus),
- az alany elmélyültsége (engagement),
- az alany kikapcsolódása (relaxation),
- az alany izgatottsága (excitement).

Az összes érték 0 és 1 közötti lebegőpontos szám.

Az eszközöktől a fenti metrikákat másodpercenkénti mintavételezéssel kérdeztük le. Ez egy jó kompromisszum volt annak érdekében, hogy kellően sűrűn is kapjunk adatot, de ne is küldjünk egyszerre túl sok eseményt a szervernek.

A sokféle metrika mérése egyrészt az eszközök képességeinek kihasználása érdekében történt, másrészt pedig amiatt, hogy a lehető legpontosabban megállapítható legyen, hogy mely metrika (esetleg mely metrikák kombinációja) indikálja legjobban a

teljesítményt, azaz melyik hasznosítható a módosított IRT modellben a mentális állapotot magában foglaló változóként.

## 5.4 Szimuláció

### 5.4.1 Algoritmus és paraméterek

A modell viselkedésének elemzéséhez a mérés előtt elkészítettem egy egyszerű szimulációs algoritmust, amit Python nyelven implementáltam. Ennek célja az volt, hogy láthassam, várhatóan mennyi mérés kell a viszonylagosan kevésbé zajos eredmények eléréséhez, illetve az egyes paraméterek finomhangolása hogyan hat a modellre. Az algoritmus a módosított, kétváltozós itemválasz függvény alapján számol, szimulálva a mérési körülményeket.

A szimuláció több feltevésen alapul. Az első és legfontosabb, hogy a szimulált személyek viselkedése, azaz az itemekre adott válaszuk pontosan az IRT modell szerinti, továbbá képességük standard normális eloszlású, azaz a képesség eloszlásának várható értéke 0, szórása 1. Ez jól megfelel a valós esetnek. A személyek mentális állapotát a  $\delta$  paraméterrel jellemezzük, ez a szimuláció alatt egyenletesen, lineárisan csökkenő, azaz eleinte nagyobb valószínűséggel adnak helyes választ az itemekre, mint később. A személyek válaszidejét egy  $1 / 3$  paraméterű exponenciális eloszlással modelleztem, ennek várható értéke 3, azaz átlagosan 3 másodperc alatt ad választ egy itemre a szimulált személy. A többi paraméter megfelel a valós méréseknek, azaz a szimulált időtartam 40 perc, a teszt itemek nehézsége azonos, diszkriminációja azonosan 1, pszeudo-találgatás értéke az intervallum hosszának (100) reciproka.

Az alábbi táblázat összefoglalja a szimuláció során használt paraméterek értékeit.

személyek képessége	normális eloszlás, $\mu = 0$ , $\sigma = 1$
személyek $\delta$ paramétere	0,95-től 0,75-re lineárisan csökkenő
személyek válaszideje	exponenciális eloszlás, $\lambda = 1 / 3$
itemek nehézsége	-3
itemek diszkriminációja	1

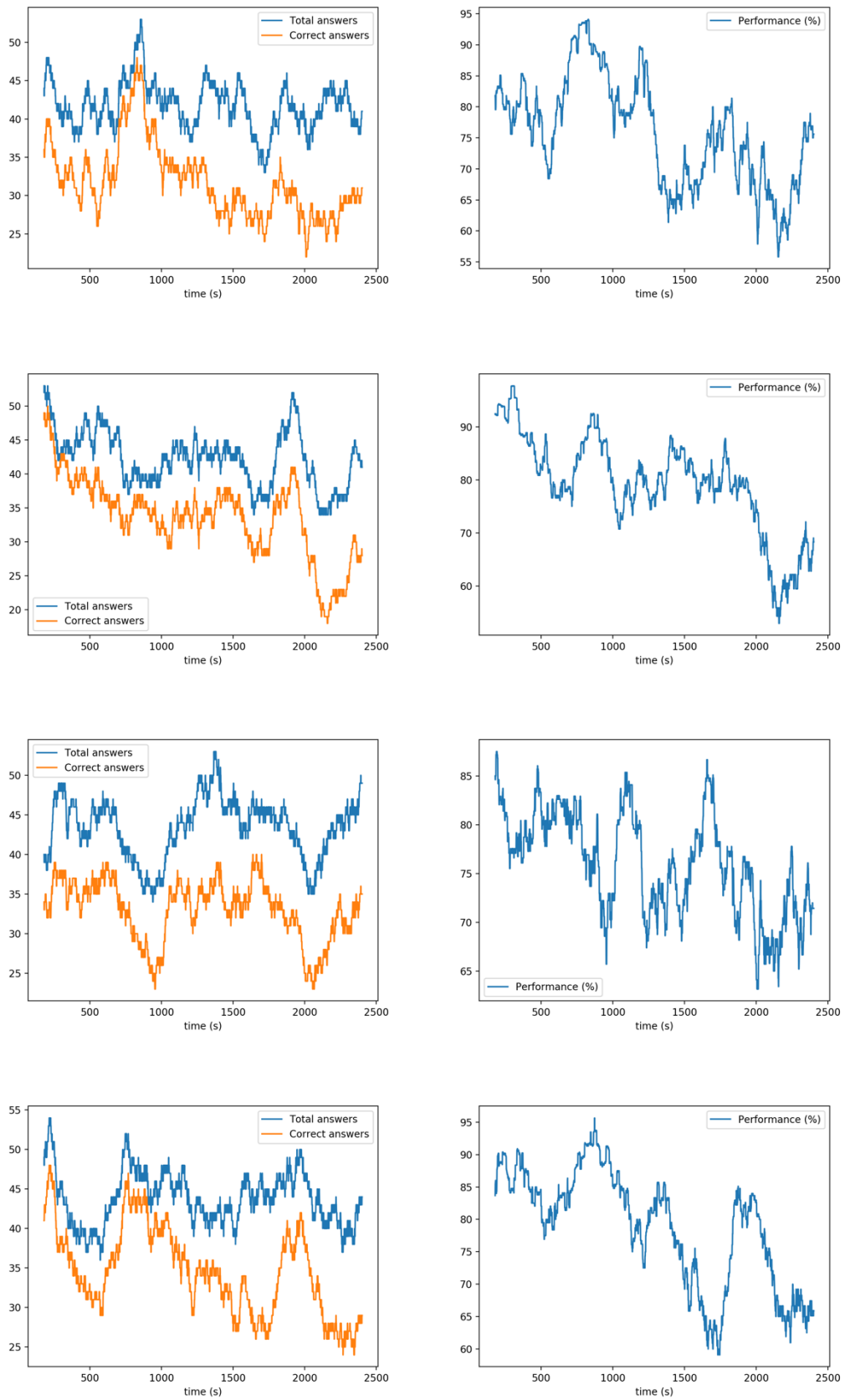
itemek pszeudo-találgatás értéke	1 / 100
szimulált időtartam	40 perc
szimulált személyek száma	1, 10, 100, 1000

### 5.4.2 Szimulációs eredmények

A szimuláció során a szimulált személyek számát fokozatosan növeltem. Eleinte egy-egy személy értékeit, később 10, 100 majd 1000 személy értékeinek átlagát vizsgáltam. Minden esetben a megválaszolt, helyesen megválaszolt teszt itemek számát, illetve a százalékos teljesítményt (a kettő hányadosának százszorosát) vizsgáltam, 3 perces időablakra vonatkozó mozgóátlagolással.

