

# Rádióműsorok elemzése a WordNetAffect érzelmi szótár segítségével

Lukács Gergely<sup>1</sup>, Martos Tamás<sup>2</sup>, Jani Máttyás<sup>1</sup>, Takács György<sup>1</sup>

<sup>1</sup> Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar  
1083 Budapest, Práter utca 50/a,

<sup>2</sup> Semmelweis Egyetem, Mentálhigiéné Intézet  
1089 Budapest, Nagyvárad tér 4.

{lukacs, jani.matyas, takacs.gyorgy}@itk.ppke.hu  
martos.tamas@public.semmelweis-univ.hu

**Kivonat:** A hang alapú tartalom személyre szabásához az elmúlt évek technikai fejlődése, elsősorban az okostelefonok és a mobilinternet elterjedése, megteremtette a technikai hátteret. Ennek megfelelően az lejátszási lista készítés (angol: „playlist generation”) fontos kutatási területté lépett elő.

Jelen munka célja a kevert beszéd-zene lejátszási listák készítésének nyelvtchnológiai vizsgálata. Előzetes kutatásunk alapján a hanganyagok szöveges leírátáblól elsősorban a hangulatnak van jelentősége a lejátszási lista készítésnél. A kérdés vizsgálatához rádióadók mintegy 2500 órányi műsorát vizsgáltuk meg. A felvételekben automatikus beszédfelismerővel a WordNetAffect érzelmi szótár szavait ismertük fel, majd az így kapott adatbázist elemeztük. Jellegzetes mintákat találtunk az érzelmi kategóriák együttes előfordulására és az érzelmek időbeli – heti, napi és óránkénti – változására vonatkozóan is.

## 1 Bevezetés

A hang alapú tartalom személyre szabásához az elmúlt évek technikai fejlődése, elsősorban az okostelefonok és a mobilinternet elterjedése, megteremtette a technikai hátteret. A közelmúltban számos személyre szabott zenei streaming szolgáltatás indult el, példaképpen a Spotify, az Apple Music, a Google Play Music vagy a Pandora. Az audió tartalmak személyre szabott kiválasztása és lineáris műsorfolyammá szerkesztése, az ún. lejátszási lista készítés (angol: „playlist generation”) ezzel együtt fontos kutatási területté vált[1]. A lejátszási lista készítés történhet a hanganyagok akusztikai tulajdonságai, a szöveges tartalma (dalszöveg illetve beszédanyag leírata) és a felhasználói interakciók alapján. A szöveges tartalom figyelembevételénél elsősorban a kapcsolódó hangulatnak van jelentősége.

A beszéd-zene lejátszási listák kutatását a megfelelő adatbázisok hiánya nehezíti. Szerkesztett (pozitív és negatív mintákat is tartalmazó) beszéd-zene műsorfolyam készítése óriási költséggel járna, ugyanakkor a professzionálisan szerkesztett rádióadók felvételei jó támpontot adhatnak. Ezeknél nehézséget jelent ugyan, hogy csak pozitív mintákat tartalmaznak, és jelen vannak torzító tényezők is, mint a csatornák tematikussága, vagy az óra által meghatározott műsorstruktúra, pl. a hírek gyakori

ismétlése. Előnyük viszont, hogy nagy mennyiségű adat készíthető automatizáltan, és az adatok elemzése jó kiindulási pontként szolgál további megfontolásokhoz.

Jelen munka keretében professzionálisan szerkesztett rádióadók felvételeiből készítettünk egy adatbázist, részben automatikus beszédfelismerő segítségével, majd az elemeztük. A munka felépítése a következő. A 2. fejezet az irodalomkutatás eredményeit, a 3. fejezet a hangulatfelismerés lehetőségeit mutatja be. A 4. fejezet az adatbázis felépítésével és elemzésével foglalkozik, az. 5. fejezet az adatelemzés eredményeit mutatja be. Az összefoglalás és kitekintés az 6. fejezetben található.

## **2 Kapcsolódó munkák**

Az első kutatások elsősorban a zenei lejátszási listákkal foglalkoztak, és főként akusztikai oldalról közelítették a feladatot. Újabb tanulmányok és vizsgálatok szerint ugyanakkor a nyelvtechnológiának is szerepe van, leginkább a számok közti hangulati harmónia figyelembevételénél [2].

