

## **Magyarnyelvű tömörített beszédminták előállítása és minőségi vizsgálata MATLAB-al**

Gelencsér Balázs, [tenere@citromail.hu](mailto:tenere@citromail.hu)  
2008 január

Feladat:

- Ismertesse röviden a virtuális valóság alkalmazási területeit, megvalósítási módjait, különösen a hanghatások irányinformációjának leképezésére (Virtual Auditory Display, auralization), a fejmozgatás szerepére és a fejhallgató lejátszórendszerek minőségére.
- Röviden ismertesse a GUIB projekt célját, állását a korábbi diplomamunkák és elvégzett feladatok alapján. Váolja, hogy a vakok milyen nehézségekkel küzdenek a PC felhasználás során, hogyan lehetne segíteni és javítani ezen és ebben milyen szerepe van a fenti dolgoknak (Earcon, Spearcon, Auditory icon).
- Rövid ismertesse a MATLAB program feladathoz szükséges lépéseit, utasításait, különösen az ausztráliai matlab rutinok működését, célját.
- Készítsen felvételeket magyar nyelven mikrofonnal wave fájl formátumban és futassa a fenti programokat le rajtuk, ezzel előállítva magyar spearcon adatbázist. Nézzon utána, hogy létezik-e (és ha igen, milyen, mit tud) text-to-speech program.

Utolsó módosítás: 2007-10-29

Dr.Wersényi György, C609  
[wersenyi@sze.hu](mailto:wersenyi@sze.hu)

Államvizsgatárgyak:

- |                                |        |
|--------------------------------|--------|
| - Híradástechnikai alkatrészek | N_ta36 |
| - Stúdiótechnika               | N_ta82 |
| - TV technika                  | N_ta55 |

# Tartalomjegyzék

|                                                                                           |    |
|-------------------------------------------------------------------------------------------|----|
| Tartalomjegyzék.....                                                                      | 1  |
| Bevezetés.....                                                                            | 3  |
| 1. A virtuális valóság alkalmazási területei, megvalósítási módjai .....                  | 4  |
| 1.1 Bevezetés .....                                                                       | 4  |
| 1.2 A hanghatások irányinformációjának leképzése .....                                    | 5  |
| 1.2.1 A leírófüggvények .....                                                             | 5  |
| 1.2.2 A helymeghatározás.....                                                             | 6  |
| 1.2.3 A fejhallgató lejátszórendszerek minősége.....                                      | 8  |
| 2. Számítógép-használat vakon.....                                                        | 9  |
| 2.1 A vakok nehézségei a számítógép használat során .....                                 | 9  |
| 2.2 A nehézségek csökkentésének eszközei.....                                             | 9  |
| 2.3 Milyen nehézségekkel kell megküzdenie egy vak felhasználónak<br>Magyarországon? ..... | 12 |
| 3. A hallható kijelző megvalósítási módjainak bemutatása .....                            | 15 |
| 3.1 Az earcon-ok, a spearcon-ok és a hallható ikonok ismertetése.....                     | 15 |
| 3.1.1 Bevezető.....                                                                       | 15 |
| 3.1.2 Az earcon-ok.....                                                                   | 16 |
| 3.1.3 A spearcon-ok.....                                                                  | 18 |
| 3.1.4 A hallható ikonok .....                                                             | 20 |
| 3.1.5 A hallható menük.....                                                               | 22 |
| 3.1.6 Az earcon-ok és a hallható ikonok gyakorlati haszna .....                           | 23 |
| 3.1.7 Az earcon-okkal és a hallható ikonokkal kapcsolatos kérdések.....                   | 24 |
| 3.1.8 A lehetséges megoldások elemzése.....                                               | 27 |
| 4. A tömörített beszédminták előállítása .....                                            | 30 |
| 4.1 A MATLAB program használatának leírása.....                                           | 30 |
| 4.2 A compress.m és a solafs.m fájlok leírása: .....                                      | 32 |
| 5. A tömörített hangminták minőségi vizsgálata .....                                      | 37 |
| 5.1 A hangminták elkészítése, editálása.....                                              | 37 |
| 5.2 A hangminták tömörítési arányának vizsgálata.....                                     | 37 |
| 5.3 A hangminták spektrális vizsgálata .....                                              | 47 |
| 5.3.1 A nyomtatás hangfájl spektrumai .....                                               | 48 |
| 6. Magyar nyelven is működő text-to-speech program ismertetése .....                      | 52 |
| 6.1 SpeakBoard - A beszélő számítógép .....                                               | 52 |
| 6.2 Wave fájlok készítése a SpeakBoard programmal .....                                   | 54 |
| 6.3 A hangminták tömörítési arányának vizsgálata.....                                     | 54 |
| 6.4 A hangminták spektrális vizsgálata .....                                              | 57 |
| 6.4.1 A nyomtatás hangfájl spektrumai .....                                               | 57 |
| 7. Összefoglalás.....                                                                     | 59 |
| 8. Irodalomjegyzék.....                                                                   | 60 |

## Bevezetés

A szakdolgozat pszichológiai háttérrel egybekötött hangjelek előállításáról, azaz az earcon-ok illetve a spearcon-ok létrehozásáról szól, melyet a MATLAB programmal végzek, valamint feladatomban volt ezen hangjelek minőségi vizsgálata.

Az earcon-ok egyrészt zajjellegűek, például különböző rövid hanghatások (például: ajtónyikorgás, ha megnyitunk egy windows ablakot) illetve beszéd és annak „gyorsított változata” a spearcon-ok (spearcon = speech + earcon).

A feladat része egy fejhallgatón keresztül történő hangtérszimulációs kísérletsorozatnak. A szimuláció egy egyszerű PC-s DSP kártyán fut, mely a bementeti mono wave fájlt a merevlemezzről veszi, a fül szűrőhatását jellemző átviteli függvényeket a memóriából, a kettő együttese pedig a szimulált hangteret eredményezi. A Rendszer része a GUIB (Graphical User Interface for Blind Persons) kutatásnak, ahol a cél vakok számára olyan rendszer kifejlesztése, mellyel tudnak PC-t (Windows-t) kezelni egérrel, hangok alapján fejhallgatóval tájékozódva. A képernyőt kell olyan mértékben felosztani, hogy a különböző eseményekhez rendelt hanghatások térben jól elkülöníthetők és az egérrel megtalálhatók legyenek.

A szakdolgozatban röviden áttekintem a virtuális valóság alkalmazási területeit és a fejhallgatók lejátszórendszerek minőségét.

Ismertetem, hogy a vakok milyen nehézségekkel küzdenek a PC felhasználás során és ezt milyen mértékben, csökkentik az earcon-ok, spearcon-ok és a hallható ikonok. A beszédmintákat és a mérési eredményeket a CD melléklet tartalmazza.

# **1. A virtuális valóság alkalmazási területei, megvalósítási módjai**

A fejezet a virtuális hangtér létrehozásával foglalkozik, melyet fejhallgatós lejátszással valósítunk meg.

## **1.1 Bevezetés**

A virtuális valóság létrehozásánál beszélhetünk a látvány, a hangélmény és a tapintás szimulációjáról is. Itt most a virtuális hangtér-szimuláció fejhallgatón keresztül történő leképzésével foglalkozunk. A valósághoz legjobban hasonlító virtuális teret hangforrásokkal hozzuk létre, melyeket irányinformációval látunk el. Legfontosabb paraméter a lokalizáció, a hangforrások irányának meghatározása.

A binaurális, kétfülű hallás technikája magába foglalja a hangfelvételi, lejátszási és szintetizálási eljárásokat, ahol a jel a dobhártyán fellépő akusztikus jel. Tehát a dobhártyákon a hangnyomás időfüggvényt kell reprodukálni úgy, hogy a hallgatónak autentikus hangélményben legyen része. Az átvinni kívánt akusztikai információ legfontosabb része az irányinformáció, azaz a hangforrások helyének meghatározása, melyet lokalizációnak hívunk. A kétfülű hallásnál a hasznos bemeneti jelet kétsatornásnak tekintjük, melyeket mérünk és reprodukálunk, de a gyakorlati megvalósításuk igen nehéz.

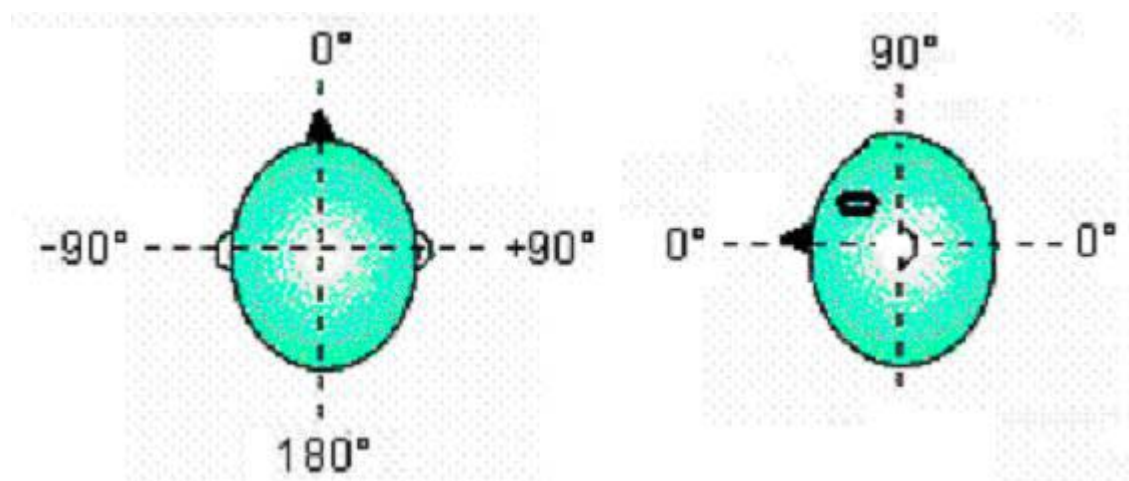
A felvételeket elméletileg a dobhártyán kellene végezni, a gyakorlatban a hallójárat bemenetén készítjük, lejátszáskor a dobhártyáig vezető átviteli utat számítógép állítja elő, és a virtuális térben elhelyezett hangforrást is a számítógép generálja.

## 1.2 A hanghatások irányinformációjának leképzése

### 1.2.1 A leírófüggvények

A hallásfolyamatban részt vesz a külső fül, a középfül, a belső fül és az agy. Az akusztikai hullám először a külső fület hozza rezgésbe, mely a dobhártyáig tart, majd a középfülben a hallócsontokon át a csigáig terjed. A frekvencia-transzformáció és az agy számára szükséges idegi impulzusok előállítása a belső fülben, tehát a csigában történik. A térbeli hallás megvalósítása szempontjából a külső fül vizsgálata a legfontosabb. A külső fület műszakilag komplex átviteli függvényével jellemezzük, melyet az angol elnevezésből HRTF-nek rövidítjük (Head-Related Transfer Function). A komplex HRTF függvénynek van amplitúdó és fázismenete is. Gyakran az időtartományban értelmezett változatát használjuk, ennek neve HRIR (Head-Related Impulse Response), ami az impulzusválasz.

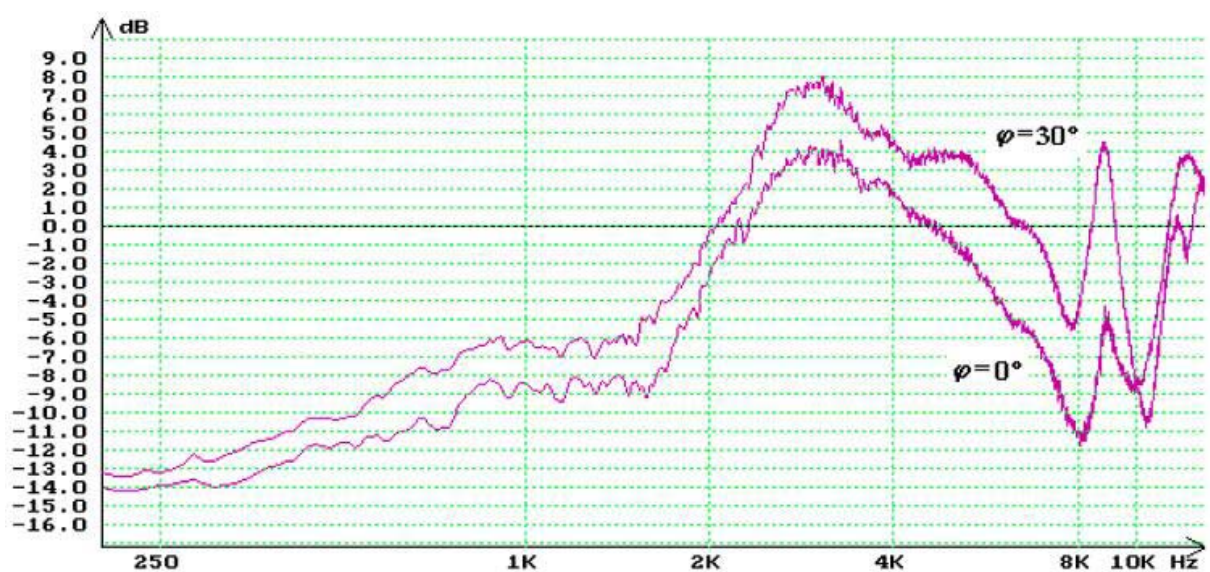
Az átviteli függvények irányfüggők így egy speciális fejhez rögzített koordináta-rendszerben ábrázoljuk. Ennek origója a fej közepében található, és az egyes függvényeket gömbi koordinátákkal ( $r$ ,  $\varphi$ ,  $\delta$ ) paraméterezzük. Az oldalirányú kitérést a horizontális (vízszintes) síkban  $\varphi$ -vel jelöljük, neve azimut. A függőleges kitérést a mediális síkban  $\delta$ -val jelöljük, neve emelkedés vagy eleváció (1.1 ábra).



1.1 ábra. A fejhez rögzített koordináta rendszer: Balra a horizontális sík ( $\varphi$ ), jobbra a mediális sík ( $\delta$ ).

Az  $r$ -rel jelölt paraméter a fej körüli gömb sugara, ez hordozza a távolságinformációt, ami a hangerősség függvénye de ezt nem használjuk, mert általában állandó távolsággal dolgozunk.

A szűrőhatást a HRTF függvénnyel írjuk le, ami személyeként és fülenként más és más, továbbá befolyásolja a felsőtest alakja és a környezet akusztikai tulajdonsága is. Ezért a HRTF függvényeket süketszobai körülmények között értelmezzük, ahol nincsenek reflexiók. A horizontális síkban mért HRTF függvényeket a 1.2-es ábra mutatja.

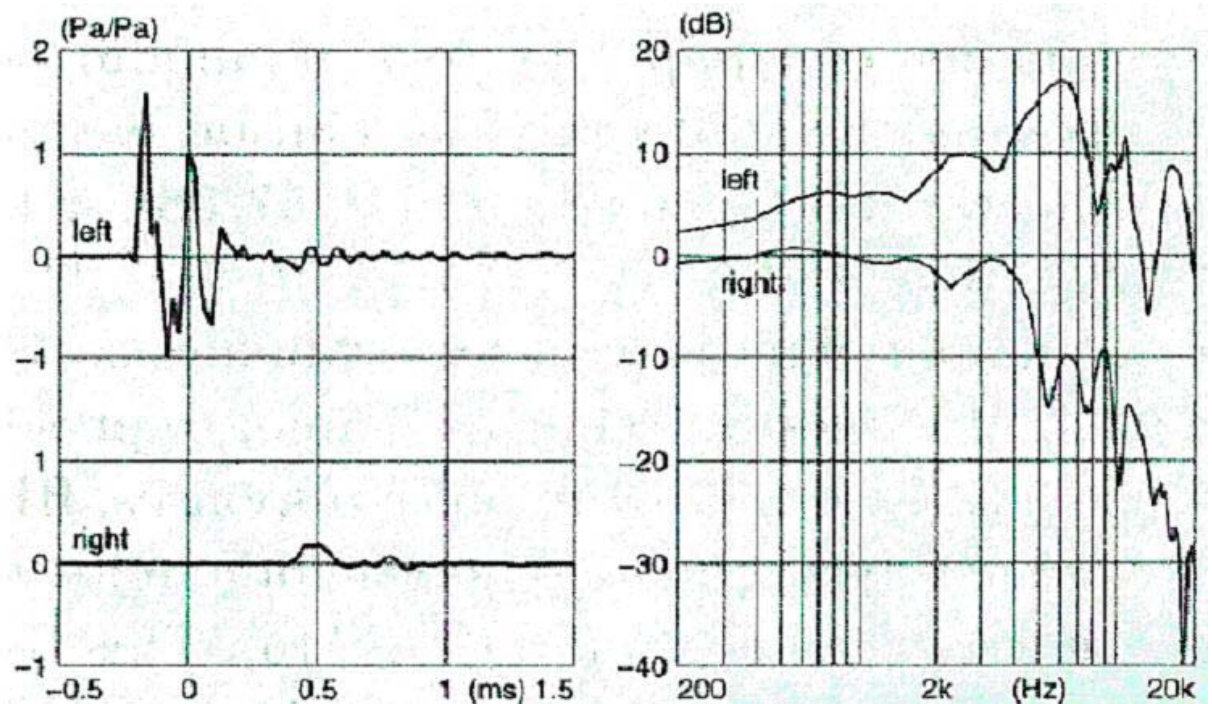


1.2 ábra. HRTF függvények a horizontális síkból a bal fül dobhártyájának helyén mérve (Brüel Kjaer 4128 műfej).

### 1.2.2 A helymeghatározás

A lokalizáció azt jelenti, hogy a fülünkbe érkező hangok irányinformációját magunkhoz viszonyítva képesek vagyunk meghatározni. Ennek pontos megállapítására nem vagyunk képesek, tehát a helyes iránytól való eltérést fokban adjuk meg, amelyet lokalizációs bizonytalanságnak nevezünk. Az irányérzékelés szempontjából nagyon fontos a füljelek közötti (interaurális) hangnyomás és időeltérés. Azt a jelet halljuk közelebbinek, amelyik előbb érkezik a fülbe, vagy hangosabb.

A két fülben ezt a folyamatot az 1.3-as ábra szemlélteti.



1.3 ábra. A külső fül rendszerleíró függvényei. Balra az impulzusválasz (HRIR), jobbra az abból számított átviteli függvény (HRTF) a bal és a jobb fülhöz, élő ember blokkolt hallójáratának bejáratán mérve „bal” ( $\varphi = -90^\circ$ ) irányból.