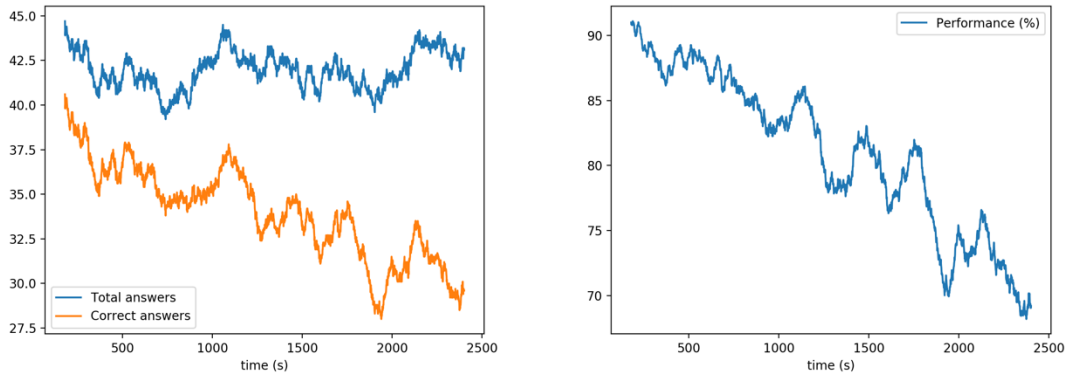
Egyetlen személy esetén a bevitt véletlen tényezőknek köszönhetően igencsak zajosak a kapott eredmények, de a trend már itt is felfedezhető: az idő múlásával valóban csökken a teljesítmény a  $\delta$  paraméter lineáris csökkentésével.



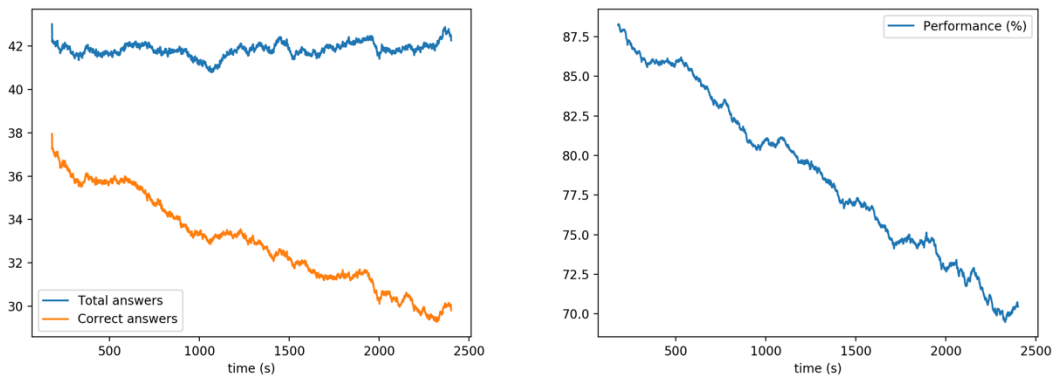


**11. Ábra: Négy különböző szimulált személy válaszainak (*total answers*) és helyes válaszainak (*correct answers*) száma, valamint százalékos teljesítménye (*performance*) az idő függvényében 3 perces időablakra mozgóátlagolva.**

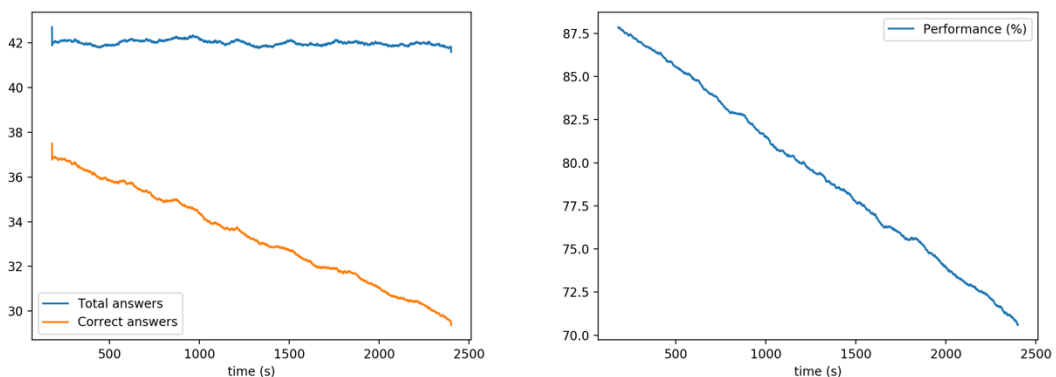
A személyek számának növelésével és az egyes időfüggvények átlagát véve az tapasztalható, hogy a grafikonok fokozatosan kisimulnak, azaz egyre jobban látszik a beállított lineáris csökkenés.



(a) 10 szimulált személy átlaga



(b) 100 szimulált személy átlaga



(c) 1000 szimulált személy átlaga

**12. Ábra:** A válaszok (total answers), helyes válaszok (correct answers) és a százalékos teljesítmény (performance) alakulása az idő függvényében 3 perces időablakra mozgóátlagolva, különböző számú szimulált személyek értékeinek statisztikai átlagát képezve.

A kapott szimulációs eredmények a modell konvergenciáját mutatják, azaz azt a tényt, hogy minél több mérést végzünk, annál jobban lesz látható a kapott eredményekből az a trend, ami szerint a teljesítmény az idő során változik. A fiziológiai tényekből kapott értékek összes mérése vonatkozó átlagát képezve, az azokban megtalálható zaj is csökkenthető, így sok mérés esetén pontos képet kaphatunk arról, mely metrika vagy metrikák függenek össze legjobban a teljesítménnyel.