A beszéd-zene lejátszási listák abban különböznek a pusztán zenei lejátszási listáktól, hogy a zeneszámok mellett beszédfelvételeket (pl. interjúk, aktuális hírek, stb.) is tartalmaznak. Ezeknek az automatikus készítésére is mutatkozik igény. Amellett, hogy a felhasználó a saját ízlésének megfelelő zenét hallgatja, össze van kötve a külvilággal és az aktuális eseményekkel[3]. Az erre vonatkozó eddigi kutatások a beszédzene átmenetek akusztikai tulajdonságait [4], a szöveges tartalom szerepét [5], valamint a tartalomfüggetlen, kizárólag a felhasználói visszajelzéseken alapuló ún. felhasználói szűrés (collaborative filtering) lehetőségeit vizsgálták [6].

Közösségi média hangulati elemzésével több kutatás is foglalkozik, példaképpen a Twitter elemzését írja le [7].

## **3 Hangulatfelismerés, a WordNetAffect hangulati szótár**

Természetes nyelvi szövegek hangulatának felismerésére, mérésére több szótár, eszköz illetve módszer is elérhető, példaképpen a WordNetAffect (WNA)[8], a Linguistic Inquiry and Word Count (LIWC)[9], az Affective Norms for English Words (ANEW)[10] és a SentiWordNet [11].

Jelen munkához egy, a leggyakrabban használt kétértékűnél (pozitív/negatív) részletesebb kategóriarendszerre volt szükségünk. Választásunk ezért a WordNetAffect érzelmi szótárra esett, mely az Ekman-féle kategóriarendszert [12] használja, hat alapérzelmet különböztet meg. Az érzelmek felismerése, mérése az WNA adott érzelmi kategóriájába tartozó szavak számának meghatározásával történik, további súlyozás nélkül. Vizsgálatunk során így az Ekman-féle kategóriákat különböztetjük meg, ezek az öröm (Ö), a meglepettség (M), a félelem (F), a düh (D), a szomorúság (Sz) és az undor (U).

## 4 Adatbázis

### 4.1 Eszközök, elkészítés

**Eszközök, az elkészítés folyamata.** Online elérhető rádióadókról készítettünk felvételeket, a hanganyagokat órás blokkokban rögzítettük. A felvételeken beszéd-zene szétválasztót és szünet detektort futtattunk. A beszéd szakaszokat beszédfelismerő segítségével leiratoztuk, valamint kulcsszókereső algoritmusok segítségével a WordNetAffect érzelmi szótár szavait kerestük.

*Beszéd-zene szétválasztó.* Beszéd-zene szétválasztóként a Xiph.Org<sup>1</sup> által készített Opus hangkodek egyszerű neuronhálával és ennek a kimenetét simító két állapotú Markov-moddellel készített beszédfelismerő algoritmusát vettük alapul és egészítettük ki további utófeldolgozással. A kiegészített algoritmus a beszéd és zeneszakaszok mellett a kevert szakaszokat is felismeri, melyekben lehalkított zene és beszéd egyszerre hallhatóak.

Az automatikus szegmentálás minőségének ellenőrzéséhez csatornánként 5-5 óra felvételen kézzel kijavítottuk a hibákat. Ezután kiszámoltuk a Cohen féle kappát és alapértelmezett módon majd súlyozva is. Az automatikus felismerő három osztályt tud felismerni (beszéd, zene, beszéd-zene), az annotátorunk viszont időnként jelölte az egyéb (pl. taps, motorzaj) kategóriát is. A súlyozott kappa számításánál a mindkettő-beszéd és a mindkettő-zene (illetve ezek fordítottjai) fele olyan súllyal szerepeltek, mint a többi tévesztés. Az 1. táblázat foglalja össze a kappa értékeket. Általánosságban 0.7-nél magasabb értékeket jónak szokták tartani. Egyedül az R4-es csatorna kappája alacsonyabb ennél. Érdekes jelenség, hogy az R3 komolyzenei adón kimagaslóan jól egyezik az automata a kézi annotálással. Ez valószínűleg részben annak köszönhető volt, hogy ritkábban szólt egyszerre beszéd és zene is, nem nagyon volt "átúsztatás", tisztábbak voltak a váltások.

1. Táblázat: A beszéd-zene felismerés minősége

Rádióadó	Cohen's kappa	Súlyozott Kappa
R1	0,713	0,814
R2	0,730	0,822
R3	0,930	0,936
R4	0,601	0,601

*Beszédfelismerő.* Az automatikus leiratot és a kulcsszókeresést a Kaldi [13] nyílt forráskódú beszédfelismerő rendszer segítségével végeztük. Az alapértelmezetten beépített TEDLIUM adatbázisra építő receptet használtuk a leiratokhoz. Csak a legutolsó GMM-HMM modellt használtuk, mély-neuron hálókat nem alkalmaztunk. Ezt leszámítva a tanítást a recept változtatása nélkül végeztük. Kulcsszókereséshez a Kaldiban található, hipotézis-gráf indexelésén alapuló kulcsszókeresési eljárást használtuk. A kulcsszavaknak a WordNetAffect érzelmi szótár szavait választottuk, célunk

<sup>1</sup> <http://www.xiph.org>

az érzelmi kategóriák előfordulásának vizsgálata volt. Az automatikus leiratot csak a szószámláláshoz használtuk. A beszédfelismerés hibájára közvetlen mérésünk nincs. A beszédfelismerő korábbi változatánál, más adathalmazon a szófelismerési hiba (WER) 47,2% volt [5], ugyanakkora hibák kiegyenlítették egymást és nem befolyásolták számottevően az érzelemfelismerés pontosságát.