A fülek felbontóképessége a szemből érkező hangokkal a legpontosabb. A füljelek idő- és szinteltérése vízszintes /horizontális/ síkban nagy lehet, ezért itt a bizonytalanság értéke alacsony, 3-5 fok körüli. A függőleges síkból, tehát a fej szimmetriasíkjából (mediális) érkező jelek a két fülbe egy időben és azonos intenzitással érkeznek, tehát itt a bizonytalanság nagyságrendekkel nagyobb, lokalizáció nehezebb, a hiba akár 10-25 fok is lehet, mert itt csak a HRTF függvények szűrőhatása érvényesül. Ebben a síkban gyakori probléma az elől-hátul döntési hiba, tehát képtelenek vagyunk meghatározni, hogy a hang előlről, illetve hátulról érkezik-e.

A binaurális, tehát kétfülű hallás ellentéte a monaurális (egyfülű) hallás. Egy füllel az irányinformáció meghatározása nagyon pontatlan, mert ebben az esetben csak a hang intenzitására hagyatkozhatunk. Fontos megjegyeznünk a fej és egyéb tárgyak árnyékoló hatását, ami az akusztikai jelet nagyon rontja, ezért

oly jelentős a kétfülű hallás, mert például ha az egyik fül alig jut információhoz, a másik a lehető legtöbbet próbálja kinyerni a csillapított jelből.

Tehát nagyon fontos a fejárnyék hatása, mely aluláteresztő jellegű, és ezeket az időbeli és szintbeli eltéréseket a HRTF és a HRIR függvények tartalmazzák.

### **1.2.3 A fejhallgató lejátszórendszerek minősége**

A fejhallgató lejátszásnál a hangszórók membránja csak pár centiméterre van a fültől, tehát így elveszítjük a légterjedés hatását, és az áthallást a két fül között. A fejmozgás hatását is elveszítjük, mivel a fejünket hiába forgatjuk, a szimulált hangkép együtt mozog a fejünkkel, így kialakulhat a fejközép-lokalizáció, tehát úgy érezzük, hogy a hangforrás a fej közepén, belül helyezkedik el. A virtuális térben így jelentősen romlik a lokalizáció, hiszen fülünket nem tudjuk beállítani a legérzékenyebb irányba. A komoly virtuális valóság szimulátorok a fej helyzetét érzékelik a fejhallgatóra szerelt adóval, így valós időben a HRTF függvények cserélgetésével és számítással a hangképet stabilan tartják. Ez jelentősen csökkenti a fejközép-lokalizáció érzetének kialakulását.

## 2. Számítógép-használat vakon

A fejezet azon eszközökkel foglalkozik, melyek segítik a vakokat a számítógép használat során.

### 2.1 A vakok nehézségei a számítógép használat során

Bizonyára mindenki előtt ismeretesebbek a számítógépek előnyei. A számítógép az élet szinte minden területén könnyítést jelent a felhasználók számára, hiszen megkönnyíti az információszerzést, olvasást, írást, kommunikációt, munkalehetőséget biztosít, hobbi célokra is használható, hogy csak néhány dolgot említsünk a számtalan lehetőség közül. A látássérültek számára számítógép önállóságot biztosít, csökkenti a ráutaltságot a látó embertársakra. Ahhoz, hogy a számítógépet megfelelően használni tudja egy látássérült, mindenekelőtt a tíz ujjas gépelés elsajátítására, ezenkívül speciális programokra, perifériákra van szükségük. Néhány példa ezekre a speciális kisegítő lehetőségekre.

### 2.2 A nehézségek csökkentésének eszközei

#### ➤ Speciális beszélőegység:



Magyarországon a vakok körében dr. Arató András és munkatársai (KFKI RMKI SZHK) Brailab-PC-je terjedt el leginkább. Ez egy több szoftverből álló programcsomagot, és egy a nyomtatóportra köthető hardver elemet tartalmaz. A billentyű echo mellett a BIOS-on keresztül átment szövegeket olvassa, hangsúlyozza a pontot, a vesszőt, kettőspontot, pontosvesszőt, felkiáltójelet, kérdőjelet is. A beszéd paraméterei is változtathatóak bizonyos mértékig: hangmagasság, sebesség, tagoltság, stb. A Brailab PC és a Brailab CD adapter kiváló berendezés. A CD azért van a

nevében, mert ezzel az eszközzel nemcsak a magyar beszédszintézis generálódik, hanem az új típusú digitális hibrid hangoskönyvek is elolvashatók. A számítógépek mai fejlettsége már lehetővé teszi, hogy a hangrögzítés ne analóg módon egy magnóval, hanem egy hangkártyával digitálisan történjék, ezért születhetett meg a digitális hangoskönyv ötlete. A felvett anyag a könyv írott (html-szerű) változatával együtt egy CD-re kerül, s a kimondottan erre a célra kifejlesztett olvasóprogram segítségével meghallgatható, illetve a képernyőn el is olvasható.

Sok előnye mellett a Brailabok hátránya, hogy DOS alatt működnek, és Windows operációs rendszer alatt csak képernyőolvasó programokkal együtt.

➤ Braille- kijelző:



A képernyőn megjelenő szöveges információt alakítja át Braille-be. A felhasználó egy kijelzőn olvashatja el a képernyő üzeneteit. Piezoelektromos elven működik, feszültség hatására a kijelzőn a pontok helyét a Braille pontrendszernek megfelelően kiemelkedő tűskék jelzik. Magyarországon alig található belőle, méregdrága eszköz. 20, 40 és 80 karakteres változatban gyártják.

➤ Braille nyomtató:



Hat vagy nyolc pontos Braille funkcióval rendelkezik. A számítógépen begépelte szöveget alakítja át Braille-be. Olyan nyomtató is létezik, amivel kétdimenziós úgynevezett kontúrgrafika nyomtatható. Ezzel az eljárással geometriai ábrákat, képeket ismerhetnek meg a vak felhasználók.

➤ Képernyőnagyító programok:

Olyan szoftverek, amelyek segítségével a képernyőn található képeket, szövegeket fel lehet nagyítani, lehetőség van a kontraszt beállítására, egérmutató méretének megváltoztatására stb., támogatják a windows alatti beszélőrendszereket és/vagy saját, hangkártyán keresztüli felolvasó funkcióval is rendelkezhetnek.

➤ Windows alatti képernyőolvasó programok:

A képernyőolvasó programok olyan szoftverek, amelyek vak emberek számára hozzáférhetővé teszik a képernyő üzeneteit. Ez történhet hangkártyán, beszélőegységen, avagy Braille-kijelzőn keresztül, attól függően milyen eszközt részesít előnyben a felhasználó, illetve a képernyőolvasó programok milyen eszközöket támogatnak. A legismertebb programok Magyarországon: a Window-Eyes, Wintalker, Jaws for Windows, SpeakBoard szoftverek.

Magyarországon is óriási erőfeszítéseket tettek tudósok, elszánt emberek annak érdekében, hogy a számítástechnika eszközeivel folyamatosan bővítsék a vakok lehetőségeit. A számítástechnika alkalmazása vakok életminőségének javulása érdekében gyakorlatilag az első személyi számítógép megjelenése óta folyik.

### **2.3 Milyen nehézségekkel kell megküzdenie egy vak felhasználónak Magyarországon?**

A hazai beszédtechnológiai fejlesztések eredményeként a magyar látássérültek már többféle beszélőegység közül választhatnak. Jelenleg a legelterjedtebb a fentebb említett Brailab PC, melynek segítségével már közel kétezer vak felhasználó számára nyílt meg az informatika világa Magyarországon.

A karakterfelismerés minősége évről-évre javul ugyan, de még az átlagos nyomdai minőség esetén a felismertetett szövegek gondos javításától egyelőre nem lehet eltekinteni, amennyiben az így nyert szöveget más célokra is fel kívánjuk használni. Átfutó információk, számlák, újságcikkek, szórakoztató regények esetén persze ettől eltekinthetünk, mivel itt nem cél a pontos szöveg, elégséges az is, ha tudjuk, miről is szól az írott anyag. Sokan el sem tudják képzelni, micsoda nagy segítség adott esetben, ha egy postai értesítéssel nem kell átkopogtatni a szomszédba, aki talán otthon sincs. A vakok évszázados álma vált valóra a számítógép, a szkennер, a nyomtató, a CD lemez és az Internet jóvoltából. Megnyílt a lehetőség arra, hogy egy vak ember önállóan könyveket olvashasson, ezzel befogadhassa a látó világban évezredek alatt felhalmozott tudományos és kulturális kincseket, a számítógép segítségével papíron vagy elektronikusan levelezzen, barátokat szerezzen, érdeklődésének és napi igényeinek megfelelő naprakész információkhoz jusson. A Brailab beszélőeszköz család elsősorban a DOS alatt futó karakteres képernyőt használó programokat támogatja igazán.

A mindennapi munka elengedhetetlen része a szövegszerkesztő. DOS alatt futó, karakteres képernyőt használó változatból rengeteg áll rendelkezésre, igen sokfélét használnak a vakok ezek közül. Azok bizonyultak eddig a leginkább használhatónak, amelyek azonnal visszaechózzák a leütött karaktert, jól támogatják az egész sor kimondását a sor elejére állva, a szókimondást a kurzorral szavanként lépkedés közben, az ékezetes karaktereknél sem akad meg a beszéd, a vezérlőfunkciók forrógombokkal hívhatók vagy a menük pontjain

sétálgatva a beszélőrendszer valamilyen hallható információt ad a felhasználónak.

Sokkal rosszabb a helyzet az internetes kliensekkel. Ezek közül is a DOS alatt futó, karakteres képernyőt használó változatokat tudják használni olyan gyorsan, mint a látók, ilyen WEB-böngésző például a LYNX és a Nettamer programon kívül szinte nincs is, mert a többi - egyébként jól használható - programok megírásakor nem gondoltak a látássérült felhasználókra, s olyan direkt képernyőírási módokat használtak a gyorsaság érdekében, amit a beszélőrendszerek képtelenek érzékelni.

A World Wide WEB megjelenése nagymértékben elvitte a NET fejlődését a grafika és a multimédia irányába, s ez bizony nem kedvez a látássérülteknek. A mi eszközeink ugyanis nem segítik elő a grafikus megjelenítést, így nélkülöznünk kell a grafikus felületek nyújtotta előnyöket.

A Magyar Vakok és Gyengénlátók Szövetségének honlapja több más, a látássérültekkel is foglalkozó honlappal együtt igyekszik úgy kialakítani arculatát, hogy azon egy vak felhasználó is képes legyen tájékozódni. Sajnos ez a WEB-lapok döntő többségéről nem mondható el. Pedig ez nem igényelne semmi plusz pénzt, a tárigénye sem lenne sok, pusztán egy kis empátiára és odafigyelésre lenne szükség. Az is megkeseríti a vak felhasználók életét, ha a weblap három-négy oszlopban hozza az információt. A WinEyes képernyőolvasó ugyanis nem oszloptalanít, (a Jaws igen) az oszlopok első sorait ilyenkor egyben hallják. Ez az akusztikus puzzle egy idő után nagyon fárasztó tud lenni. (Ha csak két oszlop van, az ember agya ezt még képes szétválasztani különösebb nehézség nélkül.)

A másik nagy probléma: Windows-os képernyőolvasók igen drágák, 400-1200\$ közötti áron szerezhetők be, a Window Eyes az olcsóbb, a Jaws for Windows a drágább. Ezért a magyar vakok többsége ezeknek a screen reader programoknak a demó változataival kénytelenek beérni.

Gyors munkaeredményt Windows alatt nem lehet várni a vakoktól. Mivel a látók számára szimultán megjelenő információkat (a szövegek, grafikák képei,

színe, elhelyezése, mérete stb.) a vak ez esetben is csak hallás után, tehát lineárisan, időbeli egymásutániségben képes kinyerni a rendszerből s, csak amikor már az összes részinformáció a rendelkezésére áll, akkor tud dönteni, hogy mi is legyen a következő művelet, amit azután újra le kell ellenőriznie minden szempontból, ha biztosra akar menni. Ezért Windows alatt a szövegszerkesztés és a Web böngészés is rendkívül időigényes.

Ezek a nehézségek mégis eltörpülnek azokkal az előnyökkel szemben, amit a számítástechnika tud nyújtani egy vak felhasználónak. A beszélőeszközök lehetővé tették, hogy a CD lemezeken megjelenő adattárakat is használhatják a vakok, ez például nagy segítség a jogászként dolgozó látássérültek számára. Egyáltalán az, hogy a PC gépeket, mint mindennapi munkaeszközöket használni tudják a beszélőeszközeikkel, nagymértékben megnövelte a látássérültek esélyeit arra, hogy magasabb képzettséget igénylő állásokat is betölthessenek.

### **3. A hallható kijelző megvalósítási módjainak bemutatása**

A fejezet a hangos kijelző lehetséges megvalósítási módjait ismerteti a hallható ikonok, az earcon-ok, a spearcon-ok és a TTS kifejezések segítségével. (text-to-speech: olyan számítógépes program, amely felolvassa a begépett szöveget, és ez a fájl elmenthető wave formátumban)

#### **3.1 Az earcon-ok, a spearcon-ok és a hallható ikonok ismertetése**

##### **3.1.1 Bevezető**

Manapság a magas technikai fejlettség lehetővé teszi, hogy egyre kisebb méretű vizuális megjelenítőket alkalmazzunk a mobil számítógépes eszközökben. Amíg ez egyrészt igény a szokványos felhasználók körében, másrészt problémát jelent ezeknek a technikáknak a használata azon felhasználóknak, akik nem látják megfelelően ezeket a szokványos vizuális felületeket. A vakok és a gyengén látó felhasználók számára különösen fontos lenne megismerni és alkalmazni azokat a módszereket és technikákat, amik javítják a használhatóságot a nem grafikus felületek esetében, mint például a menükben.

Ha ezeket a célokat jól valósítjuk meg a különféle eszközökön, akkor a hallható menü nagyszerű megoldás lehet mindkét, a látással bíró és gyengén látó felhasználóknál is.

### 3.1.2 Az earcon-ok

Az earcon-ok (fülikon) olyan zenei motívumok, amik szisztematikusan csoportokba vannak rendezve, mint egy család, ami zenei hangokból van összeállítva.

(earcon: a menühelyek hierarchikus bemutatása, zenei hangok használatával)

Például egy rövid trombitahang lejátszható egy pontos hangmagasságban, ezt a hangmagasságot emelhetjük fél tónussal, azzal a céllal, hogy előállítsunk egy hangcsaládot, különböző, de összefüggő egytónusú earcon-okkal. Az earcon-oknak az alap építő blokkjait összeállíthatjuk bonyolultabb hangokká azzal a lehetőséggel, hogy létrehozzuk egy komplett hierarchiáját a hangoknak, amik különböző hangszínűek, hangmagasságúak és tempójúak. Ezeket a hangokat használhatjuk utasításként, hogy bemutassunk egy rangsor szerinti menüszerkezetet.

A menünek a felső szintjét például reprezentálhatjuk különböző hangszíneknek egy tónusával, (különböző zenei hangszereléssel). Minden hangszín vagy hangszerelés bemutatathat egy almenüt. Így minden elemen belül az almenüt reprezentálhatjuk egy azonos hangszínnel vagy hangszerelésnek a tónusaival. A különböző elemeket az almenüben jelezhetjük különböző hangmagassággal vagy időben különböző hangmintákkal.

A felhasználók megtanulják, hogy minden egyes hangutasítás mit ismertet és összefüggésbe hozzák az adott hanggal, ami a beszéddel egyenértékű és végül képesek lesznek használni a hangokat a saját tájékozódásukra a TTS kifejezések igénybevétele nélkül.

A kísérletben résztvevők segítették a hierarchikus információ hatásos felismerését és megértését, tehát ha az earcon-okat kombináljuk a beszéddel, akkor növeljük a menüben való tájékozódás hatásosságát és pontosságát anélkül, hogy a felhasználókat jobban terheljék.

Az előnye az earcon-ok használatának, hogy mint menüelem utasítás magában foglalja azt a lehetőséget, hogy alkalmazzuk néhány típusú

menüszerkezethez, figyelmen kívül hagyva a menüelemnek a jelentését és képesek ismertetni a hierarchiát a hangcsaládok építésénél.

Az earcon-ok azonban limitálva vannak és tetemes mennyiségű tréninget igényel a hallható elemek jelentésének a megtanulása.

Nagyon bonyolult és nehéz egy új menüelem hozzáadása az előzően létrehozott hierarchiához, továbbá a rendszerek közötti hordozhatóság sem teljesen megoldott.

Úgy tűnik, hogy az earcon-ok önkényessége egyben az erőssége és gyengesége is lehet.

### 3.1.3 A spearcon-ok

A spearcon-ok olyan rövid hangok, amiket úgy állítunk elő, hogy felgyorsítjuk a beszélt kifejezéseket (gyakran a szintetikus TTS kifejezéseket), és az így nyert hangok megértése nem tart tovább, mint az eredeti szavak felfogása.

Akusztikus vonatkozásban a spearcon az eredeti beszélt kifejezés „ujjlenyomata”.

A spearcon-ok beszéd alapú earcon-ok, melyek rövid, tömör hangutasítások, könnyen előállíthatók és hasonló feladatokra használhatók, mint az earcon-ok és a hallható ikonok. De összességében a spearcon-ok hatásosabbak, melyek mind a dinamikus vagy változtatható menükben, mind a statikus vagy fix menükben is jó eredménnyel szerepelnek.

A spearcon-ok egyedülálló specifikus menüelemek úgy, mint a hangos ikonok. Az ikon hangja és jelentése között egyértelmű akusztikus összefüggés van és nem jelentéstani vagy átvitt értelmű.

A spearcon-okat készíthetjük automatikusan úgy, hogy átalakítjuk a monitoron lévő szöveget (a menüelemre nézve) beszéddé, hasonló módon, mint a text-to-speech (TTS) programok a szöveget. Ezt a spearcon-ok úgy gyorsítják fel, hogy nem romlik a szövegérthetőség.

A hallható menüben a spearcon-okat úgy használjuk, hogy a menüelem szövegét átalakítjuk beszéddé (TTS programmal felolvastatjuk), majd felgyorsítjuk és ezt az egész folyamatot, automatizáljuk. Az így előállított spearcon-t utasításként rendeljük a menüelemhez.

Például a spearcon-ok a „Mentés”-re, a „Mentés másként”-re, vagy a „Mentés weblapként”-re, egyedülálló kulcsszavak, azonban ezek akusztikusan hasonló kezdetű hangok, ami lehetővé teszi, hogy ezeket az utasításokat együtt csoportosíthassuk.

A különböző hanghosszúságok segítenek a hallgatónak megtanulni a leképzést, és vezetőként szolgálnak a fülnek, amíg „átnézi” a menüt, mint a

vizuális menükben az ablakok jobb széle, a gördítő sáv, ami segítségül szolgál a vizuális keresésben.

Mivel a spearcon-ok és a menü elemei között a leképzés nem önkényes, így használatuk kevesebb tanulást igényel, mint a tisztán önkényes leképzéssel készült menü használata.