## **5.5 Mérési eredmények**

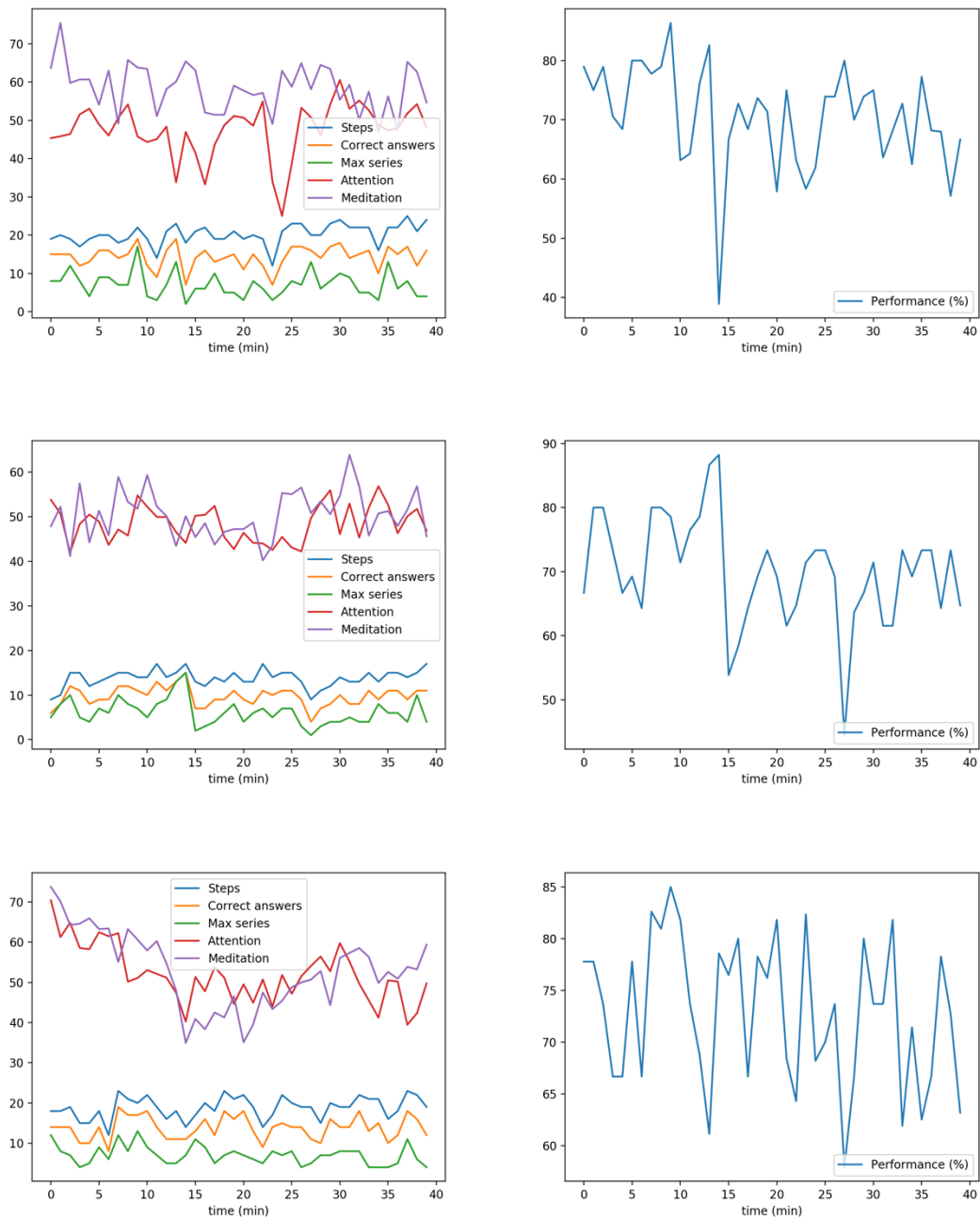
### **5.5.1 Körülmények**

A mérésekben összesen 11 felnőtt személy vett részt. Saját bevallása szerint minden résztvevő egészséges volt, egyiküknél sem diagnosztizáltak szív- és érrendszeri megbetegedést vagy idegrendszeri rendellenességet. Mindegyiküket megkértük, hogy a mérés előtti 8 órában mellőzzék a stimulánsok (koffein, alkohol, dohányzás stb.) használatát. Minden mérés 40 percig tartott, a korábban részletezett elrendezésben, a kognitív játékok közül a számolós játék használatával. A méréseket minden esetben egy kétperces, relaxációs fázis előzte meg.

### **5.5.2 Eredmények, értékelés**

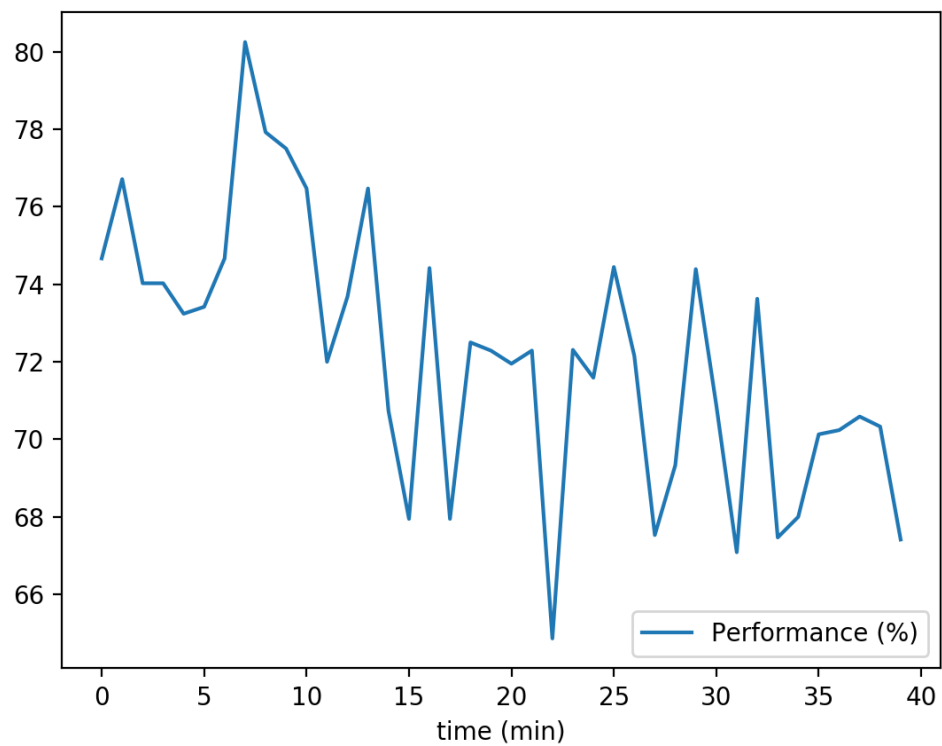
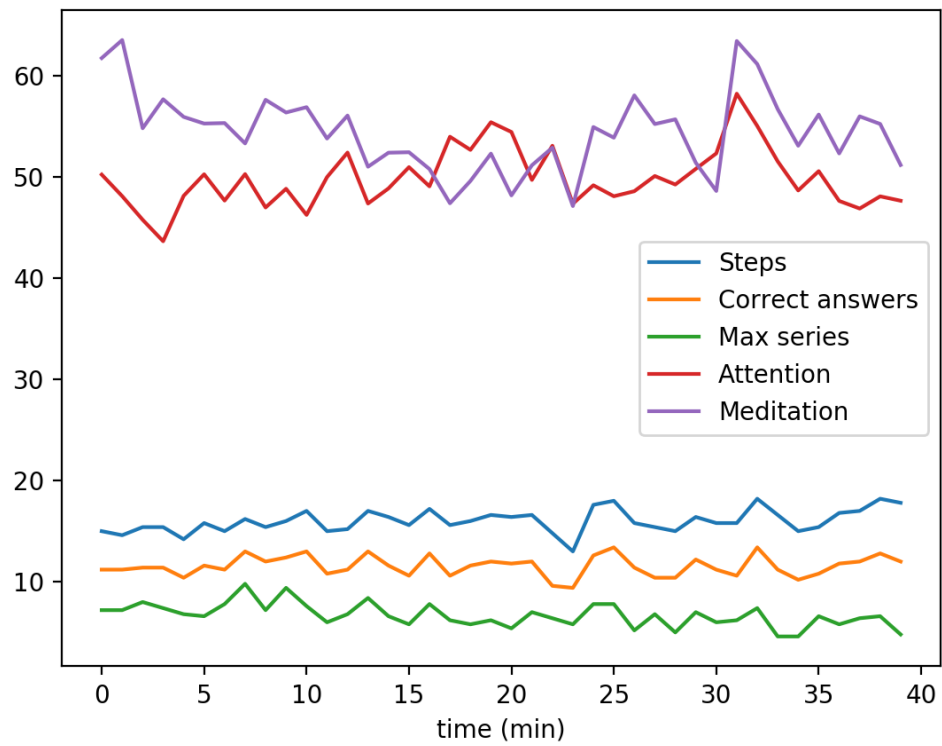
Kezdetben, az alanyok első csoportjánál kizárólag az egycsatornás NeuroSky MindWave Mobile EEG eszközt használtam, másodpercenként mintavételezve az általa mért figyelem (attention) és nyugalom (meditation) értékeket. Ezekben az esetekben az alanyok teljesítményét percenkénti felbontásban hat metrika segítségével mértem: hány műveletet (összeadás, kivonás) oldottak meg, ebből hányszor volt helyes a megoldás, mennyi volt a százalékos teljesítményük, valamint mennyi volt az egymás után legtöbb megoldott művelet száma, valamint mennyi volt a figyelem és nyugalom értékek átlaga az adott percben.