## 4.2 Adatok leírása

Az adatbázis a BBC rádióadó négy csatornájának adásait tartalmazza. A Radio 1 és a Radio 2 könnyűzenei adók, az előbbi fiatalok számára, az utóbbi felnőtteknek. A Radio 3 komolyzenei adó, a Radio 4 pedig irodalmi és aktuális eseményekkel kapcsolatos műsorokat ad.

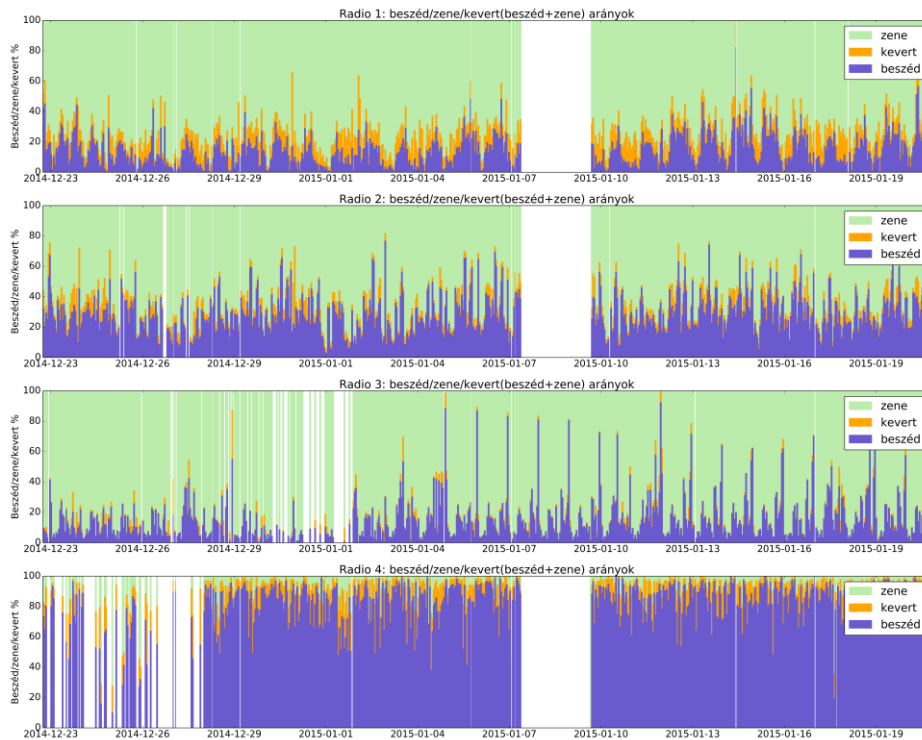
A felvételek 2014. december 22. és 2015. január 20. között készültek, összesen 2 444 óra terjedelemben. Technikai okok miatt a felvételek nem fedik le a teljes időtartományt, szünetek előfordulnak. A beszéd, kevert beszéd-zene és beszéd arányát az egyes adóknál a 2. táblázat mutatja. A hangadatok áttekintése az 1. ábrán látható.

2. Táblázat: Zene, beszéd/zene és beszéd arány a vizsgált rádióadóknál

	Zene	Kevert beszéd-zene	Beszéd
R1	75%	10%	15%
R2	65%	6%	29%
R3	83%	2%	16
R4	7%	9%	83%

A teljes adatbázis 10 572 398 szót tartalmaz. A WNA szavak száma 233 760, arányuk így 2.2%. A WNA érzelmi kategóriái között a WNA szavak megoszlása a következő: öröm: 45%, düh: 14%, félelem: 12%, szomorúság: 15%, meglepetés: 12%, undor: 3%. Az örömhöz tartozó, kiemelkedően magas érték egybecsenghet [14] megfigyelésével. A többi érzelmehez tartozó érték hasonló egymáshoz, kivéve az undorra vonatkozó alacsony számot.