A spearcon-ok használatának előnye továbbá, hogy képesek újrendezni, válogatni és elemet beilleszteni vagy törölni anélkül, hogy megváltoztatnák a leképzést a különböző hangok és menüelemek között.

A spearcon-okat készíthetjük algoritmikusan, dinamikusan (bár néha a kézi beavatkozás hatékonyabb). A spearcon-ok nem egy rideg, hanem „intelligens”, rugalmas, automatizált menüszerkezetet eredményeznek.

A spearcon-ok a helymeghatározás szempontjából nem olyan hatásosak, mint a rangsor szerint rendezett ikonok, a de a hierarchikus menükeresés gyorsabb, amikor spearcon-okat használunk. Azonban a spearcon-ok még mindig több egyértelmű leképzést szolgáltatnak a hang és a menüelem között, mint az earcon-ok, továbbá nagyobb tartalmi területet fednek le, rugalmasabbak, mint a hallható ikonok.

A spearcon-ok könnyen megtanulhatóak, ami csökkenti a felhasználó frusztráltságát, növeli a használhatóságot, továbbá ez a fajta felület-kiemelés elfogadhatóbb az eszközgyártók számára is.

### 3.1.4 A hallható ikonok

Az earcon-ok egyik alternatíváját hallható ikonoknak nevezzük. Ezek esetében rendszerint nem zenei hangokat alkalmazunk, hanem olyan hangokat, melyek hasonlítanak ahhoz a dologhoz, amit bemutatnak.

A hallható ikonokat zaj előállításával ábrázolják, a zaj asszociál azzal a dologgal, amit képvisel. Ebben az esetben egy hangos menünél a hallható ikon úgy hangzik, mint a menüelem. A hallható ikonokat azzal a céllal használják, hogy a leképzés nagyon egyértelmű legyen, így az elsajátíthatóság csaknem azonnali.

(ha az elem egy kutya, és ezt egy kutyaugatással illusztráljuk, akkor ennek leképzése nem igényel különösebb tanulást)

Ebből a szempontból hasznosak a hallható ikonok. Sajnos a leképzés közvetlensége tekintélyesen eltér. Például ha az írógép hangjával ismertetjük a menüelemet, ami a „Dokumentum nyomtatása”, ez a leképzés meglehetősen közvetlen, de nem pontos.

Egy hallható ikon reprezentálhat például egy nyomtatót, ami hangozhat úgy, mint egy pontmátrix nyomtató, vagy írógép. A lényeg az, hogy a felhasználó számára az adott hang egyértelműen utaljon arra a tárgyra, eszközre, amit be szeretnénk mutatni, akár egy grafikus ikon esetében. Tisztán a direkt hasonlóság a hallható ikon és a bemutatott tárgy között úgy változhat, mint egy vizuális ikonnál.

Azokon a területeken, ahol a hallható menük hasznosak, mint például a mobil eszközöknél és asztali számítógépeknél gyakran előfordul, hogy nem áll valós hang a rendelkezésre, hogy ismertesse a menüelemet. Például egy fájl törléséhez természetes hangot nem tudunk társítani.

Így néhány esetben inkább képletes ismertetést szükséges használni, mint a szándékosan közvetlen ikonos illusztrációt.

A leképzés teljesen önkényessé válhat, ami kiterjedt tanulást igényel és megnyitja a lehetőséget arra, hogy az előre kitalált hangutasítások jelentésének a megértésében zavart keltsen.

Néhány esetben, ez a csak ikonszerű bemutatás hordozhat átvitt értelmű jelentést is. Ebből adódóan az earcon-ok, vagy a hallható ikonok jelentős teljesítményfórlénnyel bírnak a tájékozódásban és a menü elem felismerésében a hallható menüfelületekkel szemben. Ezért a valódi hallható ikonok limitált használhatóságúak, különösen a hallható menüalkalmazásokban.

### 3.1.5 A hallható menük

Az olyan alkalmazásokban, mint a változatos mobiltelefon alapú operációs rendszerekben és az asztali számítógépes környezetekben a menübeállítások reprezentálhatók hangok útján.

Általánosságban elmondható, hogy a menüelemek szöveg címkékből vannak konvertálva, bevált kifejezéseket, automata beszéd szintézist vagy TTS programot használnak.

Gyakran a felhasználó a hallható menüben úgy tájékozódik, hogy megnyomja a fel és a le navigációs billentyűt és hallgatja a szemléltető TTS kifejezéseket, akkor a select és a return billentyűvel belép a kiválasztott menüpontba ahelyett, hogy olvasná a menüelem szövegét. A felhasználó, amikor kiválasztja a menüelemet, hallja a hangutasítást és a TTS kifejezést. Bizonyos menükben mindig van TTS kifejezés, más rendszerben a felhasználó kijelöli a soron következő elemet, vagy átmegy a következő menüpontra anélkül, hogy hallaná TTS kifejezést.

Minden menüelem meghallgatása időigényes, és a gyors és hatékony mozgás megvalósítása a menüszerkezeten keresztül bonyolult feladat.

Nehézséget okozhat a kezelés tudatossága a menürangsorban egy egyszerű mozgásnál, hogy tisztában legyünk azzal, melyik menü vagy almenü aktív éppen. Nehéz úgy elkészíteni egy hallható menüfelületet, - figyelembe véve a felhasználó memóriaterhelését, a sebességet és a pontosságot - hogy a hallható menüszerkezet megtanulása könnyű és általánosan jól használható legyen.

Bebizonyosodott, hogy a menü navigációban a spearcon-ok teljesítményben felülmúlják a hallható ikonokat és az earcon-okat.

### **3.1.6 Az earcon-ok és a hallható ikonok gyakorlati használata**

Ezek az új technikai eszközök javítják a teljesítményt és a használhatóságot a menü alapú felületeknél. Segítségükkel egy jobban áttekinthető felületet készíthetünk, ráadásul egy igen széles felhasználói kör számára. Ezeken a felületeken az alkalmazások sokkal szélesebb skálán tudnak mozogni az információáramlás és a kommunikáció tekintetében, hiszen a vizuális információn túl egy másik érzékelésünket, a hallást is kiaknázzuk az adatok továbbításában.

Az earcon-ok rövid, tömör zenei kivonatok, melyek néhány hangjegyből és csengésből állnak. A hangok üteme szisztematikusan változik, így hordoz információt. Ezzel a módszerrel felépítjük a hangoknak egy családját, ami a rendszerezett információszerkezeteknek a sajátosságait jelzi a felhasználó számára. Earcon-okat használva a menüszerkezetek kialakításánál megállapíthatjuk, hogy plusz információt adnak a menühöz a felhasználói felületen. Segítségükkel a felhasználó könnyebben eligazodik a menürendszerben, akusztikusan is tájékoztatást kap arról, hogy hol van éppen a könyvtárfában. Az earcon-okat ezért számos menürendszerben használják, mint például a grafikus felületű alkalmazásokban, mobiltelefonokban és a telefon alapú felületek esetében.

A grafikus felhasználói felületeken is fejleszthetők a menük earcon-okkal. Megelőzhetjük például, hogy a felhasználó kiválasszon egy rossz menüelemet, vagy lecsússzon a kiválasztott elemről.

Az earcon-ok egy másik ajánlott felhasználási lehetősége a felgyorsított beszéd alapú felületek segítése, magában foglalva ezek tervezését a látássérült felhasználók számára.

Earcon-okat használhatunk a járművek kijelzőjénél, ahol a sofőr vizuális csatornáit a gépjármű vezetése foglalja le. Ezekben az alkalmazásokban a hang azt is közölheti a felhasználó számára, hogy mit tartalmaz a menüelem, nem csak azt, hogy hol helyezkedik el a menüben.

### 3.1.7 Az earcon-okkal és a hallható ikonokkal kapcsolatos kérdések

Felmerül bennünk a kérdés, mikor alkalmazzunk earcon-okat, és mikor hallható ikonokat, hogy jól használható és alakítható, valamint hallható menüfelület hozzunk létre.

Mint megállapítottuk az earcon-ok képesek a hely megmutatására a menüben, valamint a menüelem tartalmát is reprezentálják.

Ezek informatívabbak lehetnek, mint a hallható ikonok, amelyek abban hatékonyak, hogy jobban megmutatják az elem tartalmát, mint a rangsor szerinti pozíciót.

Az önkényes leképzés alkalmazásával, tulajdonképpen néhány beállítás, elképzelés megvalósítható earcon-okkal, miután a menük fejlesztéséhez a hallható ikonok előállítása bonyolult. Különösen a számítógép felületeknél, ahol nincsenek valós hangok. Azonban az önkényes leképzés használatával az earcon-ok több tanulást igényelnek, de kifejezőbbek lehetnek, jobb hatást eredményezhetnek. Két rendszer között az átvitel korlátozott lehet, ha eltérő earcon nyelveket használnak.

A rendszermérnökök perspektívájából a menük, amik earcon-okat, vagy hallható ikonokat használnak, rugalmatlanok abból a szempontból, hogy hangokkal nem támogatott a menürendszer vagy a menü elemeinek módosítása. Ha a menüt vagy a menüelemet bővíteni kell, új hangok előállítása szükséges.

A rangsor szerinti earcon-ok bővítése automatikusan kezelhető új menü vagy menüelem hozzáadásával. Például ha hozzáadunk egy elemet a menü aljához, akkor a soron következő hangot vagy tempót rendeljük az elemhez a beállítási listából, amit az earcon-nak korábban deklaráltunk.

Az earcon-okat előrelátóan úgy kell létrehozni, hogy megfelelően nagy változatosságot mutassanak a menüben, valamint emellett a menü fejlődésének lehetőségét is biztosítani kell. Ez egy nehezen megvalósítható feladatot jelenthet, különösen a sokrétű rendszerek esetében, vagy ha hosszú távú felhasználásról van szó. Nehéz olyan hangokat találni ugyanis, ami széleskörűen használható: különböző rendszerekben és hosszútávon is.

Nagy problémát jelent az is, amikor egy menüelem belép a menü közepére. Például ha az első menüelem a fájl listában C-vel kezdődik, és ha a menülistába beillesztünk egy B-vel kezdődő fájlt, akkor az a fájl listában automatikusan a C elé ugrik.

(ha egyszerre sok elem kerül kijelölésre, akkor az intelligens menürendszerekben ezek fölre ugrik egy tájékoztató buborék)

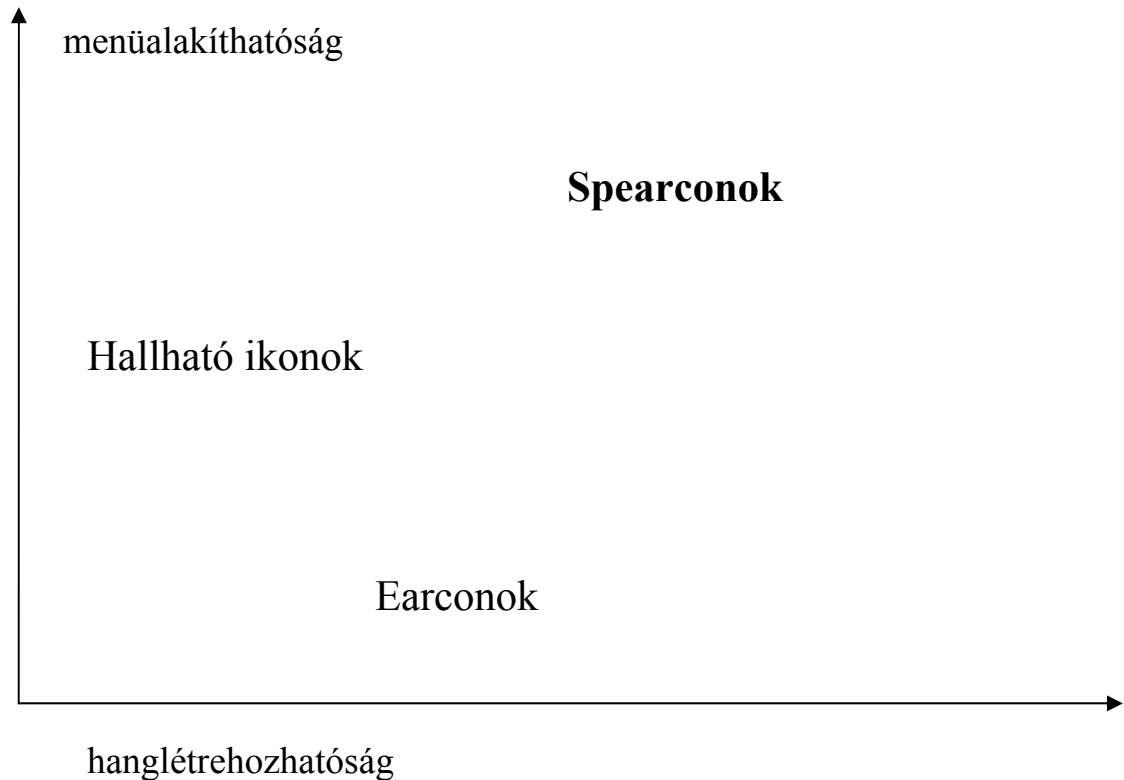
Az earcon-ok ezt a szituációt rosszul kezelik, továbbá a menü újraválogatását és rendszerezését is nehezen oldják meg. Ilyenkor a rangsor szerinti earcon-oknak újra kell rendeződniük, ami csökkenti szerepüket az értelmi összefüggés meglátásában. Az eddig elsajátított earcon-ok alatti elrendezést az új menüelemekre újra meg kell tanulni.

A hallható ikonok rugalmasabbak a menüelem beillesztésére és újrendezésére, de minden új elem létrehozásához kézi beavatkozás szükséges. Ez egyértelmű probléma a dinamikus rendszerekben.

(az is elvárás, hogy az ikonszerű hangot megtaláljuk az új elemre is, tehát, az ikon hangja az új elemre is értelmezve legyen)

Az összegzett eredményeket a 3.1-es ábra szemlélteti az alábbi dimenziókban:

- hanglétrehozhatóság
- menüalakíthatóság



**3.1 ábra. Az earcon-ok, a hallható ikonok és a spearcon-ok elhelyezkedése a grafikonon a hanglétrehozhatóság és a menüalakíthatóság szempontjából.**

Mindkét dimenzióban, sem az earcon-ok, sem a hallható ikonok nem szerepelnek magas rátával.

Az optimális megoldás, hogy az akusztikus utasítások úgy hangozzanak, hogy:

- egyszerűen és automatikusan középre generálódjanak
- kevésbé önkényes leképzést szolgáltatson, mint az earcon-ok
- egy szélesebb tartományt fedjen le a menütartalomra nézve, mint a hallható ikonok
- eléggye rugalmas legyen ahhoz, hogy támogassa az újrendezést, az újraválogatást, a közbeékelést és a menüelemek törlését

Tehát, a cél az ilyen hangok létrehozása, melyekkel elérhetjük a sebesség és a pontosság növekedését a menü kiválasztásában.

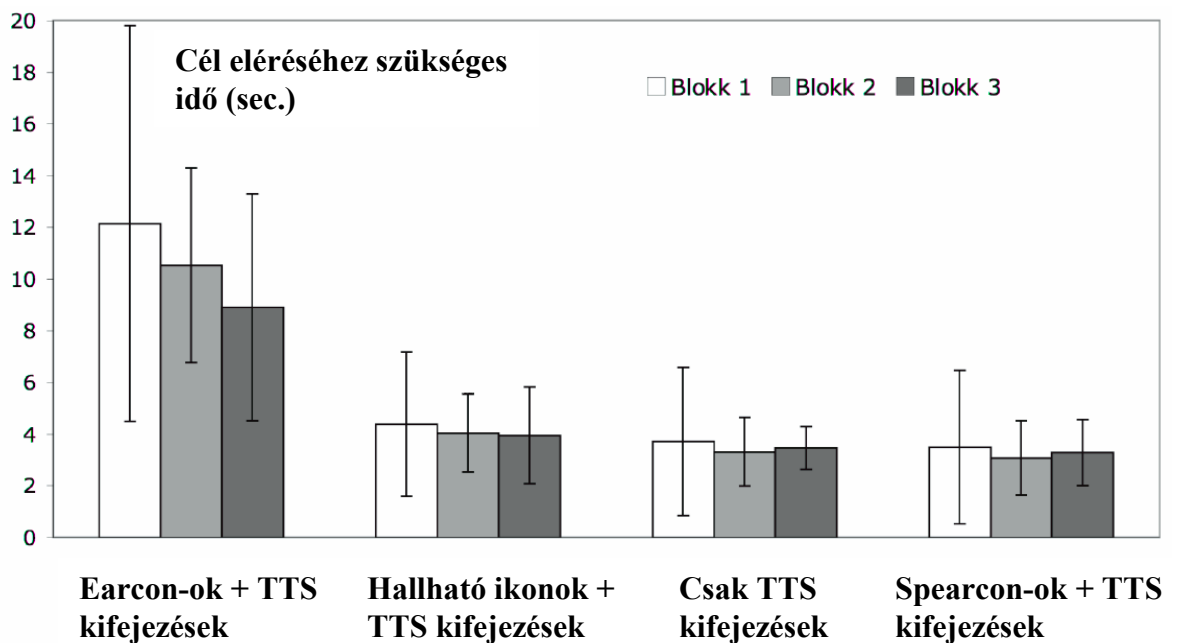
### 3.1.8 A lehetséges megoldások elemzése

A menüben való navigációhoz négy típusú hallható ábrázolást használhatunk:

- csak beszéd
- rangsor szerinti earcon-ok
- hallható ikonok
- és az új típus, amit spearcon-oknak hívunk

Spearcon-okat úgy állíthatunk elő, hogy felgyorsítjuk a beszélt kifejezéseket annak szem előtt tartásával, hogy a beszéd érthetősége ne romoljon.

Az alábbi vizsgálat során összehasonlítottuk a felhasználó által igényelt keresési időt és az elért pontosságot (3.2 és 3.3-as ábrák).



3.2 ábra. A hallható kijelző típusok.

| Hangutasítás típusa                                   | Cél eléréséhez szükséges átlagos idő (SD) sec. | Átlagos pontosság (SD) % |
|-------------------------------------------------------|------------------------------------------------|--------------------------|
| Spearcon-ok + TTS kifejezések                         | 3.28 (.517)                                    | 98.1 (1.5)               |
| Csak TTS kifejezések                                  | 3.49 (.486)                                    | 97.6 (2.0)               |
| Hallható ikonok + TTS kifejezések                     | 4.12 (.587)                                    | 94.7 (3.5)               |
| Rangsor szerint rendezett earcon-ok + TTS kifejezések | 10.52 (11.87)                                  | 94.2 (5.4)               |

**3.3 ábra. A különböző hangutasításokkal mért eredmények.**

A kapott eredmények szerint a spearcon-ok és a „csak beszédhangok használata” gyorsabb és pontosabb menühasználatot eredményeznek, mint a hallható ikonok és a rangsor szerinti earcon-ok. Ezen túl a gyakorlatban jelentős eltérés mutatkozott a keresési időben is a különböző típusú hallható reprezentációknál.