A következő ábrán három kiválasztott alany (A, B és C) eredményei látszanak ebben az elrendezésben. Bár az eredmények erősen zajosak, a százalékos teljesítmény mindhárom esetben enyhén csökken, viszont a figyelem és nyugalom értékek változását nehéz megfogni egy-egy mérési eredményből.



**13. Ábra: Az egyes metrikák változása az első mérési elrendezésben, három különböző alanynál (A, B, C) az idő függvényében.**

Ahhoz, hogy jobban észre vegyem a trendeket, az egyes mérések átlagát vettem, ez látható a következő ábrán. Így már egyértelműbben kirajzolódott egy csökkenő trend a teljesítménynél. A figyelem értéke átlagosan viszonylag állandó volt, viszont a nyugalom értéke enyhén csökkenő trendet mutatott. Egyértelműen látszik viszont, hogy a mérések számának növelésével a kapott grafikonok simulnak. Emiatt további mérések szükségesek a trendek pontos megállapításához, valamint a teljesítmény és az EEG által mért értékek közötti pontosabb összefüggés felfedéséhez.



14. Ábra: Az egyes metrikák változása az első mérési elrendezésben, a mérések átlagából számolva.

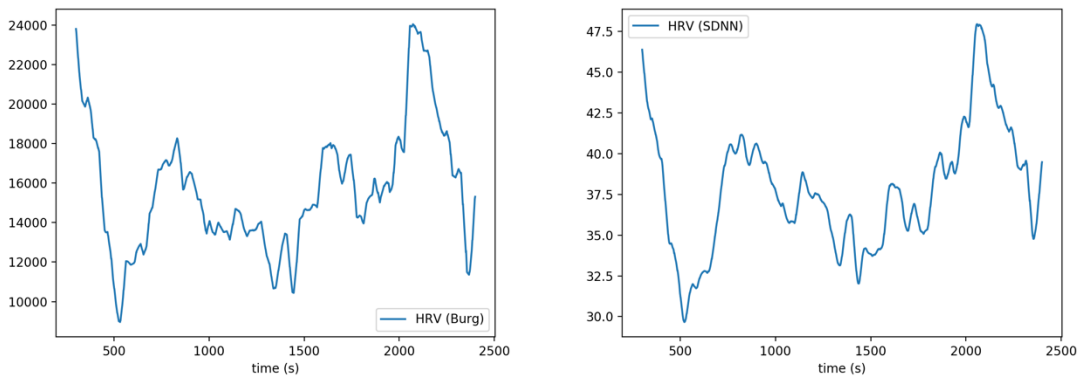
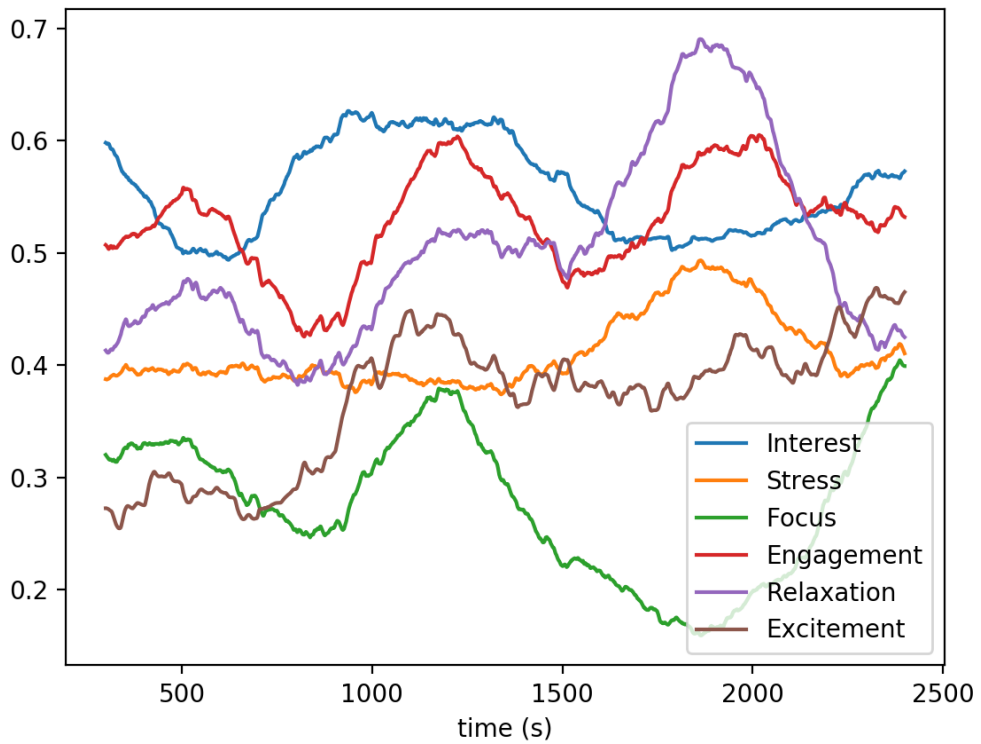
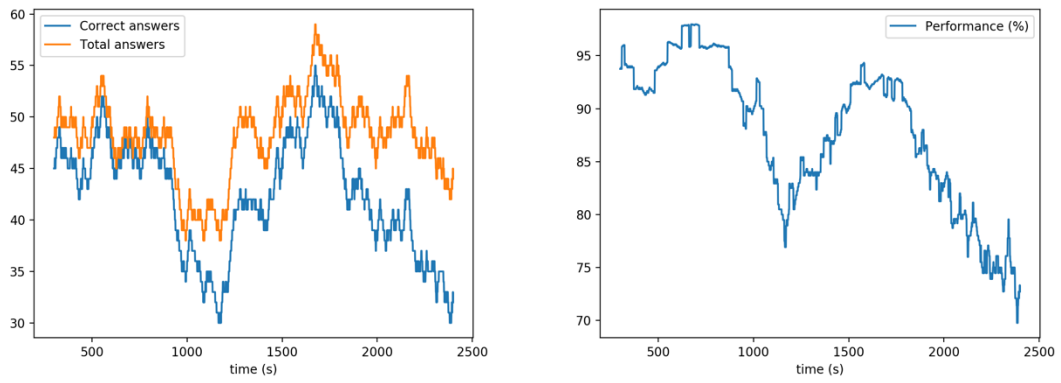
Egy második mérési elrendezésben az ötcsatornás Emotiv Insight eszköz felelt az EEG értékek méréséért. Ez a már említett hatféle teljesítménymetrika mérésére képes. Ebben az elrendezésben a szívritmus is mérésre került, amiből a HRV-re vonatkozó teljesítménymetrikát kétféle algoritmussal is kiszámoltam: az egyik Burg algoritmus, a másik az SDNN (standard deviation of RR intervals, azaz az R–R távolságok szórása) számítás. Mindkettő eredményeként hasonló alakú görbét kaptam.