Az adatbázis a jelen munkában bemutatott általános vizsgálatokhoz elegendő nagy. Speciális kérdésfelvetések esetén (pl. undor vizsgálata hétféle éjszaka) ugyanakkor a jelen adatok alapján végzett megfigyelések bizonytalansága magas.



1. ábra Rádiófelvételek áttekintése, beszéd-zene arány az idő függvényében.

## 5 Vizsgálatok

### 5.1 Kiértékelési mérték

A vizsgálatoknál a következő mértéket vettük figyelembe. Kiindulási pontként a WNA szavak és az összes szó aránya szolgált, példaképpen egy adott vizsgálatnál (adónál, időtartományban, stb.) a örömhöz tartozó WNA szavak aránya az összes elhangzott szó 1.1%-a.

Mivel az egyes érzelmi kategóriákhoz tartozó szavak előfordulási gyakorisága nagyban különbözik, ezért célszerű volt a WNA szavak arányának változását nem abszolút értékben mérni, hanem az adott esetre vonatkozó átlagos arányhoz viszonyítva. A WNA szavak (átlagos arányához mért) relatív arányán a trendek világosabban értékelhetők, és az egyes érzelmi kategóriákra vonatkozóan is összehasonlíthatóak.

## 5.2 Érzelmi kategóriák aggregálása

Az adatokon ellenőriztük, hogy az érzelmek együttes előfordulása mutat-e valamilyen mintázatot. Konkrétabban arra voltunk kíváncsiak, hogy az azonos, illetve hasonló érzelmi jelentésű szavak nagyobb valószínűséggel fordulnak-e elő egymáshoz közeli időpontban. Ha igen, akkor ez két előzetes feltevést igazol. Egyrészt azt jelzi, hogy az egyes érzelmi jelentésű szavak nem véletlenszerűen fordulnak elő egymáshoz közel, hanem az adott szöveg adott szöveggörnyezetére általánosabban jellemző érzelmi-hangulati állapotot tükröznek önmagukban is és az együttes előfordulásukból sejthetően összességében is. A második igazolt feltevés az elsőből következik: amennyiben az egyes szavak érzelmi jelentése egy általánosabb érzelmi mintázatot jelez, akkor az elkülönült előfordulások összegzése is érvényes eljárás arra, hogy nagyobb, egybefüggő hangzó szöveg érzelmi jellegének leírására használjuk az így számított aggregált értékeket. (Az együttesen összegzendő szakaszok kijelölése természetesen szintén fontos gyakorlati kérdés, de ezt a problémát jelen érvelés most nem érinti.)

A következő ellenőrzést végeztük el: az adatbázisban megkerestük azokat a kódolt érzelmi szavakat, melyek elhangzása között legfeljebb 3 mp volt a különbség – azaz feltételezhető volt, hogy amennyiben létezik a szövegnek valamilyen tartósabb hangulati állapota, akkor a két érzelmi jelentésű szó körülbelül azonos érzelmi-hangulati állapotú beszélőtől származik.

A kódolt anyagban összesen 91255 ilyen szó pár volt. Az együttes előfordulások mintázatának feltárására korrespondencia-elemzést (correspondence analysis, CA) végeztünk az SPSS programcsomag felhasználásával. A CA kategóriális adatok együttes előfordulásainak mintázatát jeleníti meg egy többdimenziós térben. Az egyes lehetséges párok együttes előfordulásainak valószínűségét a párok tagjainak egymástól való távolsága reprezentálja. (Azaz a közeli párok nagy valószínűséggel fordulhatnak elő 3 mp-en belül, míg a távoli párok együttes előfordulásának valószínűsége csekély.)