Az earcon-ok, a hallható ikonok és a spearcon-ok mind azt a célt szolgálják, hogy bemutassanak egy egyéni, sajátos elemet.

Az eszközök menüjének használata fokozható a hallható menük megtanulásával, tehát a felhasználó elsajátítja a leképzést a hangok és a menüelemek között, így a felület használhatósága jelentősen javul. Ezt a tanulást összehasonlítva az earcon-ok elsajátíthatóságával és a spearcon-okkal (ami tömörített beszéd), úgy találtuk, hogy a spearcon-ok teljesítményben felülmúlják az earcon-okat.

Jelentős különbség van a megtanulhatóságukban is. A spearcon-ok könnyebben elsajátíthatóak, megtanulásuk folyamata sokkal egyszerűbb ugyanazon alkalmazásnál, mint ami earcon-t használ.

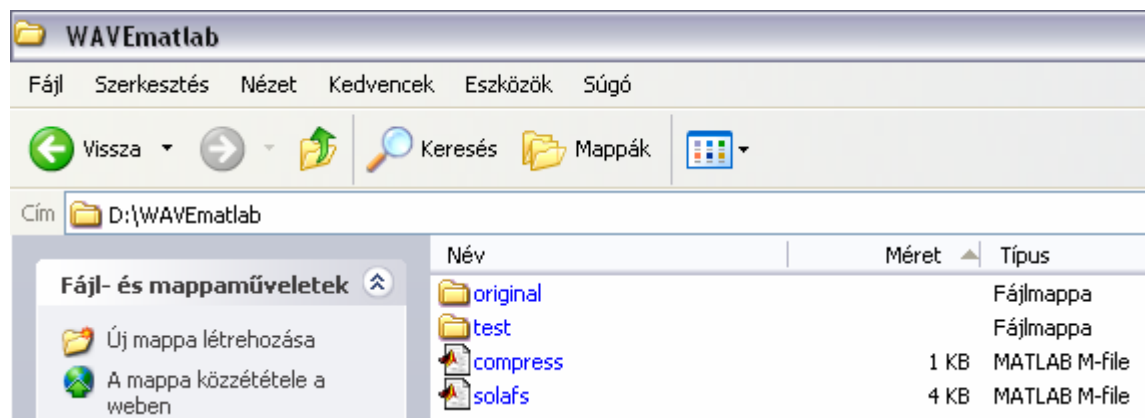
Ezek az eredmények azt mutatják, hogy a spearcon-ok hatékonyabban használhatók a menü alapú felületeken, mint a korábban alkalmazott hallható utasítások. Segítségükkel jobb teljesítményt, pontosságot, valamint jobban alakítható menüfelépítést tudunk létrehozni.

## 4. A tömörített beszédminták előállítása

A fejezet a wave fájlok spearcon-ná tömörítéséről szól, melyet a MATLAB programmal végzek.

### 4.1 A MATLAB program használatának leírása

Létre kell hozni egy mappát a helyi meghajtón. Ez a mappa bárhol lehet és tetszés szerint elnevezhető (Pl.:WAVEmatlab). Majd ezen mappába bele kell másolni a compress.m és a solafs.m fájlokat továbbá szintén ugyan itt el kell helyezni egy original és egy test nevű mappát is (4.1 ábra).



4.1 ábra. A MATLAB program használatához szükséges mappák és fájlok elrendezése.

Az original mappába kell másolni a tömöríteni kívánt wave fájlokat, majd a program a tömörített hangmintákat a test mappában helyezi el.

A wave fájlok elejéről és végéről a csendet wave editorral el kell távolítani. Ha ezt nem végezzük el, akkor a tömörített hangminta túl gyors lesz és így az érthetőség jelentősen csökken.

Fontos megjegyeznem, hogy a program csak mono wave fájlal működik, tehát sztereóval a következő hibaüzenetet adja:

??? Error using ==> solafs

X must be a single-row vector

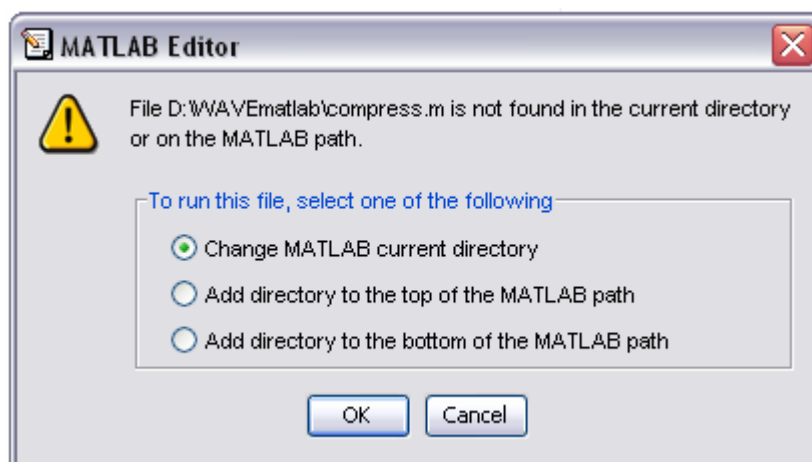
Error in ==> compress at 14

```
xc = solafs(x',ratio,win);
```

Ha létrehoztuk a szükséges mappákat és végrehajtottuk a megfelelő beállításokat, nincs más teendőnk, mint hogy elindítsuk a MATLAB-ot! A File – Open... menüponttal nyissuk meg a compress.m és a solafs.m fájlokat, melyek a MATLAB Editor ablakában jelennek meg. Be kell állítani az általunk létrehozott mapparendszer (jelen esetben a WAVEmatlab) elérési útját a MATLAB ablak File – Set Path... menüpontja alatt.

Előfordulhat, hogy a program nem ismeri fel a beállított mapparendszert, és nem végzi el a tömörítést. Ekkor a MATLAB ablak File – Set Path... menüpontjánál beállított elérési utat töröljük ki.

Ha ez megtörtént, akkor futtassuk a már megnyitott compress.m fájlt a MATLAB Editor ablakának Debug – Run menüpontjával. Ekkor a program a következő üzenetet adja (4.2 ábra):



4.2 ábra. A hibaüzenet.

Válasszuk a Change MATLAB current directory lehetőséget, majd a program elvégzi a tömörítést és a test mappában, elhelyezi a tömörített beszédmintákat.

Ha a program felismeri a MATLAB ablak File – Set Path... menüpontjánál beállított elérési utat, akkor futtassuk a már megnyitott compress.m fájlt a MATLAB Editor ablakának Debug – Run menüpontjával. A program ebben az esetben is a test mappában helyezi el a tömörített beszédmintákat.

## 4.2 A compress.m és a solafs.m fájlok leírása:

➤ a compress.m fájl felépítése:

```
fs = 16e3;  
nbits = 16;  
l = 0.1;  
c = 2;  
win = 200;  
  
y = dir('./original/*.wav');  
  
for i = 1:length(y)  
  
    [x,fs,nbits] = wavread(sprintf('./original/%s',y(i).name));  
    r = length(x)/fs/l;  
    ratio = c*log10(r)+1  
    xc = solafs(x',ratio,win);  
    wavwrite(xc,fs,nbits,sprintf('./test/%s',y(i).name));  
  
end
```

➤ a solafs.m fájl felépítése:

```
function Y = solafs(X, F, W, Wov, Kmax, Wsim, xdecim, kdecim)
% Y = solafs(X, F, W, Wov, Kmax, Wsim, xdec, kdec) Do SOLAFS timescale
mod'n
% Y is X scaled to run F x faster. X is added-in in windows
% W pts long, overlapping by Wov points with the previous output.
% The similarity is calculated over the last Wsim points of output.
% Maximum similarity skew is Kmax pts.
% Each xcorr calculation is decimated by xdecim (8)
% The skew axis sampling is decimated by kdecim (2)
% Defaults (for 22k) W = 200, Wov = W/2, Kmax = 2*W, Wsim=Wov.
% Based on "The SOLAFS time-scale modification algorithm",
% Don Hejna & Bruce Musicus, BBN, July 1991.
% 1997may16 dpwe@icsi.berkeley.edu $Header:
/homes/dpwe/matlab/dpwebox/RCS/solafs.m,v 1.3 2006/04/09 20:10:20 dpwe
Exp $
% 2006-04-08: fix to predicted step size, thanks to Andreas Tsiartas

if (nargin < 3) W = 200; end
if (nargin < 4) Wov = W/2; end
if (nargin < 5) Kmax = 2 * W; end
if (nargin < 6) Wsim = Wov; end
if (nargin < 7) xdecim = 8; end
if (nargin < 8) kdecim = 2; end

Ss = W - Wov;

if(size(X,1) ~= 1) error('X must be a single-row vector'); end;

xpts = size(X,2);
ypts = round(xpts / F);
```

```

Y = zeros(1, ypts);

% Cross-fade win is Wov pts long - it grows
xfwin = (1:Wov)/(Wov+1);

% Index to add to ypos to get the overlap region
ovix = (1-Wov):0;
% Index for non-overlapping bit
newix = 1:(W-Wov);
% Index for similarity chunks
% decimate the cross-correlation
simix = (1:xdecim:Wsim) - Wsim;

% prepad X for extraction
padX = [zeros(1, Wsim), X, zeros(1,Kmax+W-Wov)];

% Startup - just copy first bit
Y(1:Wsim) = X(1:Wsim);

xabs = 0;
lastxpos = 0;
lastypos = 0;
km = 0;
for ypos = Wsim:Ss:(ypts-W);
    % Ideal X position
    xpos = F * ypos;
    % disp(['xpos=',num2str(xpos),' ypos=',num2str(ypos)]);
    % Overlap prediction - assume all of overlap from last copy
    kmpred = km + ((xpos - lastxpos) - (ypos - lastypos));
    lastxpos = xpos;
    lastypos = ypos;
    if (kmpred <= Kmax && kmpred >= 0)
        km = kmpred; % no need to search
    end
end

```

```

else
    % Calculate the skew, km
    % .. by first figuring the cross-correlation
    ysim = Y(ypos + simix);
    % Clear the Rxy array
    rxy = zeros(1, Kmax+1);
    rxx = zeros(1, Kmax+1);
    % Make sure km doesn't take us backwards
    %Kmin = max(0, xabs-xpos);
    Kmin = 0;
    % actually, this sounds kinda bad. Allow backwards for now
    for k = Kmin:kdecim:Kmax
        xsim = padX(floor(Wsim + xpos + k + simix));
        rxx(k+1) = norm(xsim);
        rxy(k+1) = (ysim * xsim');
    end
    % Zero the pts where rxx was zero
    Rxy = (rxx ~= 0).*rxy./(rxx+(rxx==0));
    % Local max gives skew
    km = min(find(Rxy == max(Rxy))-1);
end
xabs = xpos+km;
% disp(['ypos = ', int2str(ypos), ', km = ', int2str(km), '(base = ', int2str(ypos-
xabs), ')']);
%
% subplot(311);
% plot(ysim);
% subplot(312);
% plot(padX(Wsim + xpos + ((1-Wsim):Kmax)))
% subplot(313);
% plot(Rxy);
% pause;

```

```

% Cross-fade some points
Y(ypos+ovix) = ((1-xfwin).*Y(ypos+ovix)) +
(xfwin.*padX(floor(Wsim+xabs+ovix)));
% Add in remaining points
Y(ypos+newix) = padX(floor(Wsim+xabs+newix));
end

```

## **5. A tömörített hangminták minőségi vizsgálata**

A fejezet a hangminták felvételéről, szerkesztéséről és a MATLAB-al spearcon-ná alakított wave fájlok tömörítési arányának méréséről, spektrális elemzéséről és minőségi vizsgálatáról szól.

### **5.1 A hangminták elkészítése, editálása**

A magyar nyelvű felvételek a Széchenyi István Egyetem Távközlési Tanszékének Audio és Videotechnikai oktató és kutatólaboratóriumában készültek. A felvételhez Sennheiser BSN-ME64 kardiod mikrofont és LUNA II audio kártyát használtam és 16 Bit 48 kHz PCM Wave formátumban rögzítettem. Ezek után a Wave fájlból a hangmintákat egyenként kivágtam, elmentettem és a csendet a fájlok elejéről és végéről eltávolítottam.

A beszédminták tömörítési arányának mérése céljából a 16 Bit 48 kHz PCM Wave formátumból készítettem 16 Bit 44.1 kHz, 16 Bit 22.05 kHz, 8 Bit 44.1 kHz-es PCM Wave formátumokat is.

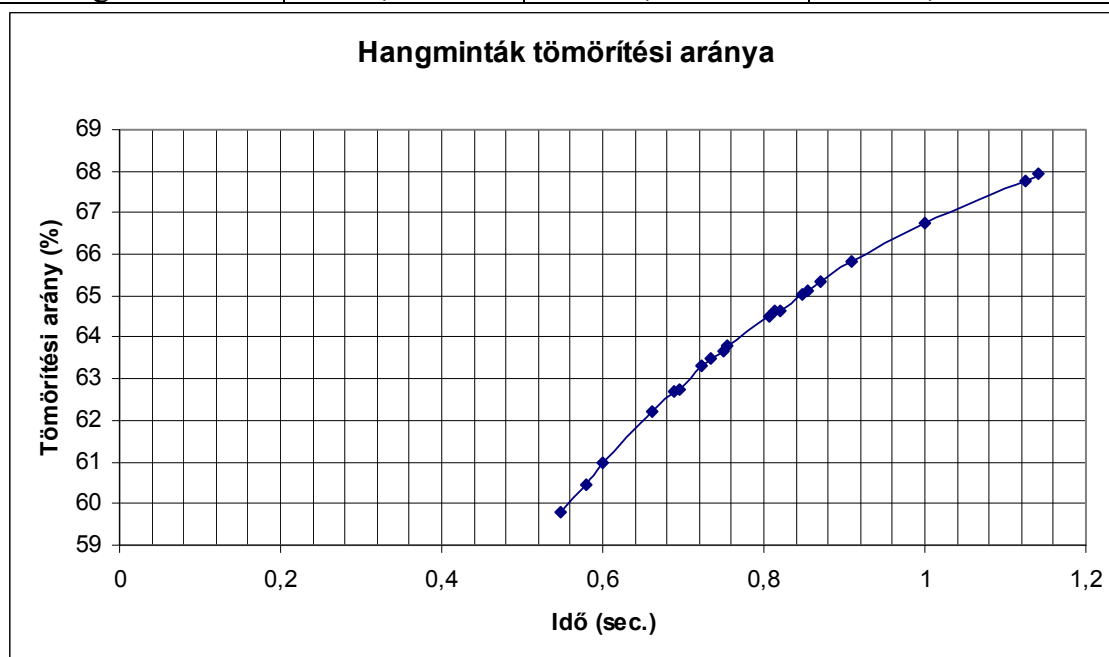
### **5.2 A hangminták tömörítési arányának vizsgálata**

A tömörítési arányt úgy mérem a különböző formátumokon, hogy megvizsgálom az eredeti és a tömörített wave fájl időbeli hosszát és kiszámítom, hogy a tömörített beszédminta hány százalékkal kisebb az eredetinél, majd az eredményeket táblázatban, grafikonon és diagramon is megjelenítem.

Az eredmények táblázatban és grafikonon megjelenítve:  
(a grafikonon az eredeti hangfájl időbeli hosszához tartozó tömörítési arányt ábrázolom, így nyomon követhető, hogy a kettő közötti kapcsolat lineáris-e vagy sem)

➤ 16 Bit 48 kHz PCM Wave: 5.1 ábra.

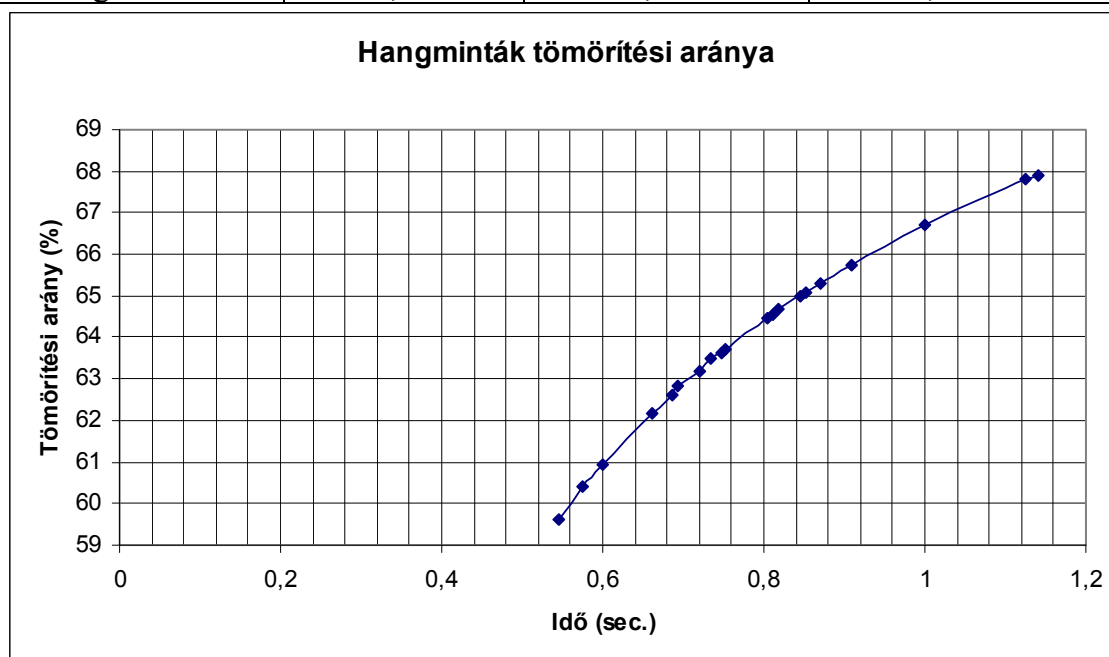
| Beszédminták     | Eredeti (sec.) | Tömörített (sec.) | Tömörítési arány (%) |
|------------------|----------------|-------------------|----------------------|
| Email            | 0,547          | 0,22              | 59,78062157          |
| Word             | 0,579          | 0,229             | 60,44905009          |
| Excel            | 0,6            | 0,234             | 61                   |
| Törlés           | 0,662          | 0,25              | 62,23564955          |
| Mentés           | 0,689          | 0,257             | 62,69956459          |
| Keresés          | 0,695          | 0,259             | 62,73381295          |
| Levelező         | 0,722          | 0,265             | 63,29639889          |
| Startmenü        | 0,734          | 0,268             | 63,48773842          |
| Mozgatás         | 0,749          | 0,272             | 63,68491322          |
| Nyomtatás        | 0,754          | 0,273             | 63,79310345          |
| Sajátgép         | 0,806          | 0,286             | 64,51612903          |
| Megnyitás        | 0,814          | 0,288             | 64,61916462          |
| Másolás          | 0,82           | 0,29              | 64,63414634          |
| Böngésző         | 0,847          | 0,296             | 65,05312869          |
| Letöltés         | 0,854          | 0,298             | 65,10538642          |
| Bezárás          | 0,871          | 0,302             | 65,3272101           |
| Leállítás        | 0,91           | 0,311             | 65,82417582          |
| Táblázatkezelő   | 1,001          | 0,333             | 66,73326673          |
| Mentés másként   | 1,126          | 0,363             | 67,76198934          |
| Szövegszerkesztő | 1,142          | 0,366             | 67,95096322          |



5.1 ábra.

