



M Ű E G Y E T E M 1 7 8 2

Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Infokommunikációs hálózatok és alkalmazások

FŐKOMPONENS ELEMZÉS ALKALMAZÁSA ÉRZELEMFELISMERÉS TERÜLETÉN

Tulics Miklós Gábor

Villamosmérnöki szak, BSc képzés

KONZULENS

Dr. Vicsi Klára

BUDAPEST, 2012

Tartalomjegyzék

Összefoglaló	5
Abstract.....	6
1. Bevezetés	7
2. Érzelmek	8
3. Főkomponens elemzés elméleti háttere	10
3.1. Bevezetés	10
3.2. Szemléltető példa	10
3.3. Bázisváltás	12
3.3.1 Naiv bázis	12
3.3.2 A bázis megváltoztatása.....	13
3.3.3 Fennmaradó kérdések	15
3.4. A cél és a variancia	15
3.4.1 Zaj és rotáció.....	15
3.4.2 Redundancia.....	17
3.4.3 Kovariancia mátrix	18
3.4.4 Kovariancia mátrix diagonalizálása.....	20
3.4.5 Kikötések összefoglalása	21
3.5. Főkomponens elemzés megoldása a sajátvektorokra bontás módszerével.....	22
4. Érzelmi tartalmú hanganyag adatbázis.....	24
5. PCA Matlab-os megvalósítása	27
5.1. Bemenő adatok	27
5.2. A főkomponens elemzés Matlab-beli előkészítése	30
5.3. Az eredmények ismertetése	33
5.3.1 Düh.....	35
5.3.2 Öröm	39
5.3.3 Nevetve beszél	41
5.3.4 Semleges	43
5.3.5 Szomorú.....	48
5.3.6 Az összes érzelemre egyszerre végzett analízis.....	50
5.4. Összefoglalás	54
6. Felmerülő kérdések, további célok, egyéb tapasztalatok	55

7. Köszönetnyilvánítás	57
Irodalomjegyzék.....	58
Függelék.....	60

HALLGATÓI NYILATKOZAT

Alulírott **Tulics Miklós Gáriel**, szigorló hallgató kijelentem, hogy ezt a szakdolgozatot/ diplomatervet meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerző(k), cím, angol és magyar nyelvű tartalmi kivonat, készítés éve, konzulens(ek) neve) a BME VIK nyilvánosan hozzáférhető elektronikus formában, a munka teljes szövegét pedig az egyetem belső hálózatán keresztül (vagy hitelesített felhasználók számára) közzétegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetővé.

Kelt: Budapest, 2012. 12. 05.

.....
Tulics Miklós Gáriel

Összefoglaló

A korábbi féléves „Önálló laboratórium” munkám témája a meglévő érzelem adatbázis feldolgozásában és javításában való részvétel, gépi felismerővel (SVM) és a Praat programmal való megismerkedés, valamint a teljes mondatok érzelem-felismerésének vizsgálata volt.

Szakedolgozatomban az eddig tanultak fényében egy új eljárásra, a főkomponens elemzésre teszek kísérletet, melynek segítségével meghatározhatjuk, mely akusztikai paraméterek lényegesek az érzelem-felismerés során. Ennek segítségével remélhetőleg újabb optimalizációs lehetőségek nyílnak meg előttünk, amik által jobb eredményekre juthatunk. A feladat végrehajtásához a Beszédakusztikai Labor által rendelkezésemre bocsátott érzelmi tartalmú hanganyag adatbázist használtam.

A dolgozat első részében általánosan írok az érzelmekről, később az érzelemadatbázissal foglalkozom.

A munkám nagyobb része a főkomponens elemzés alapjainak leírásával foglalkozik, valamint annak gyakorlati megvalósításával. A PCA - t Matlab segítségével viszem át a gyakorlatba, majd szemléltetem ennek lépéseit.

A szakdolgozat végén az eredményül kapottakat értelmezem, illetve leírom az általam nyert tapasztalatokat.

Abstract

In the past semesters during „Individual Laboratory” I participated in the improvement of the existing emotion – database, I tried to get familiar with the speech recognizer machine (SVM) and the program called Praat. I also dealt with the examination of recognition of emotions on sentence-based speech units.

With this knowledge I will attempt a new procedure, the principal component analysis which is used to determine which acoustic parameters are relevant to emotion recognition. This will hopefully help further optimization opportunities and we can achieve better results. To fulfill my task I used the database based on emotions provided by the Speech Technology Laboratory.

In the first part of my thesis I discuss generally about emotions, then I turn to emotion database.

The biggest part of my work deals with description of the basis of principal component analysis and it's practical implementation. The implementation is done in Matlab, I illustrate the steps.

At the end of the thesis I interpret the results and write about my gathered experience.