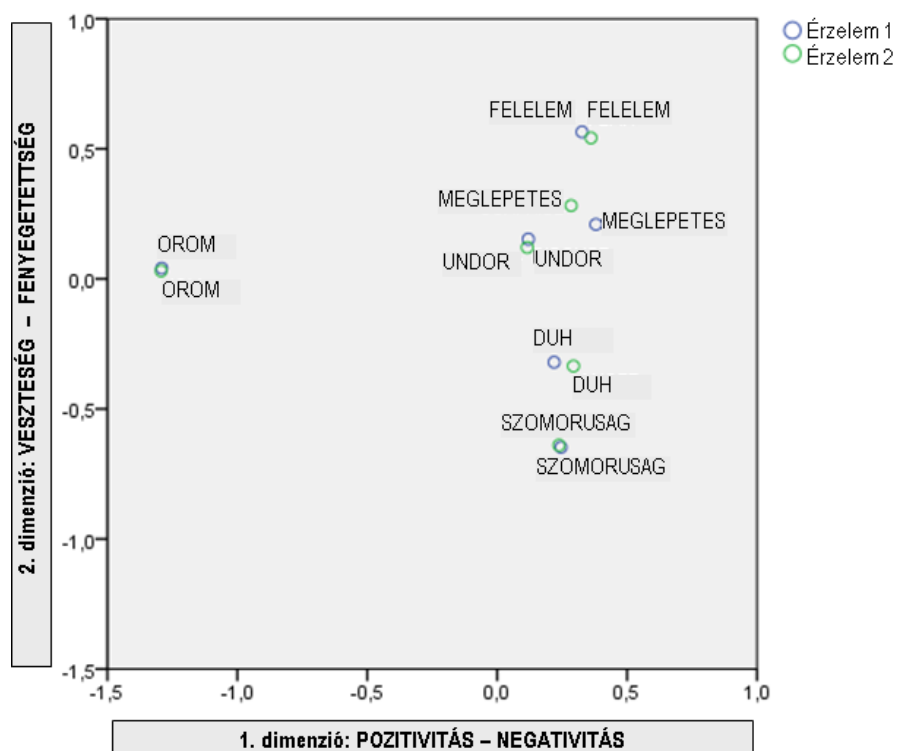
**Eredmények:** A CA alapján két dimenzió írja le az adatok varianciájának összesen 92,4 %-át. Ebből az 1. dimenzió magyarázott 77,1 %-ot, míg a második 15,3 %-ot. Az egyes érzelmi kifejezés-párok együttes előfordulásának valószínűségét a 2. ábrán mutatjuk be. Az ábra alapján két fontos megállapítást tehetünk.

1. A legnagyobb valószínűséggel ugyanazon érzelmi kifejezések jelennek meg egymástól legfeljebb 3 mp távolságban. (Azaz az időben első és a második előfordulást reprezentáló kétféle színű pont gyakorlatilag átfedésben van az ugyanazon érzelmeket reprezentáló szavak esetében.) Ez az eredmény arra utal, hogy a kódolt szövegek – viszonylag rövid időhatárokon belül – valóban egységes hangulati-érzelmi állapotokkal jellemezhetők. Másfelől pedig jelzi a kódolási rendszer érvényességét is.

2. A két dimenziót is megkísérelhetjük értelmezni. Az ábrán vízszintes dimenzió egyértelműen a pozitív (öröm) és a negatív érzelmeket jelölő szavak elkülönülését jelzi. A másik (az ábrán függőleges) dimenzió értelmezése kevésbé egyértelmű, de alább megkíséreljük. Itt a pozitív oldalon természetesen nem volt változatosság. A negatív oldalon az egyik póluson a félelem, a másikon a szomorúság jelzi az egymástól elkülönülő érzelmeket. A félelemhez közelebb helyezkedik el a meglepetés és az undor, míg a szomorúsághoz a düh. Ez a dimenzió értelmezésünk szerint annak az érzelemfelosztásnak felel meg, ami a negatív érzelmeket aszerint különbözteti meg, hogy milyen tapasztalatra adott választ jeleznek. A félelem egy fenyegetésre, averzív

inger közeledésére adott reakció, míg a szomorúság valamilyen pozitív élmény elvesztésére, távolodására adott válasz. A köztes érzelmek is besorolhatók ebbe a modellbe, hiszen a meglepetés és az undor inkább valamilyen közelítő averzív ingerhez kapcsolódik, míg a düh a szomorúsághoz hasonlóan a frusztrációra, azaz veszteségre adott reakció. A második tengelyt ezért veszteség vs. fenyegetés tengelynek neveztük el.

Összességében mindkét szintű mintázat (azonos érzelmek közelsége és a kétdimenziós érzelmi tér értelmezhetősége) arra utal, hogy a kódolási rendszer érvényes adatokat szolgáltat és az egyes érzelmek összesített gyakorisága is érvényes eljárás – beleértve akár a pozitív és negatív érzések elkülönült aggregálását is.



**2. ábra** A 3 mp-en belül együttesen előforduló érzelmi jelentésű szavak korrespondancia-elemzésének eredménye (91255 szópár alapján). Az egyes dimenziók elnevezése a szerzők javaslata.