➤ 16 Bit 44.1 kHz PCM Wave: 5.2 ábra.

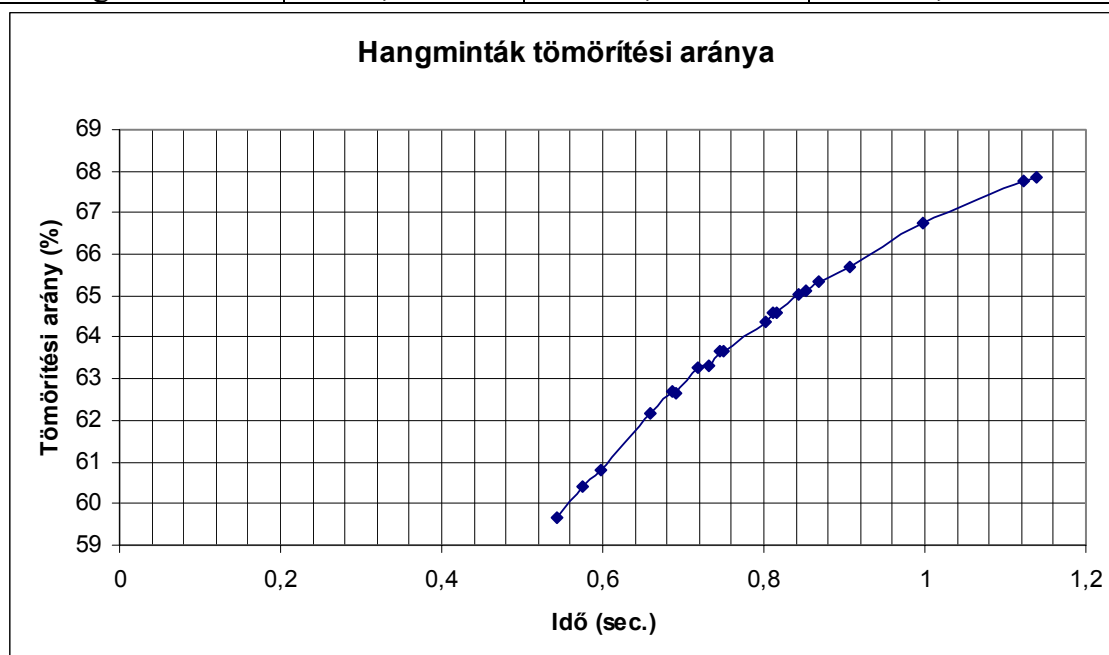
| Beszédminták     | Eredeti (sec.) | Tömörített (sec.) | Tömörítési arány (%) |
|------------------|----------------|-------------------|----------------------|
| Email            | 0,545          | 0,22              | 59,63302752          |
| Word             | 0,576          | 0,228             | 60,41666667          |
| Excel            | 0,599          | 0,234             | 60,93489149          |
| Törlés           | 0,661          | 0,25              | 62,1785174           |
| Mentés           | 0,687          | 0,257             | 62,59097525          |
| Keresés          | 0,694          | 0,258             | 62,82420749          |
| Levelező         | 0,72           | 0,265             | 63,19444444          |
| Startmenü        | 0,734          | 0,268             | 63,48773842          |
| Mozgatás         | 0,748          | 0,272             | 63,63636364          |
| Nyomtatás        | 0,752          | 0,273             | 63,69680851          |
| Sajátgép         | 0,805          | 0,286             | 64,47204969          |
| Megnyitás        | 0,812          | 0,288             | 64,5320197           |
| Másolás          | 0,818          | 0,289             | 64,66992665          |
| Böngésző         | 0,845          | 0,296             | 64,9704142           |
| Letöltés         | 0,853          | 0,298             | 65,06447831          |
| Bezárás          | 0,87           | 0,302             | 65,28735632          |
| Leállítás        | 0,908          | 0,311             | 65,74889868          |
| Táblázatkezelő   | 1              | 0,333             | 66,7                 |
| Mentés másként   | 1,125          | 0,362             | 67,82222222          |
| Szövegszerkesztő | 1,14           | 0,366             | 67,89473684          |



5.2 ábra.

➤ 16 Bit 22.05 kHz PCM Wave: 5.3 ábra.

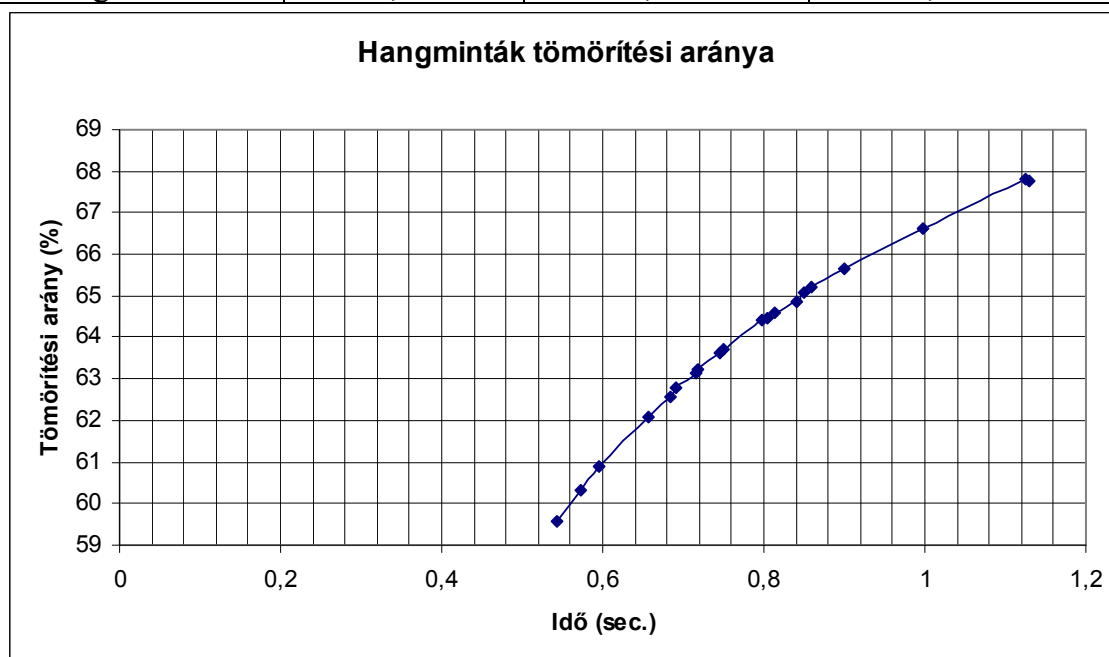
| Beszédminták     | Eredeti (sec.) | Tömörített (sec.) | Tömörítési arány (%) |
|------------------|----------------|-------------------|----------------------|
| Email            | 0,543          | 0,219             | 59,66850829          |
| Word             | 0,576          | 0,228             | 60,41666667          |
| Excel            | 0,597          | 0,234             | 60,8040201           |
| Törlés           | 0,658          | 0,249             | 62,15805471          |
| Mentés           | 0,686          | 0,256             | 62,68221574          |
| Keresés          | 0,691          | 0,258             | 62,66280753          |
| Levelező         | 0,719          | 0,264             | 63,28233658          |
| Startmenü        | 0,731          | 0,268             | 63,3378933           |
| Mozgatás         | 0,746          | 0,271             | 63,67292225          |
| Nyomtatás        | 0,751          | 0,273             | 63,64846871          |
| Sajátgép         | 0,803          | 0,286             | 64,38356164          |
| Megnyitás        | 0,811          | 0,287             | 64,61159063          |
| Másolás          | 0,816          | 0,289             | 64,58333333          |
| Böngésző         | 0,844          | 0,295             | 65,04739336          |
| Letöltés         | 0,852          | 0,297             | 65,14084507          |
| Bezárás          | 0,868          | 0,301             | 65,32258065          |
| Leállítás        | 0,907          | 0,311             | 65,71113561          |
| Táblázatkezelő   | 0,998          | 0,332             | 66,73346693          |
| Mentés másként   | 1,123          | 0,362             | 67,76491541          |
| Szövegszerkesztő | 1,139          | 0,366             | 67,8665496           |



5.3 ábra.

➤ 8 Bit 44.1 kHz PCM Wave: 5.4 ábra.

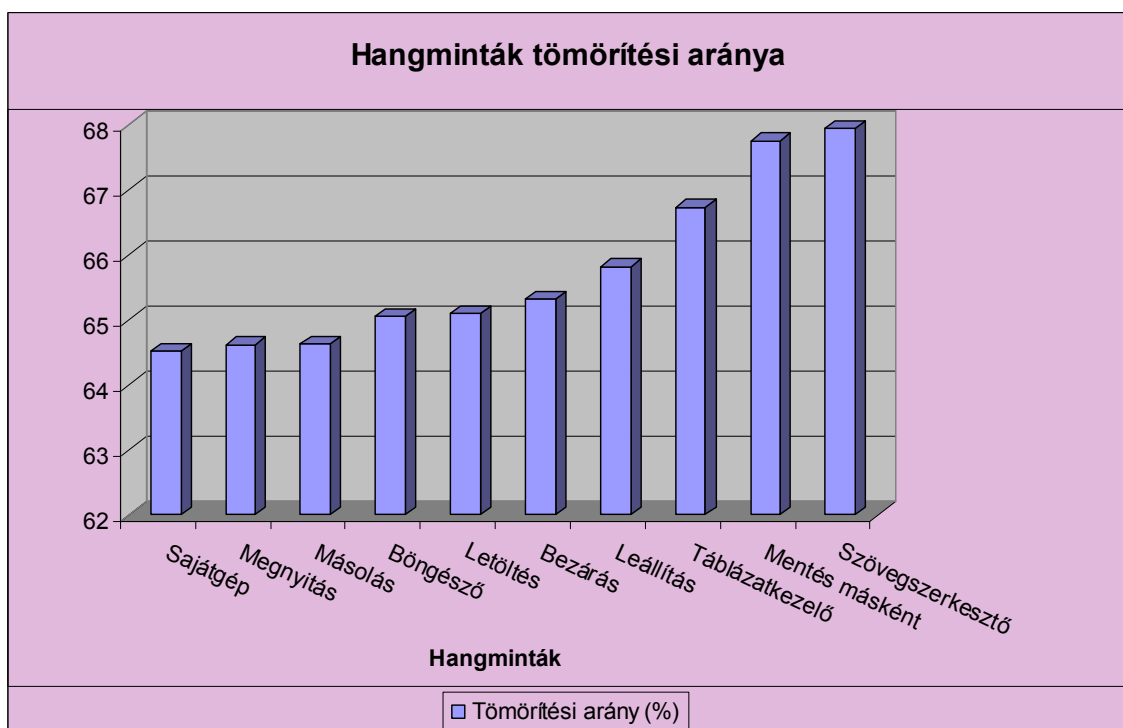
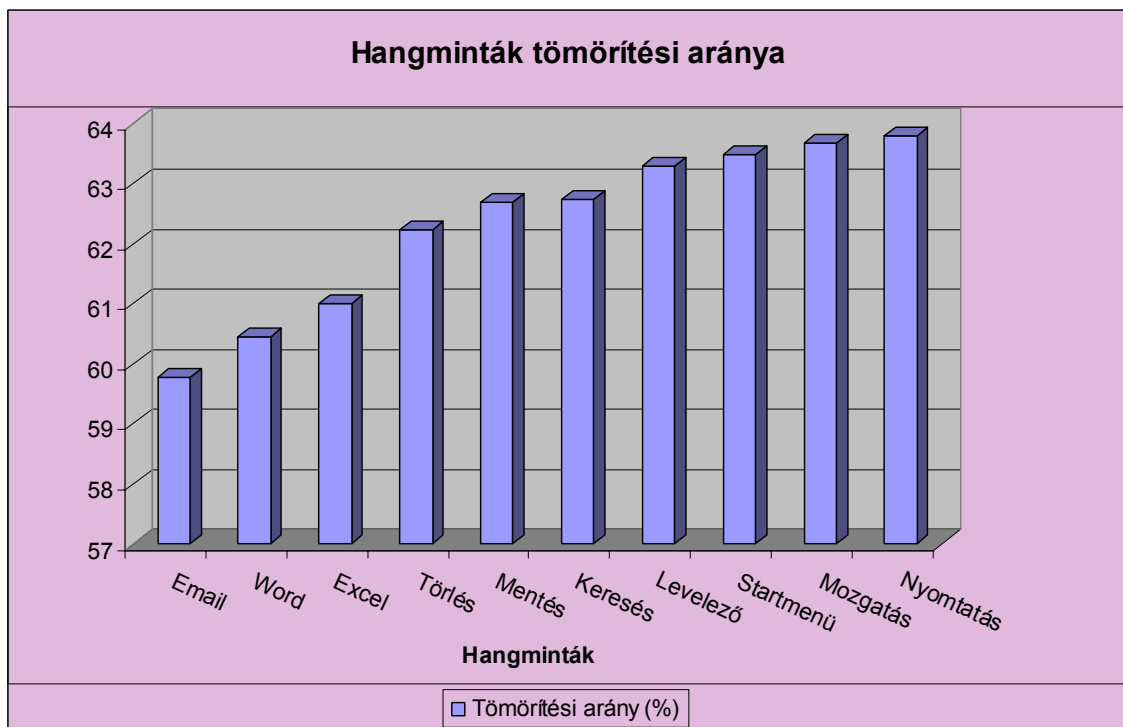
| Beszédminták     | Eredeti (sec.) | Tömörített (sec.) | Tömörítési arány (%) |
|------------------|----------------|-------------------|----------------------|
| Email            | 0,544          | 0,22              | 59,55882353          |
| Word             | 0,572          | 0,227             | 60,31468531          |
| Excel            | 0,596          | 0,233             | 60,90604027          |
| Törlés           | 0,657          | 0,249             | 62,10045662          |
| Mentés           | 0,684          | 0,256             | 62,57309942          |
| Keresés          | 0,691          | 0,257             | 62,80752533          |
| Levelező         | 0,716          | 0,264             | 63,12849162          |
| Startmenü        | 0,718          | 0,264             | 63,23119777          |
| Mozgatás         | 0,745          | 0,271             | 63,62416107          |
| Nyomtatás        | 0,75           | 0,272             | 63,73333333          |
| Sajátgép         | 0,798          | 0,284             | 64,41102757          |
| Megnyitás        | 0,805          | 0,286             | 64,47204969          |
| Másolás          | 0,813          | 0,288             | 64,57564576          |
| Böngésző         | 0,84           | 0,295             | 64,88095238          |
| Letöltés         | 0,85           | 0,297             | 65,05882353          |
| Bezárás          | 0,86           | 0,299             | 65,23255814          |
| Leállítás        | 0,9            | 0,309             | 65,66666667          |
| Táblázatkezelő   | 0,998          | 0,333             | 66,63326653          |
| Mentés másként   | 1,124          | 0,362             | 67,79359431          |
| Szövegszerkesztő | 1,13           | 0,364             | 67,78761062          |



5.4 ábra.

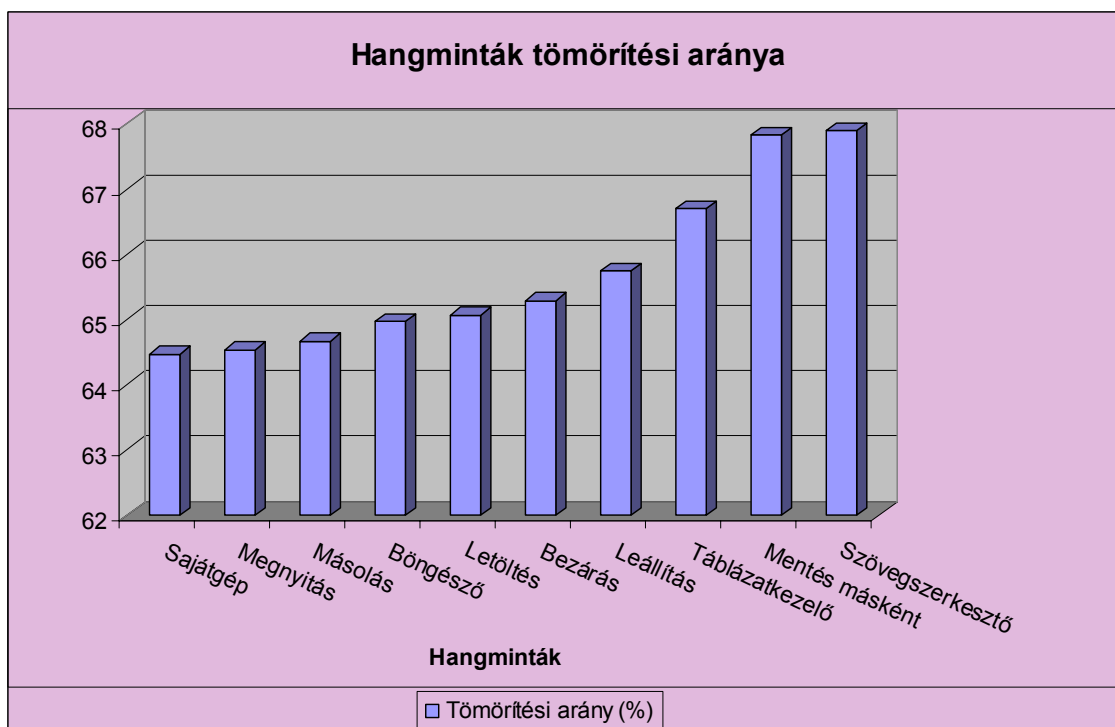
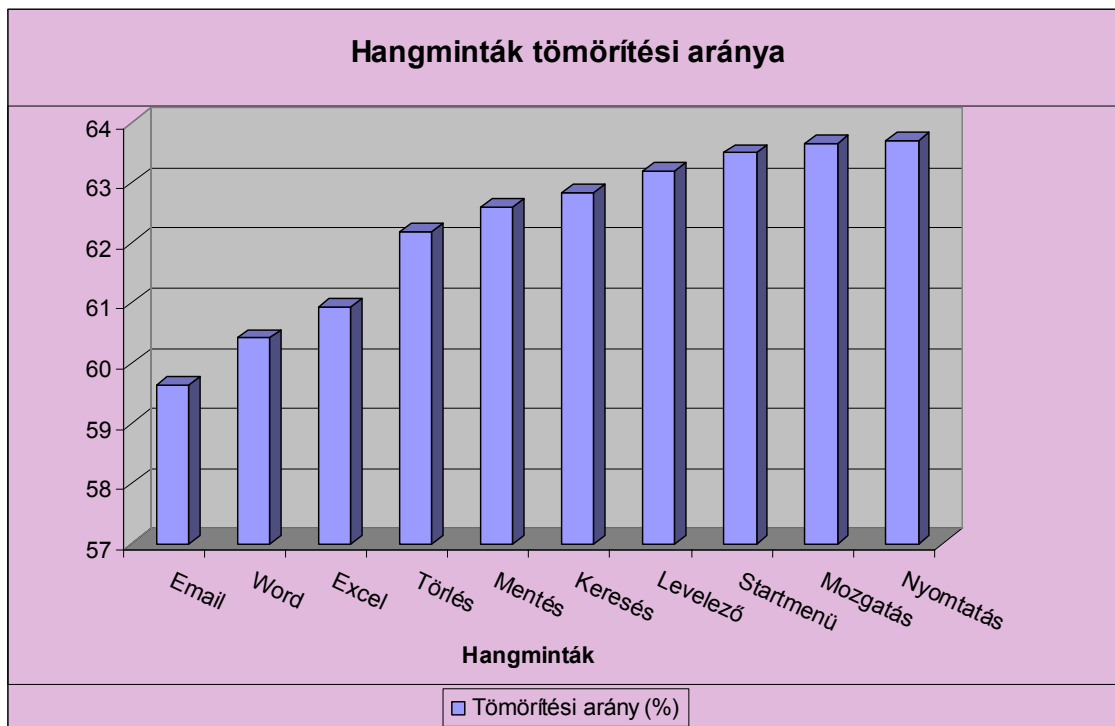
Az eredmények diagramon ábrázolva:

➤ 16 Bit 48 kHz PCM Wave: 5.5 ábra.