A következő ábrákon ismét három kiválasztott személy (D, E és F) mért adatai láthatók. Az értékeket itt nem egypercenként átlagoltam, mivel a mérési kódkönyvtár kis módosításával lehetővé vált minden egyes számítás eredményének pontos feljegyzése. Emiatt itt mind a teljesítményre (összes válasz, helyes válaszok, százalékos teljesítmény), mind a mentális állapotra (hatféle EEG és kétféle HRV metrika) vonatkozó értékek 5 perces időablakkal mozgóátlagolásra kerültek a pontosság, valamint a rövid időtartamú, pillanatnyi változások kiszűrése érdekében.

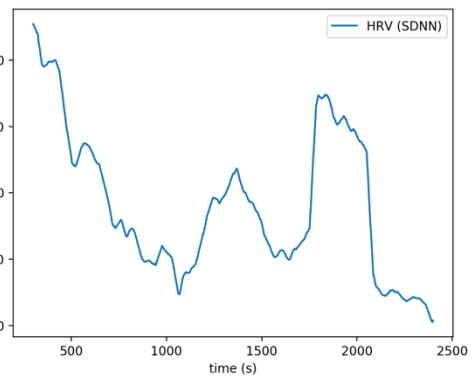
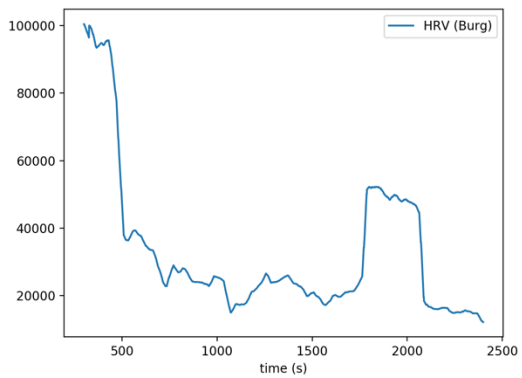
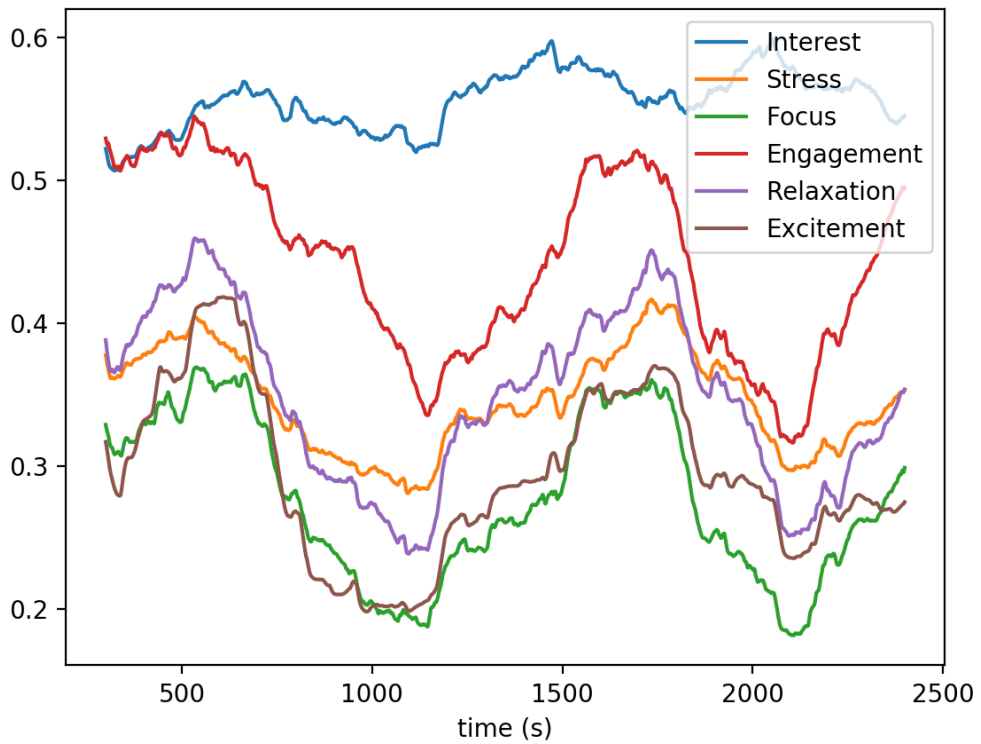
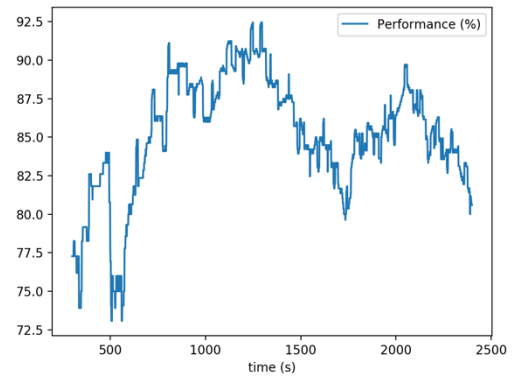
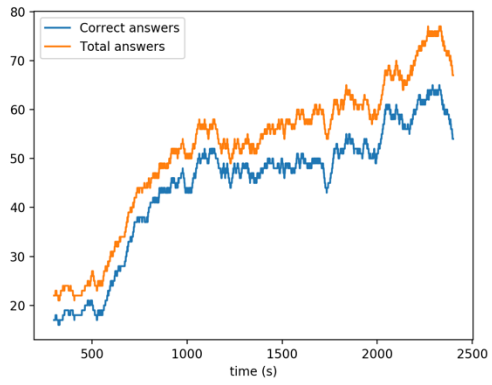
Az összes mérés egyenkénti kiértékelése helyett itt is az egyes mérésekből számítottam statisztikai átlagot, ezt mutatja a 18. Ábra. Mind az egyes mérések egyenkénti vizsgálata, mind az átlag vizsgálata alapján a százalékos teljesítmény esetén egyértelműen kirajzolódik egy csökkenő trend, viszont a mérés elején (a szimulációnál feltettekkel ellentétben) emelkedés figyelhető meg ebben a tekintetben, amelynek lehetséges magyarázata lehet az, hogy a felhasználó akkor „tanul bele” a játék működésébe. Emiatt célszerű lehet egy kb. 15 betanulási fázis beiktatása a mérés elején, amikor még nem mérjük az alanyt.