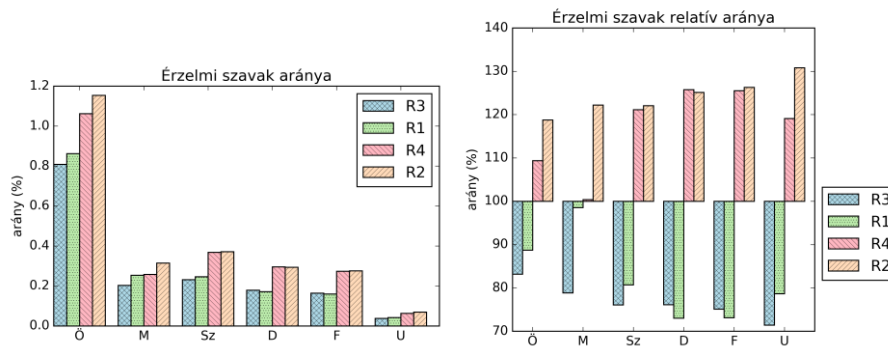
### 5.3 Érzelmi kategóriák időben aggregált aránya

Elsőként időben aggregálva vizsgáltuk a hangulati szavak előfordulását az egyes rádióadóknál, két mértéket használva: (1) a hangulati szavak aránya az összes elhangzott

szóhoz viszonyítva, (2) és a hangulati szavak relatív aránya a négy csatorna átlagértékéhez viszonyítva (3. ábra).

Az R2 és az R4 esetében jelentősen gyakoribbak a hangulati szavak. Egy lehetséges magyarázat, hogy ezek több, mélyebb tartalmat akarnak közvetíteni, amire nagyobb beszédarányuk is utalhat. (A nagyobb beszéd arány, vagy akár a beszédtempó csatornánkénti különbözősége közvetlenül nem befolyásolja a vizsgált értékeket, hiszen a hangulati szavak és az elhangzott szavak arányát, illetve ennek az arálynak a változását vizsgáljuk.)

Az R2 és R4 örömmre vonatkozó jellemzői ugyanakkor kevésbé kiugróak, mint a szomorúság, düh, félelem és undor esetében. Egy lehetséges magyarázat, hogy az örömet viszonylag könnyebb kommunikálni, és a több tartalmat szolgáltató adók (az R2 és az R4) az amúgy inkább háttérbe szorított negatív érzelmeknek is nagyobb teret ad. A meglepetésre vonatkozóan az R1 és az R4 esetén nem áll a fenti minta, az okok felderítéséhez további vizsgálatokra van szükség.

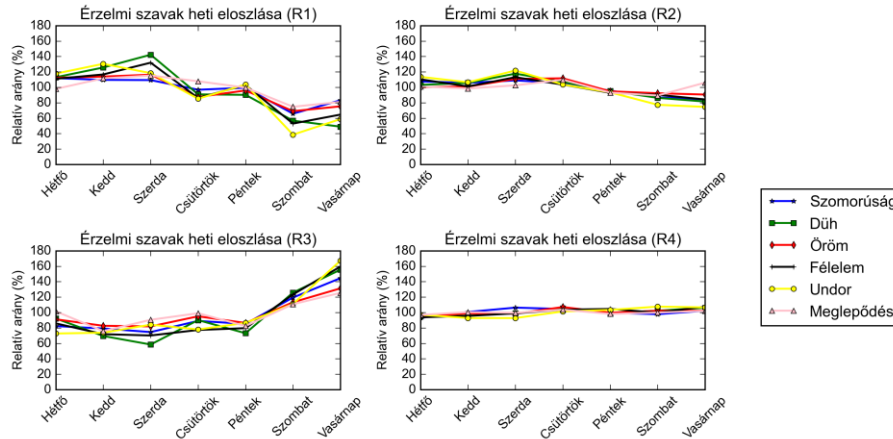


3. ábra WNA érzelmi kategóriák aránya és relatív aránya a vizsgált rádiófelvételekben.

### 5.4 Időbeli minták

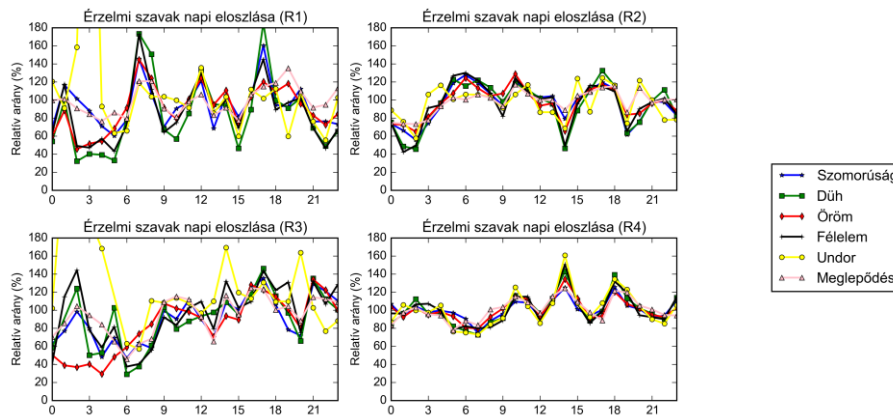
**Heti minta.** A WNA szavak arányának heti ritmusú változását a 4. ábra mutatja. Az érzelmi szavak az érzelmi kategóriától függetlenül együtt változnak. Az R1 esetén hétköznap magasabb, hét végén alacsonyabb a WNA szavak aránya. Az R3 esetében ellenkező minta figyelhető meg, a hétvégéken jelentősen magasabbak az értékek. Az R2 esetén jóval mérsékeltőbb a heti ritmusú változás, ennek iránya pedig az R1-hez hasonló: hét közben magasabb értékek. Az R4 esetében alig észrevehető a heti minta, tendenciáját tekintve az R3-hoz hasonlóan hétvégén emelkednek, bár csak kis mértékben az értékek.