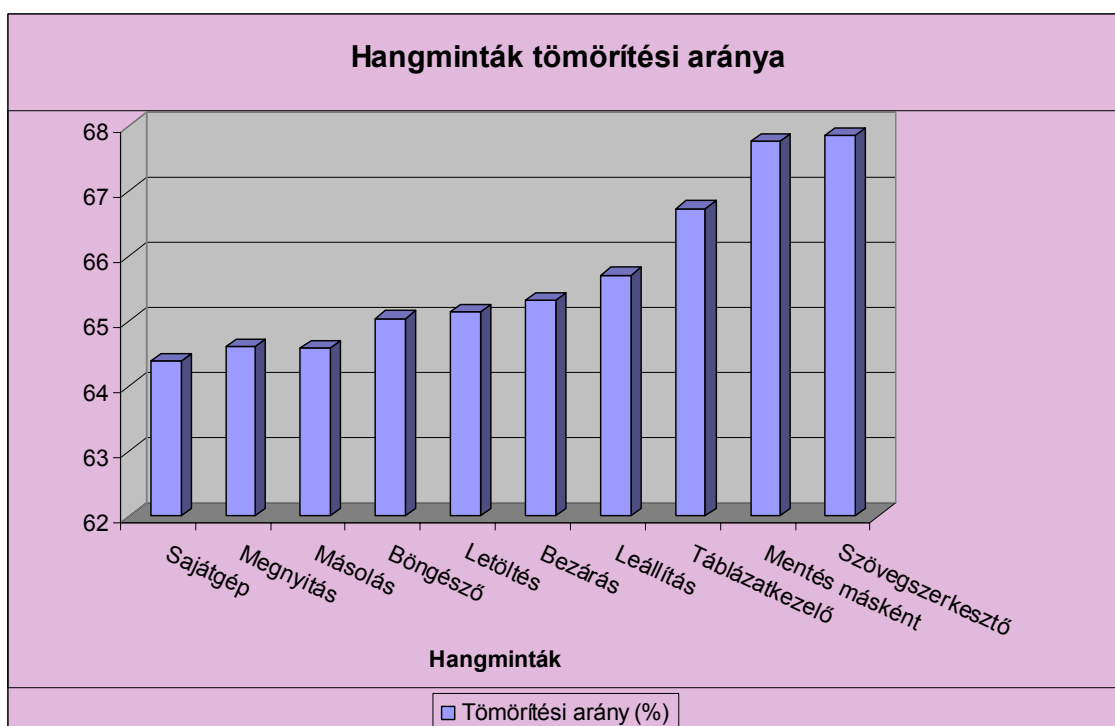
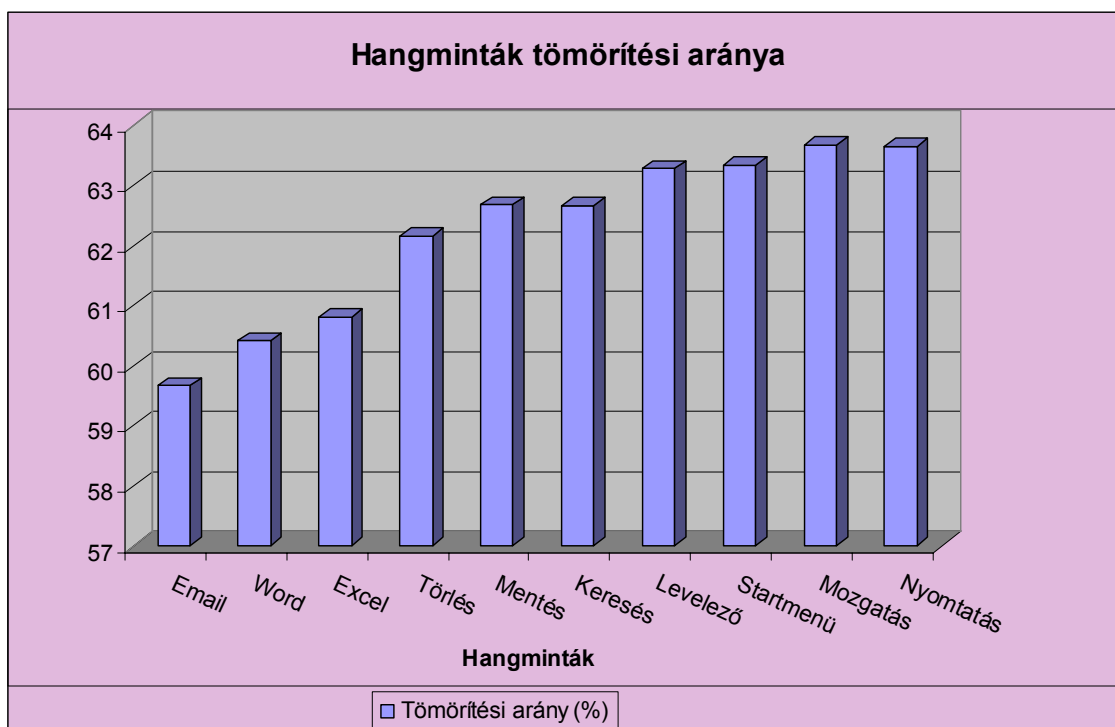
5.5 ábra.

➤ 16 Bit 44.1 kHz PCM Wave: 5.6 ábra.



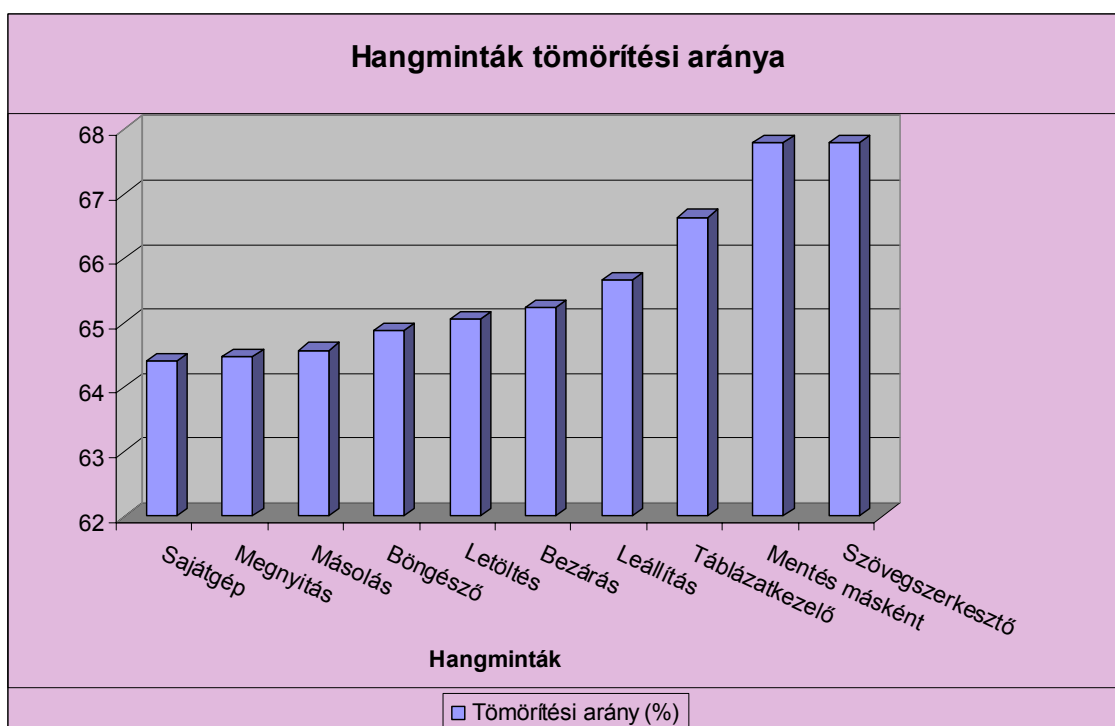
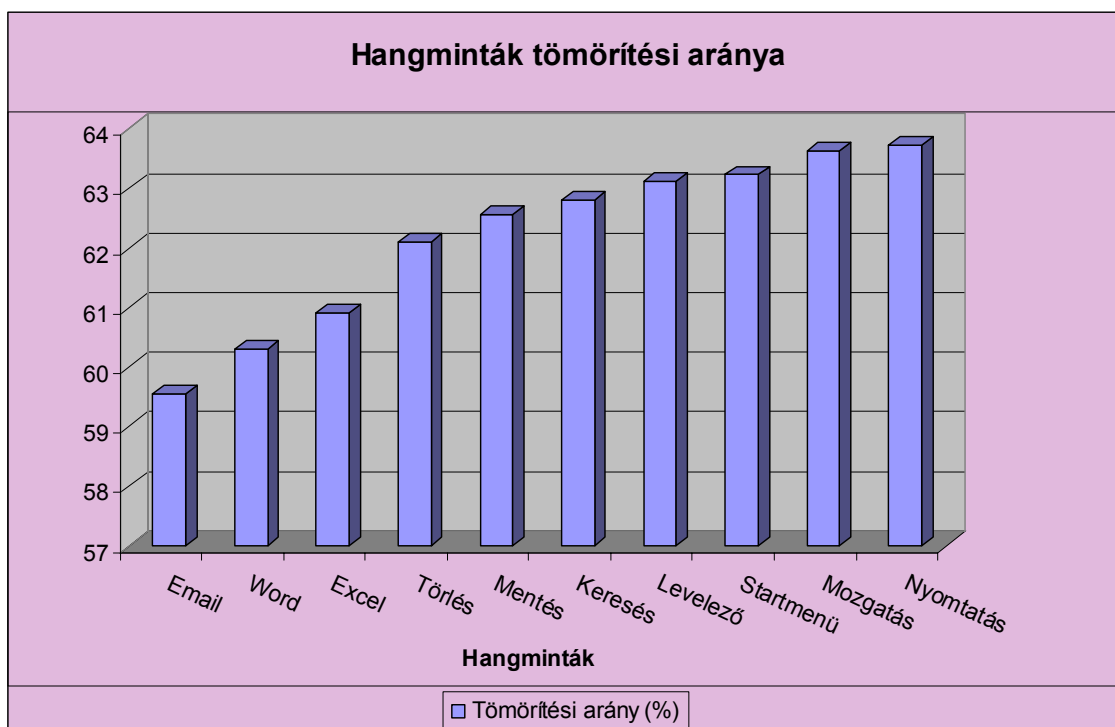
5.6 ábra.

➤ 16 Bit 22.05 kHz PCM Wave: 5.7 ábra.



5.7 ábra.

➤ 8 Bit 44.1 kHz PCM Wave: 5.8 ábra.



5.8 ábra.

Összességében elmondható, hogy a tömörítési arány nem függ az alkalmazott formátumtól. Az arány megközelítőleg lineáris, 59 % és 68 % között változik az eredeti hangfájl hosszának függvényében. Minél nagyobb a hangfájl időbeli hossza, annál nagyobb a tömörítési arány.

A minőségi elemzés alapján elmondható, hogy a 16 Bit 48 kHz-es és a 16 Bit 44.1 kHz-es beszédminta között nincs nagy különbség. A 16 Bit 22.05 kHz már tompábban szól a csökkent frekvenciaátfogás miatt. A 8 Bit 44.1 kHz már nagyon sűg, a 8 Bit felbontású mintáknak már nagy a kvantálási zaja. A mai asztali és mobil számítástechnikai eszközök megnövekedett tárolókapacitása lehetővé teszi, hogy akár 16 Bit 44.1 kHz-es formátumot alkalmazzunk, mivel a tömörített beszédminták mérete 20 KB és 32 KB között változik.

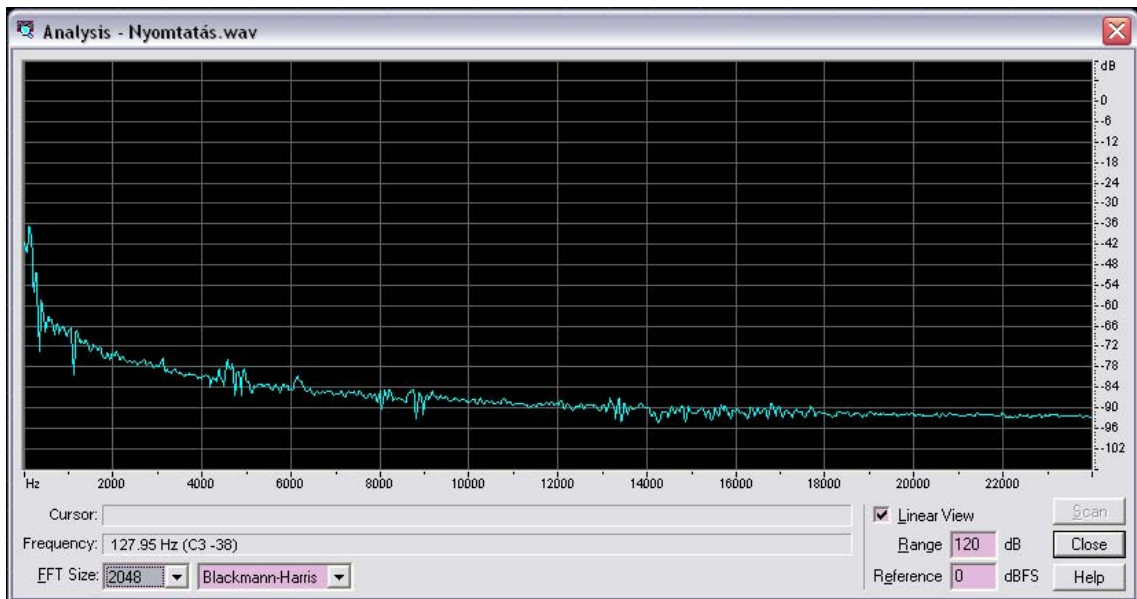
### **5.3 A hangminták spektrális vizsgálata**

A spektrális analízist a Cool Edit Pro program Analyze – Frequency Analysis menüpontjával végeztem. A vizsgálat célja annak kiderítése, hogy miként változik ugyanazon hangfájl spektruma a tömörítés során, a különböző formátumokon.

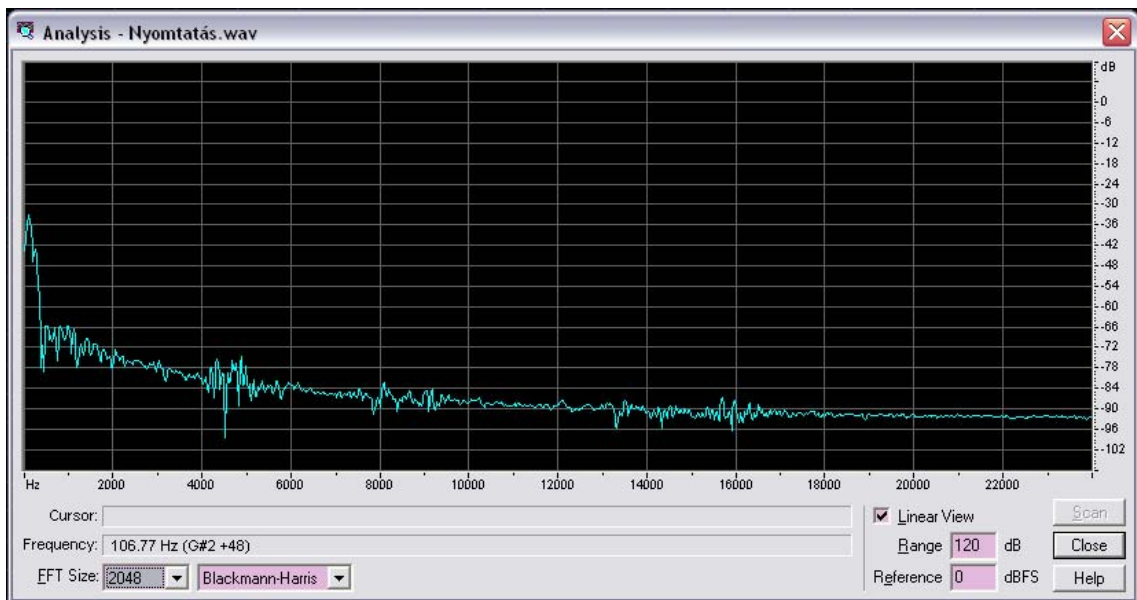
Az analízishez a nyomtatás.wave fájlt választottam, mert spektruma a teljes frekvenciasávban eléggé „sima”, nincsenek benne nagy „tüskék”, így az esetleges kiemeléseket vagy csillapításokat jól nyomon lehet követni a különböző frekvenciákon.

### 5.3.1 A nyomtatás hangfájl spektrumai

➤ 16 Bit 48 kHz PCM Wave: 5.9 és 5.10-es ábra.