Érdekes módon az 5 perces ablakban megoldott összes válasz abszolút száma átlagban növekvő trendet mutatott a mérések során (ez is különbözik a szimulációnál feltettekkel), ez a türelmetlenség növekedésével magyarázható: a mérés vége felé a fáradt felhasználó, ha nehéznek érzi a műveletet, nem kezd el hosszasan gondolkodni, inkább rávágja az eszébe jutó első megoldást, vagy adott esetben véletlenszerű megoldást ad (pl. abban az esetben, mikor nem emlékszik a megjelenő szimbólum jelentésére).

A fiziológiai tények közül az SDNN és Burg algoritmussal számított értékek eleinte csökkenő, majd növekedő trendet mutatnak, míg a mérések átlaga alapján az elmélyültség (engagement) és az összpontosítás értéke hajlamos együtt mozogni, és mindkettő csökken, viszont a feszültség (stress) szintje enyhén növekszik, az érdeklődés (interest) szintje pedig viszonylag állandó. A többi metrika esetében nem olvasható ki egyértelmű trend.

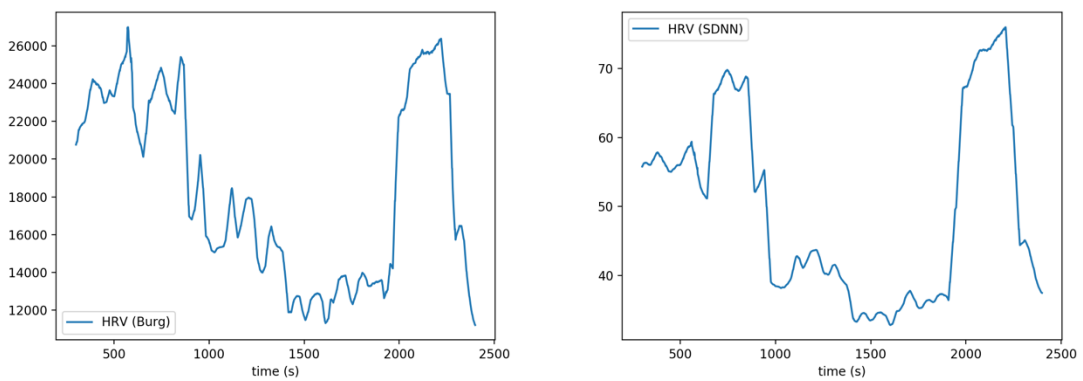
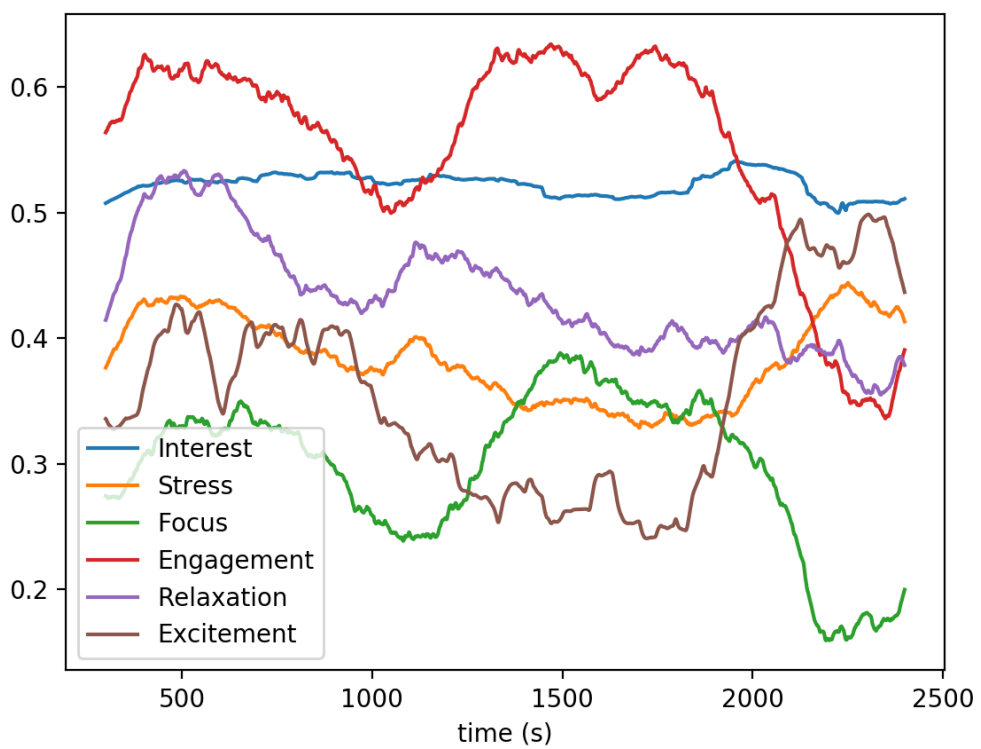
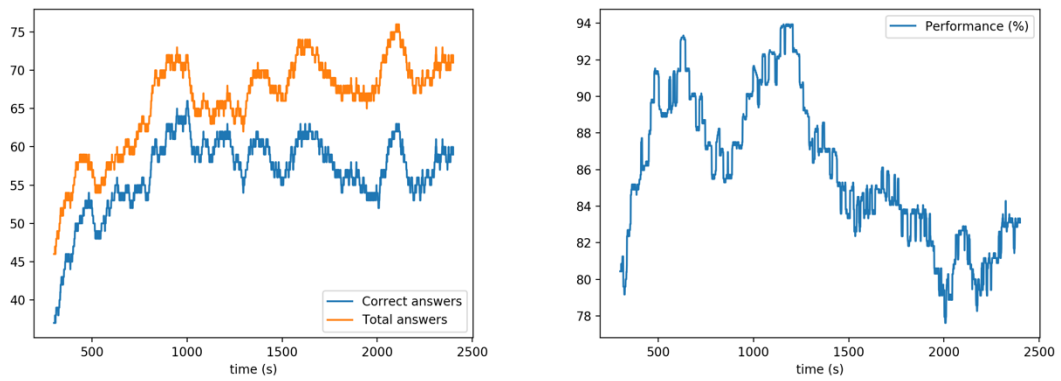