1. Bevezetés

Érzelmekről nehéz beszélni, mert személyesek és szubjektívek. Számos olyan kutatás folyik azonban amelynek célja az érzelmek gépi felismerése, elemzése. Az érzelmek gépi felismerése történhet vizuális úton - arcmimika, gesztus és egyéb testmozgás detektálásával - illetve beszéd alapján. Munkám során egy olyan matematikai eljárást mutatok be - a főkomponens elemzést - amely alkalmas nagy adathalmazokban lineáris strukturális szabályszerűségek feltárására. Az eljárást alkalmazom olyan hangmintákból gyűjtött akusztikai paraméterek esetén, amelyek különböző érzelmeket hordoznak, ezáltal megvizsgálva, hogy az említett paraméterek hordoznak-e érzelmekre jellemző tulajdonságokat.

Előbb az érzelmek pszichológiai meghatározásának mai állását mutatom be, majd személyes érdeklődésem határterületéből egy keveset, mely nem tartozik szorosan a szakdolgozat témájához. Ezt követően rátérek a főkomponens elemzés elméleti alapjainak ismertetésére, továbbá az érzelmi tartalmú hanganyag adatbázis ismertetésével foglalkozom. Végül pedig a PCA futtatásának eredményét értelmezem. Munkámat meglehetősen megnehezítette, hogy magyar nyelvű irodalom alig található ebben a témában, így angol nyelvű szakirodalomra hagytam. Bízom abban, hogy ez a szakdolgozat megfelelő alap lesz azoknak, akik a jövőben főkomponens elemzéssel szeretnének foglalkozni.

2. Érzelmek

Az érzelem feltárása határvonalat képezhet a tudomány és a filozófia területén, ugyanis sokszor kapcsolatba hozzuk a lélekkel. Valójában agyunk felelős az érzelmek generálásáért [3]. Érzelmi reakcióink szituációfüggők és sikerült kimutatni, hogy a hipotalamusz, amygdala vagy a limbikus rendszer fontos szerepet játszanak ezek kialakulásában. Amit mi boldogságnak élünk meg, azt tulajdonképpen egy dopamin nevű neurotranszmitter kelt (köznyelven „boldogsághormon”). Ezt egy specifikusan kialakult rendszer, mondhatni egy „jutalmazó rendszer” adagolja a szervezetbe [1]. Tehát az érzelmek, mint jól meghatározott fizikai és kémiai folyamatok jelentkeznek az emberi szervezetben.

Az evolúció során az érzelemkeltő és befolyásoló/kontrolláló központok is komplexebbé váltak, akárcsak az agyunk többi része. Ezáltal érzelmeink is kifinomultabbak és többszínűbbek lettek. A darwini evolúció elmélet szerint számos közismert gesztus biológiai alapja vezethető vissza adaptív állati magatartásformára. Ezek később az embereknél szimbolikus értelmet nyernek. Így bizonyos emocionális viselkedési formák csökevényként maradtak meg: „A kifejezés egyik-másik alakját alig érthetjük meg; így például a rendkívüli félelem hatása alatt a haj égne meredését, az őrjöngő düh hatása alatt a fogak vicsorgatását, hacsak fel nem tesszük, hogy az ember hajdan alsóbbrendű és állatszerű állapotban élt” [3][4].

Feltételezhetjük, hogy a Darwin által említett érzelmi mechanizmusok és kísérő jelenségeik állati mivoltunkból erednek. Bizonyos érzelmek arckifejezései, például a boldogság, szomorúság vagy a düh különböző kultúrákban is fellelhetőek és azonosak. Ezt elsőként Carroll Ellis Izard mutatta ki 1968-ban [7][3]. Ennek magyarázata az lehet, hogy evolúció során genetikailag belénk ívódtak bizonyos érzelmi mintázatok, amelyek a természetes szelekció során hasznosnak bizonyultak. Így lehet párhuzamot vonni ember és állat között is.

Az emóciókat többféleképpen próbálták megragadni. Az elméletek egy része az érzelmeket egy X-Y koordináta-rendszerben helyezi el, a polaritás (pozitív vagy negatív) és az intenzitás (gyenge vagy erős) függvényében [8][3].

Egy másik elméleti irányzat az alapvető affektív mintázatokra épít. Az univerzálisan megjelenő érzelmekre fektetik a hangsúlyt és feltételezik, hogy ezek

biológiaiilag kialakult alapérzelmek. Ezen elmélet szerint az érzelmek gyakorlatilag adaptációs mechanizmusok, amelyek az evolúció során a problémák kiküszöbölésére szolgáltak [9][3].

Az egyik legelfogadottabb és legátfogóbb meghatározás Frijda nevéhez fűződik: „Az érzelmet kiváltó ok általában abban rejlik, hogy a személy egy fontos dologgal (célal) kapcsolatban tudatos vagy tudattalanul relevánsnak tart egy eseményt. Ha az esemény segíti az ügy előrehaladását, az érzelemnek pozitív töltete lesz, ha gátolja azt, negatív”[10][3].

Az érzelmeknek cselekvéskészség beindító tulajdonságuk van. Bizonyos típusú cselekvésminták sürgősségérzettel párosítva elsőbbséghez juttathatnak, mások megzavarhatnak vagy versengést eredményezhetnek.