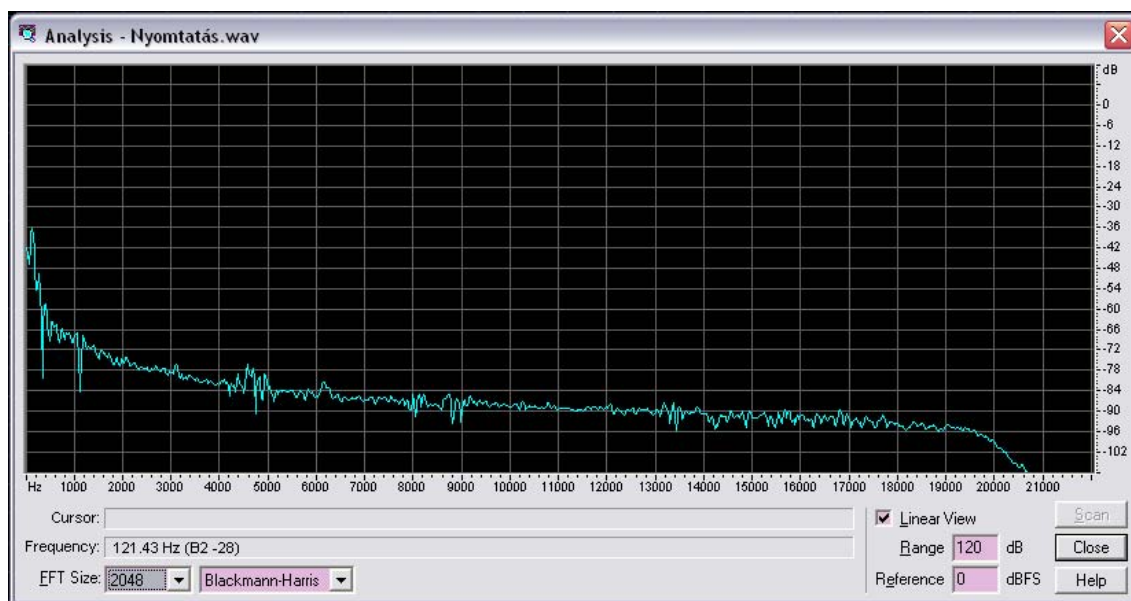
5.9 ábra. Az eredeti fájl spektruma.



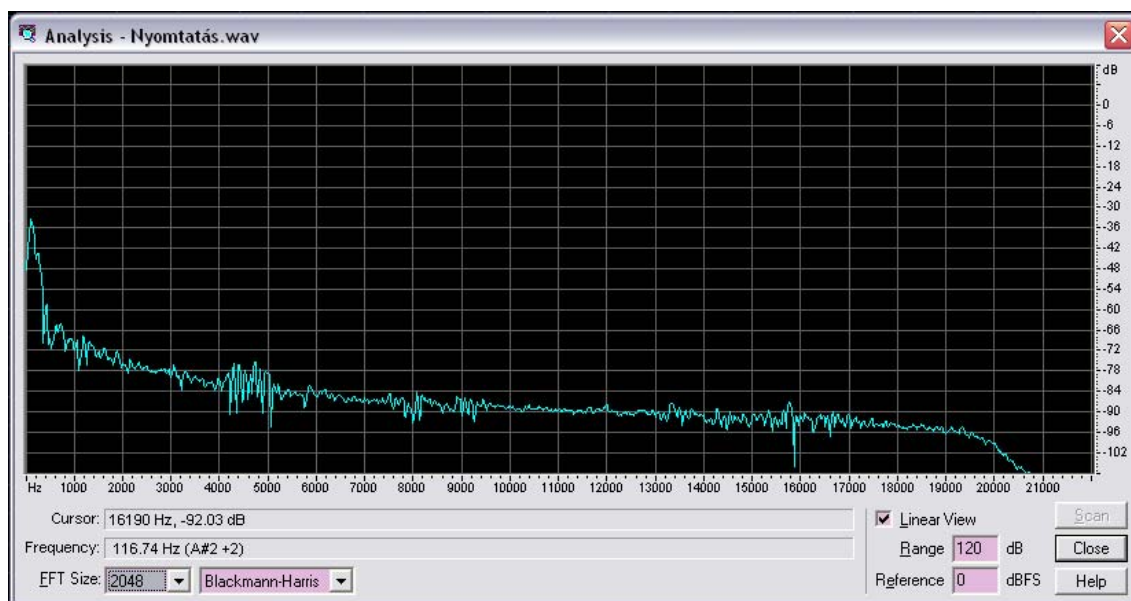
5.10 ábra. A tömörített fájl spektruma.

Itt a tömörített jelnél 4000 és 4500 Hz között, továbbá 16000 Hz-nél figyelhető meg némi torzítás, egyébként nagy eltérés nincs.

➤ 16 Bit 44.1 kHz PCM Wave: 5.11 és 5.12-es ábra.



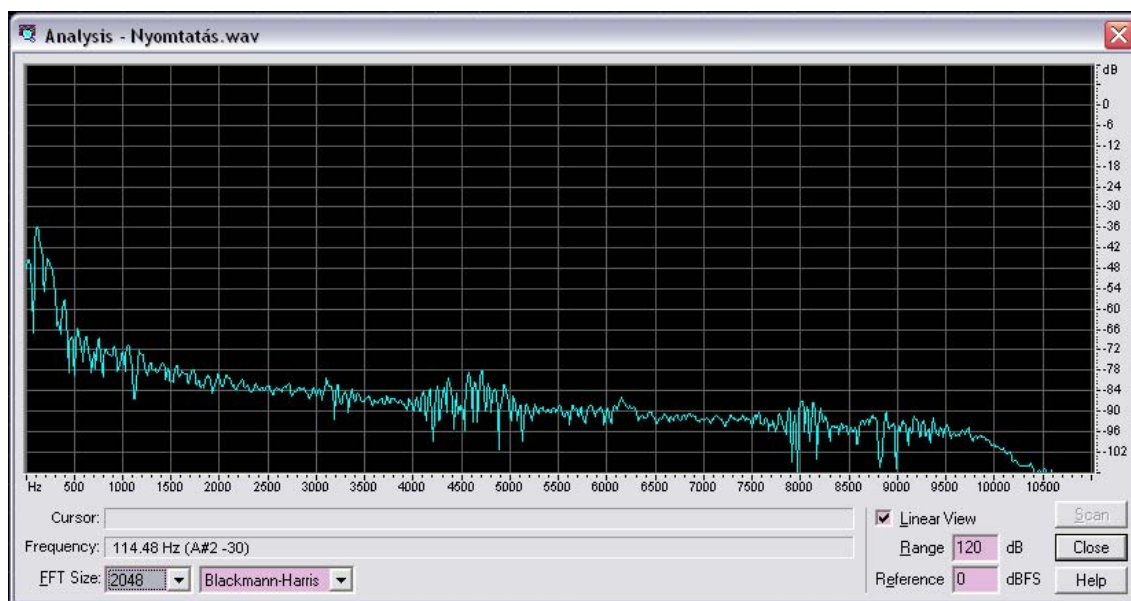
5.11 ábra. Az eredeti fájl spektruma.



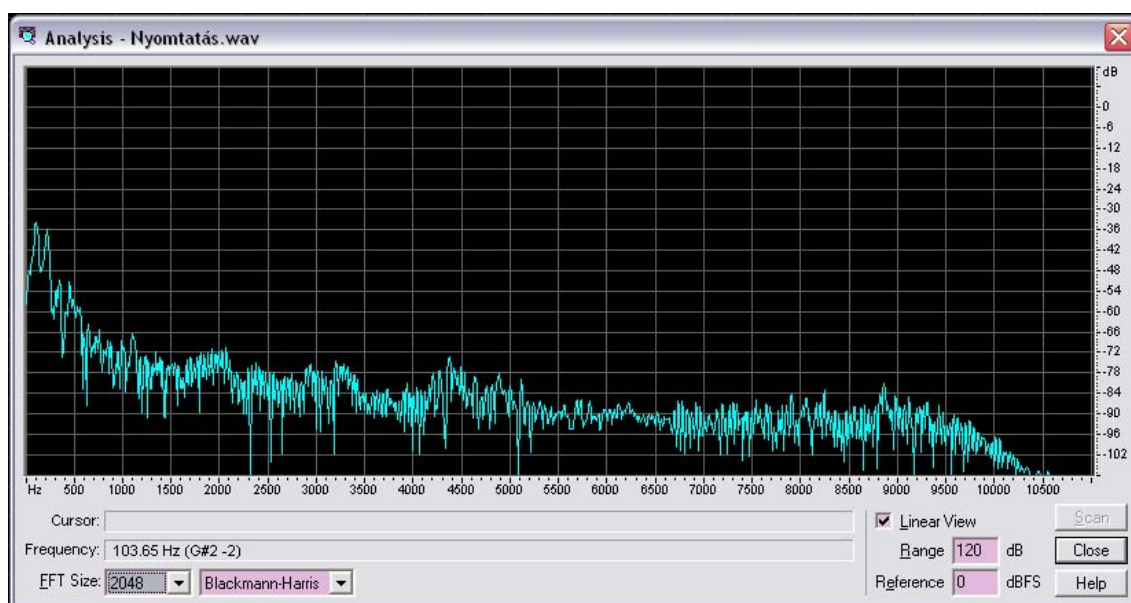
5.12 ábra. A tömörített fájl spektruma.

Itt a tömörített jelnek 4000 és 5000 Hz között, továbbá 16000 Hz-nél „tűskésebb” a spektruma.

➤ 16 Bit 22.05 kHz PCM Wave: 5.13 és 5.14-es ábra.



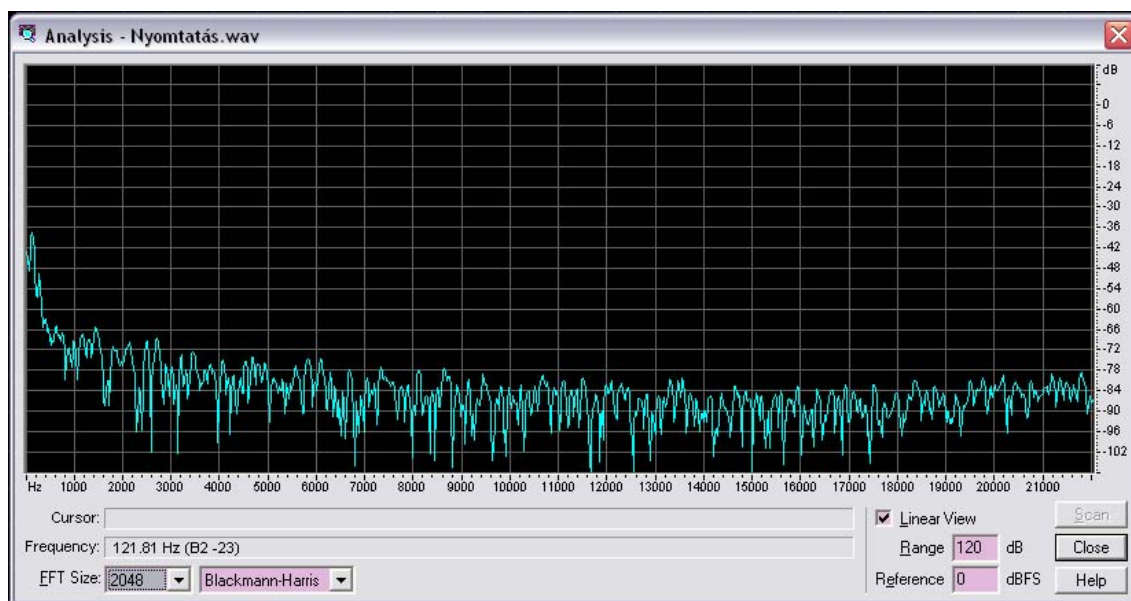
5.13 ábra. Az eredeti fájl spektruma.



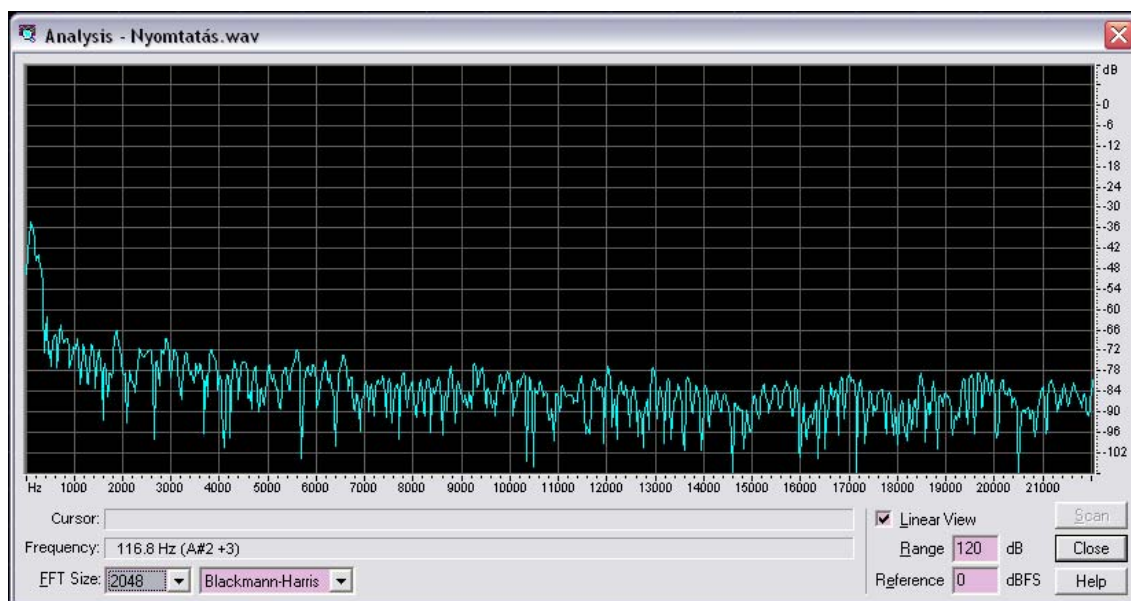
5.14 ábra. A tömörített fájl spektruma.

Itt a tömörített jel spektruma sokkal „szőrösebb”, „tüskéesebb”, tehát ez a formátum nem szerencsés választás.

➤ 8 Bit 44.1 kHz PCM Wave: 5.15 és 5.16-os ábra.



5.15 ábra. Az eredeti fájl spektruma.



5.16 ábra. A tömörített fájl spektruma.

Itt a 8 Bit felbontású minták miatt a kvantálási zaj mindkét jelben nagy. Az eredeti és a tömörített jel spektruma között nincs nagy különbség, mindegyik egyaránt „szőrös”, „tüskés”.

## 6. Magyar nyelven is működő text-to-speech program ismertetése

A fejezetben egy magyar nyelven is működő text-to-speech program kerül ismertetésre, melynek segítségével a szavakat felolvastatom és wave fájlba elmentem, majd a beszédmintákat spearcon-ná alakítom. Továbbá a hangfájlokon elvégzem a tömörítési arányok mérését, a spektrális elemzést és a minőségi vizsgálatot.

### 6.1 SpeakBoard - A beszélő számítógép



A SpeakBoard egy olyan program, aminek segítségével magyar nyelven, természetes hangzású női, vagy férfihangon, fel lehet olvasatni a Word dokumentumokat, e-maileket, az Interneten található weblapokat, pdf dokumentumokat, szöveges állományokat és minden más a képernyőn megjelenő szöveget. A programmal elsősorban az angol és magyar szavak kiejtését lehet gyakorolni.

A program hasznos lehet a tanulásban, a munkában és a szórakozásban egyaránt. Segítheti a tanulást azáltal, hogy a számítógépen tárolt szövegeket, mindig pontosan és jól érthetően olvassa fel, megkönnyítve a tanulást a diszlexiások számára is. A program a felolvasott szövegeket változatos felszínű ablakokban jeleníti meg, ahol pontosan tudja követni, hogy melyik szónál tart a program a szöveg felolvasásában.

Munka közben hasznos segítség lehet a szövegek ellenőrzésére, mivel sok hibát szemmel észre sem veszünk, meghallgatva, pedig rögtön kitűnik. Ugyancsak hasznos funkció, hogy a kimásolt szövegeket mindig felolvassa, amennyiben a SpeakBoard fut, és a vágólap automatikus felolvasása be van kapcsolva, úgy a program azonnal elkezdi felolvasni a kijelölt szöveget. Így

közvetlen visszajelzés van arról, hogy mit fogunk a dokumentumba beszúrni, valamint a billentyűzet segítségével bármikor ellenőrizhető, illetve meghallgatható újból a kimásolt szöveg.

A gyengénlátók használhatják a SpeakBoard programot a képernyőn megjelenő apró feliratok felolvasására, vagy akár hosszabb dokumentumok meghallgatására is, úgy, hogy közben a képernyőn követhetik az elhangzottakat. A felolvasás közben megjelenő szövegek tetszőlegesen nagyok lehetnek, vagy akár teljesen ki is kapcsolhatók.

A SpeakBoard a vágólapon keresztül képes elérni más programok szövegeit. Ha a beállításokban nincs letiltva, akkor a SpeakBoard minden vágólapra került adatot felolvas. Amint új adat kerül a vágólapra, a program elkezd annak felolvasását. Megjegyzendő, hogy a szövegfájl, Word dokumentum, Outlook levél és Internet Explorer weblap szövegét a felsorolt programokban megjelenő SpeakBoard ikonra kattintva is fel lehet felolvastatni.

A SpeakBoard telepítéskor beépül az Internet Explorer-be, ahova egy gombot helyez el, aminek segítségével egyszerűen fel lehet olvastatni az aktuális weblap tartalmát. Ezt egy Java szkript segítségével teszi meg, amit néhány víruskereső az Explorer védelmi réseit kihasználó kódnak vélhet.

A felolvasó program mindig azt olvassa fel, ami le van írva, ezért használható szövegek ellenőrzésére is, hiszen az elütéseket, elírásokat jobban észre lehet venni a szöveg meghallgatásakor, mint többszöri átolvasás után. A program csak a számítógépen lévő szövegeket tudja felolvasni.

A képen lévő szöveget egy karakterfelismerő programmal először át kell alakítani szöveges formátumúra. Az így átalakított szöveget viszont már fel tudja olvasni a program.

A szöveges állományt át lehet átalakítani hangfájlba, vagy pedig a program hangrögzítő funkciójával a beírt szöveget el tudja menteni wav vagy Mp3 formátumban.

## **6.2 Wave fájlok készítése a SpeakBoard programmal**

A program a Media Markt üzleteiben vásárolható meg. Én a demó verziót használtam, melyben a hangrögzítés funkció nem érhető el, így a felvételt egy Creative gyártmányú Sound Blaster Audigy 2 típusú hangkártya „What U Hear” funkciójával készítettem, mely azt jelenti, hogy amit hallunk a hangkártya kimenetén az kerül rögzítésre. A felvételeket 16 Bit 48 kHz PCM Wave formátumban készítettem el.