4. ábra WNA érzelmi kategóriák heti ritmusa adónként.

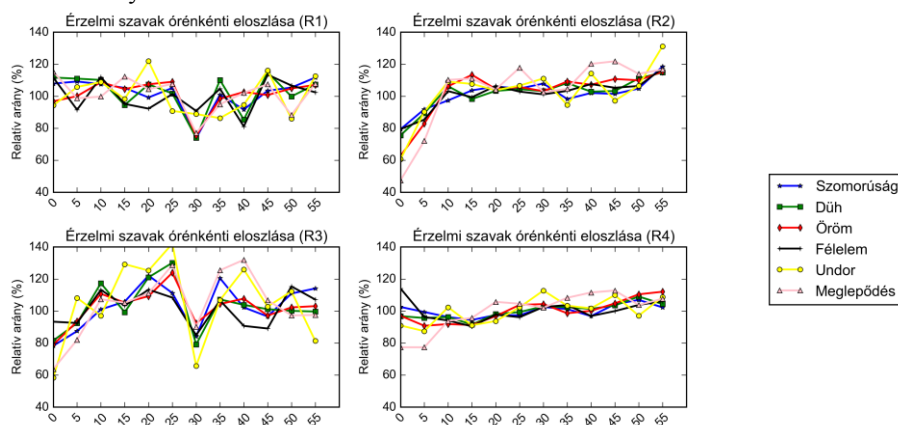
**Napi minta.** A napi ritmusú változást az 5. ábra mutatja. Az érzelmek itt is nagyjából együtt mozognak. Általában jellemző a hajnali nyugodtabb és a nappali magasabb szint (az R1 és az R3 esetén az undorra vonatkozó hajnali kiugró értékek nagyon kis számú mintából történtek kiszámolásra). Az R1 adóra jellemzőek a reggeli és az esti kiugró értékek. Az R2 esetében kisebbek a változások, de a délutáni és éjszakai alacsonyabb értékek itt is megfigyelhetők. Az R3 esetében egy délutánra, estére emelkedő tendencia látható. Ez esetleg összefügghet a heti ritmusban megfigyelt, hétfővégre emelkedő értékekkel is. Az R4 napi ritmusán csak kisebb változások figyelhetők meg, a délutáni csúcsok ennek ellenére felismerhetők.



5. ábra WNA érzelmi kategóriák napi ritmusa adónként.

**Órai minta.** Célszerűnek tűnik az óránkénti mintát is megvizsgálni, hiszen a rádióadók műsorszerkesztése ehhez igazodik. Az órák első perceiben, amikor általában a hírek hallhatóak, a WNA szavak aránya alacsony. Ugyanakkor ez még szembetűnőbb az öröm és a meglepetés esetében, ezekhez képest a félelem, düh, szomorúság, undor aránya általában viszonylag magasabb. Elsősorban a nagyobb beszédarányal rendel-

kező R2 és R4 adókra jellemző, hogy az órák folyamán enyhén növekszik a WNA szavak aránya.



6. ábra WNA érzelmi kategóriák óránkénti ritmusa adónként.

## 6 Összegzés, kitekintés

Érdekes megfigyelés, hogy a hat vizsgált érzelem általában együtt mozgott, együtt mutattak fel alacsony vagy magas értékeket. Ez alapján a műsorok egyik legalapvetőbb (rejtett, implicit) tulajdonsága az lehet, hogy mennyire „telítettek” érzelmmel. Az (implicit) érzelmek milyensége, aránya csak második lépésben árnyalhatja az érzelmi összhatást.

Az időbeli változást tekintve jellemző heti, napi és óránként visszatérő mintákat találtunk. Az adók között ugyanakkor jelentős különbségek is vannak. Több szempont szerint hasonlóképpen viselkedik az R1 könnyűzenei és az R3 komolyzenei adó egyfelől, és a magasabb beszédarányal rendelkező R2 és R4 másfelől. Más szempont alapján pedig a két könnyűzenei adó, a fiatalokat célzó R1 és az idősebbeknek szánt R2 mutat hasonlóságot.