**15. Ábra: Az egyes metrikák változása a második mérési elrendezésben, D alanynál az idő függvényében.**

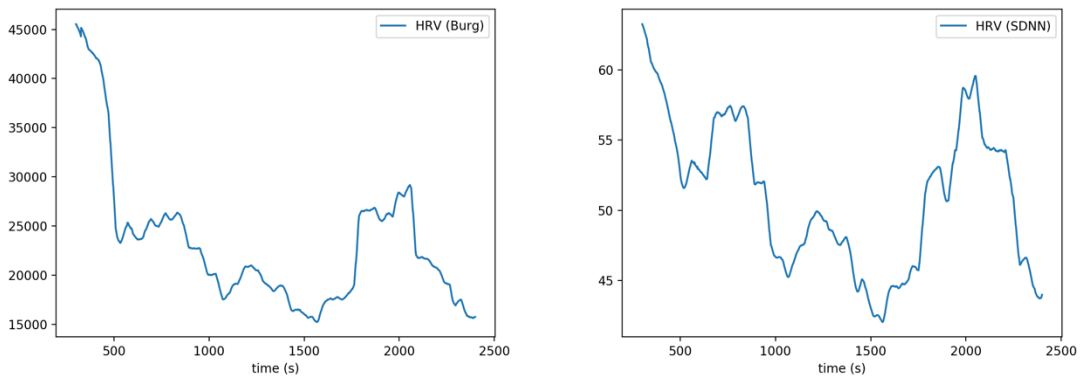
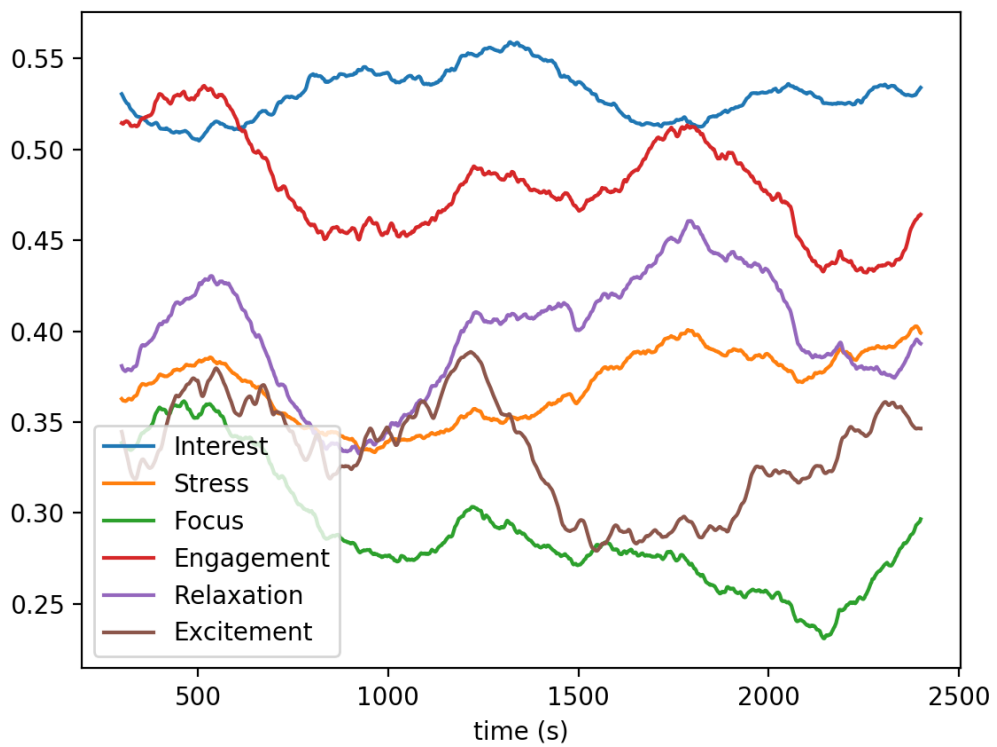
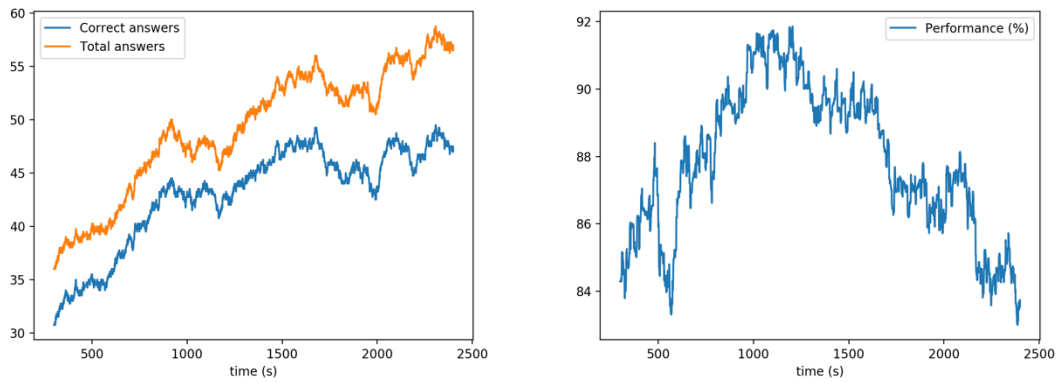


**16. Ábra: Az egyes metrikák változása a második mérési elrendezésben, E alanynál az idő függvényében.**





**17. Ábra: Az egyes metrikák változása a második mérési elrendezésben, F alanynál az idő függvényében.**



**18. Ábra: Az egyes metrikák változása a második mérési elrendezésben, a mérések átlagából számolva.**

Összességében ezen méréscsoport esetén is elmondható, hogy több mérés esetén, valamint további tudományos vizsgálatokkal valószínűleg pontosabb eredményeket lennének képesek kapni, viszont a kapott eredmények mindenképp biztatók, hiszen látható, hogy az egyes kilógó értékek több mérés átlagolásával elsimulnak, amely szintén a modell konvergenciájára enged következtetni.

Konklúzióként: a teljesítménnyel leginkább összefüggő metrikák az EEG által mért értékek közül az összpontosítás, az elmélyültség és a feszültség (amely negatívan befolyásolja a teljesítményt) voltak. A szívritmusból számolt HRV metrikák esetén is felfedezhető korreláció a teljesítménnyel, viszont itt a grafikonokat összevetve az alacsonyabb érték jelenti a jobb teljesítményt.