A SpeakBoard hangja Miklós Demó volt, a lejátszás sebességéül, pedig a -4-es beállítást választottam, mivel az alapbeállítás esetén a lejátszás túl gyors volt, mely az érthetőséget rontotta volna.

## **6.3 A hangminták tömörítési arányának vizsgálata**

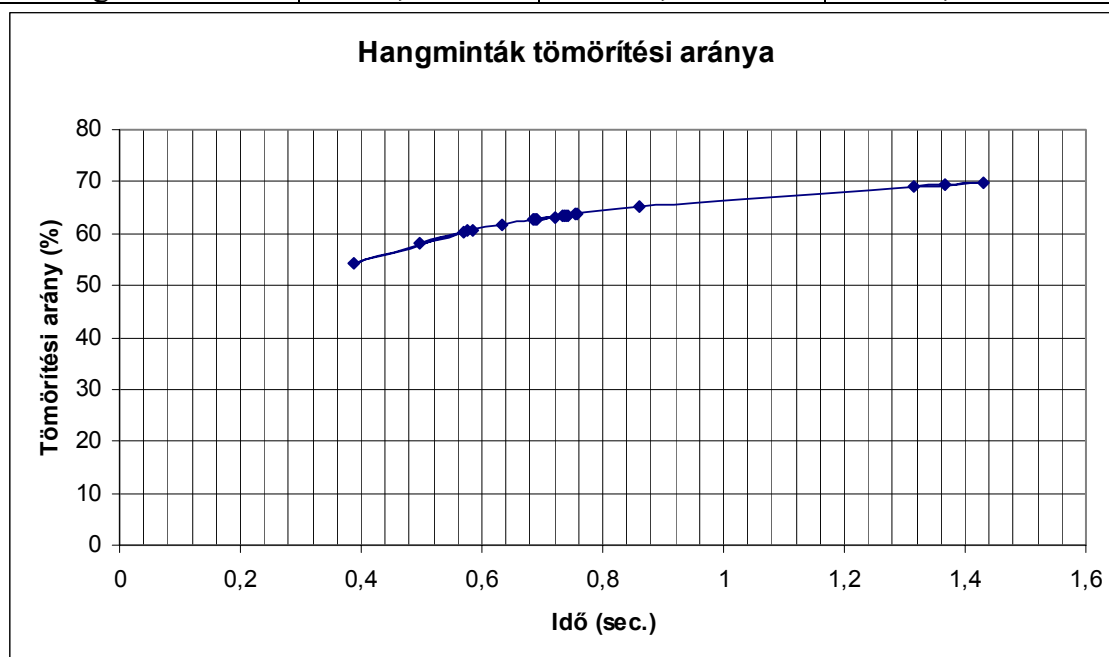
Az eredmények táblázatban és grafikonon megjelenítve:

Az elemzést az 5. fejezetben leírtak szerint végzem, tehát itt most nem részletezem az eljárásokat.

(a beszédminták tömörítési arányát illetően itt is hasonló eredményekre jutottam, mint az 5. fejezetben a magyar nyelvű mikrofonos felvételeknél)

➤ 16 Bit 48 kHz PCM Wave: 6.1 ábra.

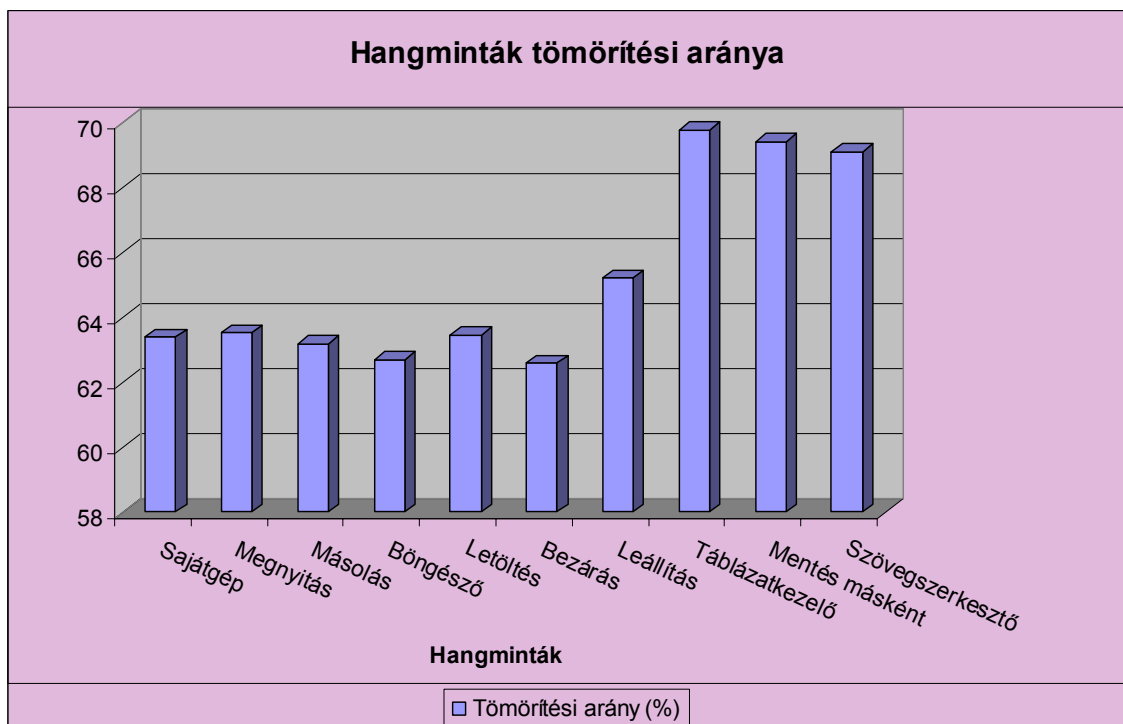
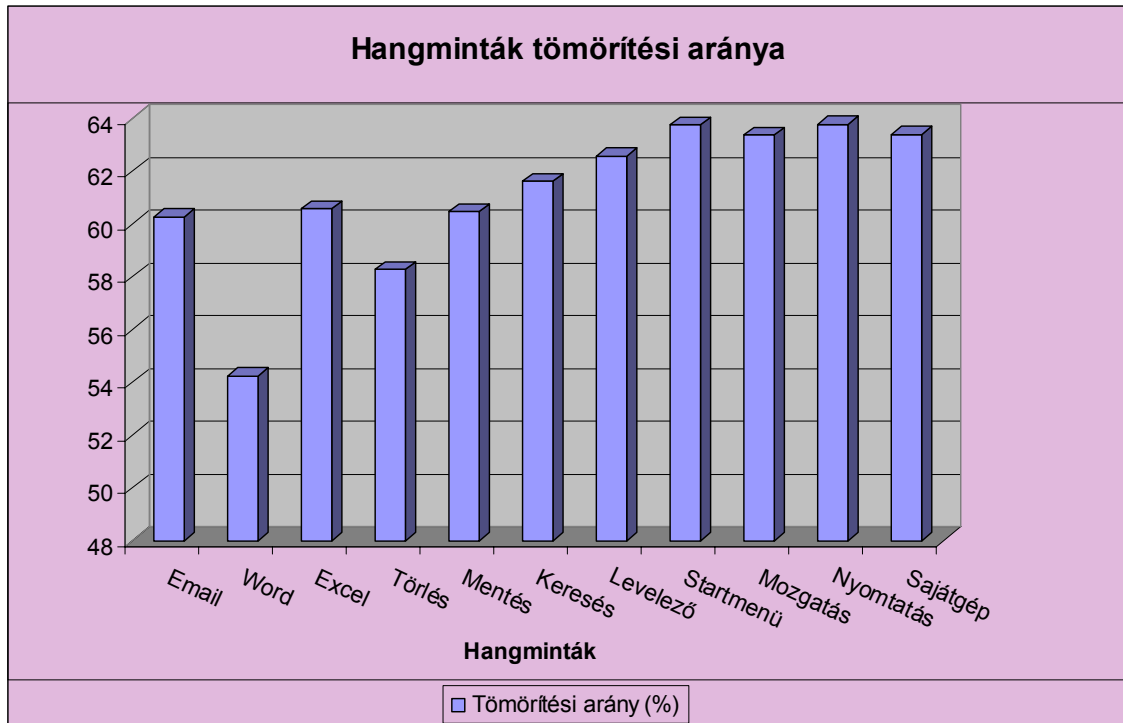
| Beszédminták     | Eredeti (sec.) | Tömörített (sec.) | Tömörítési arány (%) |
|------------------|----------------|-------------------|----------------------|
| Email            | 0,569          | 0,226             | 60,28119508          |
| Word             | 0,389          | 0,178             | 54,24164524          |
| Excel            | 0,586          | 0,231             | 60,58020478          |
| Törlés           | 0,496          | 0,207             | 58,26612903          |
| Mentés           | 0,577          | 0,228             | 60,48526863          |
| Keresés          | 0,633          | 0,243             | 61,61137441          |
| Levelező         | 0,684          | 0,256             | 62,57309942          |
| Startmenü        | 0,759          | 0,275             | 63,76811594          |
| Mozgatás         | 0,732          | 0,268             | 63,38797814          |
| Nyomtatás        | 0,754          | 0,273             | 63,79310345          |
| Sajátgép         | 0,735          | 0,269             | 63,40136054          |
| Megnyitás        | 0,743          | 0,271             | 63,52624495          |
| Másolás          | 0,72           | 0,265             | 63,19444444          |
| Böngésző         | 0,692          | 0,258             | 62,71676301          |
| Letöltés         | 0,739          | 0,27              | 63,46414073          |
| Bezárás          | 0,687          | 0,257             | 62,59097525          |
| Leállítás        | 0,86           | 0,299             | 65,23255814          |
| Táblázatkezelő   | 1,43           | 0,432             | 69,79020979          |
| Mentés másként   | 1,368          | 0,418             | 69,44444444          |
| Szövegszerkesztő | 1,315          | 0,406             | 69,12547529          |



6.1 ábra.

Az eredmények diagramon ábrázolva:

➤ 16 Bit 48 kHz PCM Wave: 6.2 ábra.

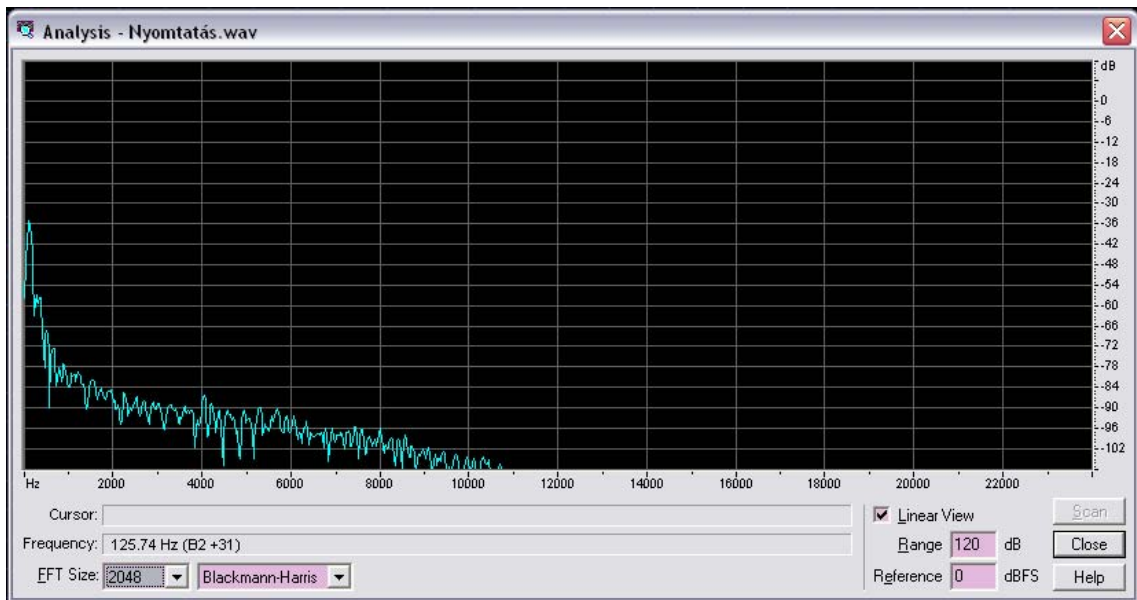


6.2 ábra.

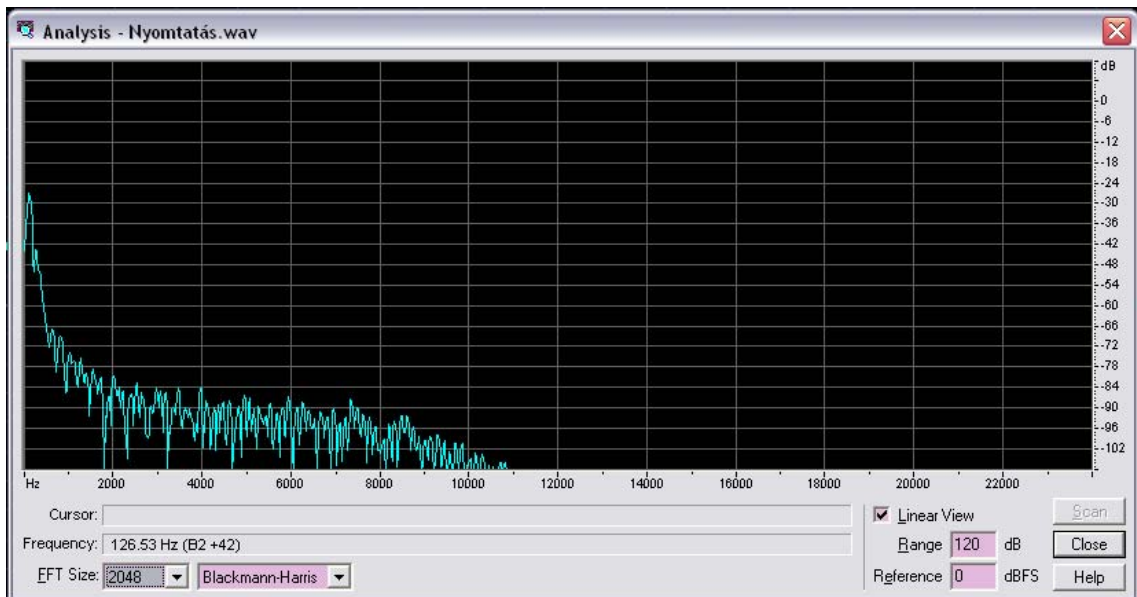
## 6.4 A hangminták spektrális vizsgálata

### 6.4.1 A nyomtatás hangfájl spektrumai

➤ 16 Bit 48 kHz PCM Wave: 6.3 és 6.4-es ábra.



6.3 ábra. Az eredeti fájl spektruma.



6.4 ábra. A tömörített fájl spektruma.

Miklós Demó hangja elég rossz minőségű volt, ez a spektrumon is látszik. A felvételhez elég lett volna a 22.05 kHz-es mintavételi frekvencia is, továbbá mind az eredeti, mind a tömörített jel spektruma nagyon „tüskés” és zajos, valószínűleg a hangminták felbontása sem túl nagy.

## 7. Összefoglalás

A szakdolgozat célja a GUIB (Graphical User Interface for Blind Persons) projekt segítése, ahol a vakok számára olyan rendszert szeretnénk létrehozni, amellyel képesek lesznek használni grafikus felületű asztali és mobil számítástechnikai háttérrel támogatott eszközöket.

A grafikus felületű eszközök menüjében a vakok irányinformációt is tartalmazó hangok alapján tájékozódnak. Ezek a hangok természetesen a menüelemről is hordoznak információt. Ezen információ értelmezésének pontosságát és sebességét a spearcon-ok alkalmazásával növelhetjük.

A feladatom volt, hogy egy MATLAB-ban írt programmal magyar nyelvű spearcon-okat hozzak létre és ezek minőségi vizsgálatát végezzem el. Az elemzés során mértem a hangfájlok tömörítési arányát, továbbá analizáltam a spektrumukat a különböző formátumokban. Az eredmények alapján elmondható, hogy a tömörítési arány nem függ az alkalmazott formátumtól. Az arány megközelítőleg lineáris, 59 % és 68 % között változik az eredeti hangfájl hosszának függvényében. Minél nagyobb a hangfájl időbeli hossza, annál nagyobb a tömörítési arány. A hangfájlok minőségét (frekvenciaátfogás, zaj) viszont az alkalmazott formátum befolyásolja. A vizsgálatok során arra az eredményre jutottam, hogy a hangminták felbontása 16 bitnél ne legyen kisebb, mert ez alatt jelentősen nő a kvantálási zaj, továbbá a mintavételi frekvencia értékénél nem mehetünk 44.1 kHz-alá, mert ennél kisebb mintavételi frekvencia alkalmazása esetén a tömörített jel spektruma sokkal „szőrösebb”, „tüskésebb” lesz.

Tehát a vizsgálatok során arra az eredményre jutottam, hogy a mai asztali és mobil számítástechnikai eszközök megnövekedett tárolókapacitása és számítási teljesítménye lehetővé teszi, hogy akár 16 Bit 44.1 kHz-es formátumot alkalmazzunk, mivel ebben az esetben a tömörített beszédminták mérete 20 KB és 32 KB között változik.

## 8. Irodalomjegyzék

Dr. Wersényi György: Virtuális hangtérszimuláció és a binaurális technológia;

Híradástechnika, Vol. LXII, Nr.2, 2007 február, pp. 25-32

<http://legeza.oszk.hu>

Bruce N. Walker, Amanda Nance, and Jeffrey Lindsay: Spearcons: speech-based earcons improve navigation performance in auditory menus; Proceedings of the 12th International Conference on Auditory Display, London, UK, June 20-23, 2006

<http://sonify.psych.gatech.edu/publications/pdfs/2006ICAD-WalkerNanceLindsay.pdf>

Dianne K. Palladino and Bruce N. Walker: Learning rates for auditory menus enhanced with spearcons versus earcons; Proceedings of the 13th International Conference on Auditory Display, Montréal, Canada, June 26-29, 2007

<http://sonify.psych.gatech.edu/publications/pdfs/2007ICAD-PalladinoWalker.pdf>

Dr. Wersényi György: Bevezetés a stúdiótechnikába; Győr 2005

[www.mathworks.com](http://www.mathworks.com)

[www.softpedia.com/get/Multimedia/Audio/Audio-Editors-Recorders/Cool-Edit-Pro.shtml](http://www.softpedia.com/get/Multimedia/Audio/Audio-Editors-Recorders/Cool-Edit-Pro.shtml)

[www.adobe.com/special/products/audition/syntrillium.html](http://www.adobe.com/special/products/audition/syntrillium.html)

<http://www.speecht.com>

[www.soundblaster.com](http://www.soundblaster.com)