A Twitter adatok vizsgálata a WNA segítségével [7] alapján jellegzetes napi mintázatot mutatott, reggel a pozitív, este a negatív érzelmek mutattak viszonylag magasabb értéket. A rádióadások napi mintája ezt részben követi. A rádióadások ugyanakkor rendelkeznek egy jellegzetes óránként visszatérő mintával is, a műsorszerkesztés sajátosságai miatt. A Twitter adatainak vizsgálatakor a téli-nyári adatok összevetése is megtörtént [7], erre a rádióadások elemzéséhez nem áll rendelkezésre adatunk. A rádióadásokban ugyanakkor egy jellegzetes heti minta is felismerhető, a Twitter adatainak vizsgálata ilyen szempontból érdekes összehasonlításokra adhatna lehetőséget.

Érdekes lenne tovább validálni emberi észleléssel végzett kísérletben, hogy a szavak gyakoriságával kimutatott hangulati jellemzők mennyire korrelálnak a hallgatók által észlelt hangulattal. További kutatási irányt jelent a hang egyéb, nem nyelvi jel-

lemzőnek (pl. hangerő, beszédtempó) a vizsgálata, illetve az egyes tényezők egymással való összefüggése.

## 6 Köszönetnyilvánítás

A munka a KAP15-059-1.1-ITK pályázat keretében valósulhatott meg, melyért a szerzők ezúton is köszönetüket fejezik ki.

## Bibliográfia

1. Fields, B.: Contextualize Your Listening: The Playlist as Recommendation Engine. PhD thesis, Goldsmiths, University of London, London (2011)
2. Hu, Xiao, Downie, J. Stephen: When Lyrics Outperform Audio for Music Mood Classification: A Feature Analysis. In Proceedings of the 10th International Conference on Music Information Retrieval , Utrecht, The Netherlands (2010) 619–624
3. Lukacs, G., Pethesné, D.B., Madocsai, B.: Impact of Personalized Audio Social Media on Social Networks. In XXXIII. Sunbelt Social Networks Conference of the International Network for Social Network Analysis Abstract Proceedings , Hamburg, Germany (2013) 210
4. Jani, M., Lukács, G., Takács, Gy.: Experimental Investigation of Transitions for Mixed Speech and Music Playlist Generation. In Proceedings of ACM International Conference on Multimedia Retrieval , Glasgow, United Kingdom (2014) 392–398
5. Benyeda, I., Jani, M., Lukács, G.: Beszéd-zene lejátszási listák nyelvtechnológiai vonatkozása. In Proc. of XI. Magyar Számítógépes Nyelvészeti Konferencia, MSZNY 2015 , Szeged, Hungary (2015) 257–268
6. Jani, M.: Fast Content Independent Playlist Generation for Streaming Media. In 12th ACS/IEEE International Conference on Computer Systems and Applications AICCSA 2015 , Marrakech, Morocco (2015) To appear
7. Strapparava, C., Valitutti, A.: WordNet-Affect: an Affective Extension of WordNet. In Proceedings of the 4th International Conference on Language Resources and Evaluation , Lisbon, Portugal (2004) 1083–1086
8. Pennebaker, J.W., Martha, E.F., Booth, R.J.: Linguistic Inquiry and Word Count. , Mahwah, NJ (2001)
9. Bradley, M.M., Lang, P.J.: Affective Norms for English Words (ANEW): Instruction manual and affective ratings, <http://www.uvm.edu/~pdodds/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf>, (1999)
10. Esuli, A., Sebastiani, F.: SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. In In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06 (2006) 417–422
11. Ekman, Paul: Facial expression of emotion. *Am. Psychol.* 48, (1993) 384–392
12. Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K.: The Kaldi Speech Recognition Toolkit. In IEEE 2011 Workshop on Automatic Speech Recognition and Understanding , Hilton Waikoloa Village, Big Island, Hawaii, US (2011)
13. Dodds, P.S., Clark, E.M., Desu, S., Frank, M.R., Reagan, A.J., Williams, J.R., Mitchell, L., Harris, K.D., Kloumann, I.M., Bagrow, J.P., Megerdumian, K., McMahon, M.T., Tivnan,

- B.F., Danforth, C.M.: Human language reveals a universal positivity bias. *Proc. Natl. Acad. Sci.* 112, (2015) 2389–2394
14. Lampos, V., Lansdall-Welfare, T., Araya, R., Cristianini, N.: Analysing Mood Patterns in the United Kingdom through Twitter Content. *CoRR*. abs/1304.5507, (2013)