## 6 Összegzés

A dolgozatban két tématerület hatott egymásra, egészítette ki kölcsönösen egymást: az egyik a pszichometriai modelleké, amelyek segítségével az egyes területeken való képesség becsülhető, a másik pedig a fiziológiai tények alapján való mentális állapot meghatározás. Ezt a kettőt összegyúrva született meg egy olyan újszerű modell, amely a hagyományos IRT modellek kiterjesztéseként az eredetnél nagyobb lehetőségekkel bír, ugyanis a teljesítmény mögött álló emberi tényezőket is képes figyelembe venni.

A kutatásom célja egy olyan adaptív információs rendszer kifejlesztése, amely akár tantermi környezetben is használható, és a fiziológiai tények mérését a tanulókat minél kevésbé zavarva, de annál pontosabban hajtja végre. Az új modellt felhasználva az IRT teszt jellegű feladatokban a heurisztikus becslések mellett vagy akár helyett a precíz matematikai háttérrel felhasználhatjuk arra, hogy egyrészt minél jobb becslést adjunk a képességre, másrészt pedig ne csak a képességet, hanem a pillanatnyi kognitív állapotot is beleszámítsuk ebbe a becslésbe. A hatékony tanulást elősegítendően pedig mindenkinek olyan nehézségű feladatot válasszunk, amely mind képességei, mind pedig az aktuális mentális állapota alapján a neki legmegfelelőbb.

A kutatott terület számos lehetőséget rejt még magában, a megfelelő nehézségű feladatot kiválasztó algoritmus esetén például célszerű lehet egy CAT-alapú módszer alkalmazása, ami ebben az esetben az új modellnek megfelelően kiegészítendő a mentális tényezőkkel. Összességében elmondható, hogy a dolgozat eredményeként egy eddig nem látott megoldás jött létre, amely jó kiindulópont lehet a további tudományos kutatások, vizsgálatok szempontjából.

## Irodalomjegyzék

- [1] M. Kimura, H. Uwano, M. Ohira, and K. ichi Matsumoto, Toward constructing an electroencephalogram measurement method for usability evaluation, in Proc. 13th International Conference on Human- Computer Interaction(HCI International 2009), pp. 95–104, Springer, 7 2009. San Diego, CA, USA.
- [2] S. Sampayo-Vargas, C. J. Cope, Z. He, and G. J. Byrne, The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game, *Computers & Education*, vol. 69, pp. 452–462, 2013.
- [3] Ronald K. Hambleton, Hariharan Swaminathan (auth.)-Item Response Theory Principles and Applications-Springer Netherlands (1985), pp. 15–31, 75–98, 125–150
- [4] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Press.
- [5] Bock, R. Darrell, Moustaki, I. (2007). *Item Response Theory in a General Framework*. In Rao, C. R., Sinharay, S. (Eds), *Handbook of Statistics, Volume 26: Psychometrics* (pp. 469-513). Amsterdam: Elsevier.
- [6] Thissen, D. & Orlando, M. (2001). *Item response theory for items scored in two categories*. In D. Thissen & Wainer, H. (Eds.), *Test Scoring* (pp. 73-140). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- [7] Nering, Michael L.; Ostini, Remo, eds. (2010). *Handbook of polytomous item response theory models*. Taylor & Francis.
- [8] Hambleton, R. K., Jones, R. W. (1993). *Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development*. *Educational Measurement: Issues and Practice*, 12, (3), 38-47.
- [9] Han, K. T. (2012). *Fixing the c Parameter in the Three-Parameter Logistic Model*. *Practical Assessment, Research & Evaluation*, 17 (1), 1-2.
- [10] Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- [11] Bock, R.D.; Aitkin, M. (1981). "Marginal maximum likelihood estimation of item parameters: application of an EM algorithm". *Psychometrika*. 46 (4): 443–459.

- [12] Weiss, D. J.; Kingsbury, G. G. (1984). "Application of computerized adaptive testing to educational problems". *Journal of Educational Measurement*. 21: 361–375.
- [13] M., Csikszentmihalyi (2000). *Beyond boredom and anxiety*. Jossey-Bass.
- [14] H. Gardner, *Intelligence Reframed: Multiple Intelligences for the 21st Century*, 1999, pp. 33-4
- [15] K. Hercegfı, Improved temporal resolution heart rate variability monitoring-pilot results of non-laboratory experiments targeting future assessment of human-computer interaction. *Int J Occup Saf Ergon*, vol. 17, no. 2, pp. 105–17, 2011.
- [16] A. Poole and L. J. Ball, Eye tracking in human-computer interaction and usability research: Current status and future," in *Prospects*, Chapter in C. Ghaoui (Ed.): *Encyclopedia of Human-Computer Interaction*. Pennsylvania: Idea Group, Inc, 2005.
- [17] B. Forstner, L. Szegletes, R. Angeli and A. Fekete, A general framework for innovative mobile biofeedback based educational games, 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom), 2013.
- [18] Szegletes, Luca, and Bertalan Forstner. Reusable framework for the development of adaptive games. *Cognitive Infocommunications (CogInfoCom)*, 2013 IEEE 4th International Conference on. IEEE, 2013.
- [19] L. Szegletes, M. Koles and B. Forstner, The design of a biofeedback framework for dynamic difficulty adjustment in games, 2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom), 2014.
- [20] L. E. Nacke, M. Kalyn, C. Lough, and R. L. Mandryk, Biofeedback game design: using direct and indirect physiological control to enhance game interaction, in *Proceedings of the 2011 annual conference on Human factors in computing systems, CHI '11*, (New York, NY, USA), pp. 103–112, ACM, 2011.
- [21] Mantri, S., Dukare, V., Yeole, S., Patil, D., & Wadhai, V. M. (2013). A Survey: Fundamental of EEG. *International Journal*, 1(4).
- [22] Moon, B. S., Lee, H. C., Lee, Y. H., Park, J. C., Oh, I. S., & Lee, J. W. (2002). Fuzzy systems to process ECG and EEG signals for quantification of the mental workload. *Information Sciences*, 142(1), 23-35.
- [23] MacLean, A., Young, R. M., Bellotti, V. M., & Moran, T. P. (1991). Questions, options, and criteria: Elements of design space analysis. *Human-computer interaction*, 6(3-4), 201-250.

- [24] Robert C. Martin: Clean Code: A Handbook of Agile Software Craftsmanship, 2008.
- [25] Android Interface Definition Language (AIDL):  
<https://developer.android.com/guide/components/aidl.html>
- [26] Retrofit: <http://square.github.io/retrofit/>
- [27] OkHttp: <http://square.github.io/okhttp/>
- [28] Requery: <https://github.com/requery/requery>
- [29] JUnit: <http://junit.org/junit5/>
- [30] Mockito: <http://site.mockito.org>
- [31] JaCoCo: <http://www.eclemma.org/jacoco/>
- [32] SonarQube: <https://www.sonarqube.org>
- [33] Dagger 2: <https://google.github.io/dagger/>
- [34] NeuroSky MindWave Mobile: <https://store.neurosky.com/pages/mindwave>
- [35] Emotiv Insight: <https://www.emotiv.com/insight/>
- [36] Zephyr HxM BT: <https://www.zephyranywhere.com/resources/hxm>