Az emóciókat mentális állapotként éljük át, amiket néha testi változások, kifejezések kísérnek [11].

Az érzelmek információkat közvetítenek rólunk, lelki állapotunkra, személyhez vagy a helyzethez való viszonyunkra utalnak. Cselekvéseinkre vagy viselkedési tendenciánkra adhatnak predikciót[3].

Munkám során konkrét érzelmi kategóriákra bontott hanganyagokkal dolgoztam. A hanganyagok érzelmi tartalma objektivitásra törekedve, mégis szubjektívan lett megállapítva. A beszédhangokat tartalmazó adatbázisból tanulmányozhattam az érzelem akusztikai paramétereit. Ez az általános módja az érzelmek osztályozásának és a gépi felismerésnek.

3. Főkomponens elemzés elméleti háttere

3.1. Bevezetés

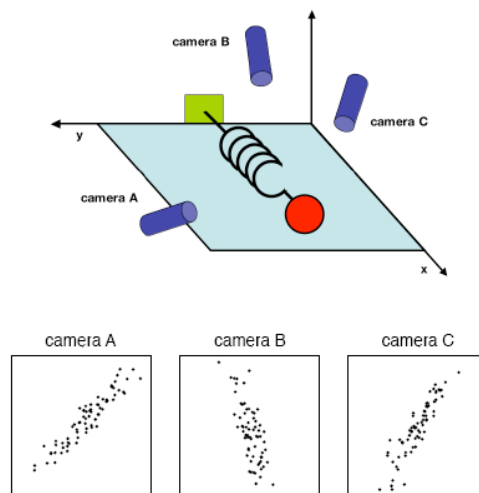
A főkomponens elemzés, vagy angolul Principal Component Analysis (**PCA**), a modern adatelemzés egy tartóoszlopa, amelyet széles körben használnak az idegtudománytól a számítógépes grafikáig, viszont olykor elméleti háttere nem feltétlenül tisztázott. Használata elterjedt, mert egy egyszerű, nem paraméteres módszer a releváns információ kinyerésére, a sokszor átláthatatlan adathalmazból [3].

A dolgozatom során igyekszem feltárni a főkomponens elemzés mibenlétét, nem kihagyva a szemléletes példákat és a matematikai leírást.

3.2. Szemléltető példa

Ahhoz, hogy a főkomponens analízist jobban megértsük, képzeljük el, hogy vizsgálók vagyunk. Szeretnénk megérteni egy jelenséget, úgy, hogy néhány kvantitást mérünk (spektrum, feszültség, szórás, stb.) egy rendszeren belül. Természetesen nem jövünk rá, mi történik pontosan a rendszer belsejében, mert az adathalmazunk felhősen, minden információt (lényegit és jelentéktelent) egyszerre tartalmaz, nem egyértelmű és redundáns. Ez nem egy triviális probléma, inkább egy fundamentális akadály az empirikus tudomány keretein belül. Erre számos példát mondhatunk komplex rendszerekben, olyanokban, mint az idegtudomány, web-indexelés, meteorológia vagy oceanográfia. A mérni kívánt változók száma nagy lehet, megtevesztők és nehezen értelmezhetőek lehetnek, és a köztük lévő reláció néha nem egyértelmű.

Nézzük a következő egyszerű példát az 1. ábráról. Tegyük fel, hogy egy ideális rugó mozgását vizsgáljuk. A rendszer egy „m” tömegű labdát tartalmaz, ami egy tömegtelen súrlódásmentes rugóhoz van illesztve. A labda (a kinyújtott rugó helyzetében) kis távolságra a tökéletes dinamikus egyensúlyi helyzetből van elengedve. Mivel a rugó ideális, az x tengely mentén, adott frekvencián, az egyensúlyi helyzet körül fog oszcillálni.



**1. ábra – egy rugóra helyezett
labda oszcillálásának
felvétele A,B és C
kameraállásából.**

Ez a fizikában egy standard probléma, amiben az x irányú mozgás kifejezhető és megoldható egy időfüggvénnyel. Más szavakkal, a rendszer dinamikája kifejezhető egyetlen x szerinti függvénnyel.

A gyakorlati helyzetekhez hasonlóan nem tudjuk a rendszer tényleges tulajdonságait, nem tudjuk melyik vagy hány tengely és dimenzió fontos a mérésünkhöz. Úgy döntünk, hogy a labda pozícióját egy háromdimenziós térben akarjuk vizsgálni (abból kifolyólag, hogy egy háromdimenziós világban élünk). Három kamerát helyezünk el az elrendezésünk körül. 120 Hz-es frekvencián a kamerák egy-egy kétdimenziós felvételt vesznek fel a labda helyzetéről. A tudatlanságunkból az is következik, hogy az igazi x, y, z koordinátákat sem ismerjük, így tetszőleges tengelyeket és szabadon választott szögeket rendelünk a rendszerünkhöz. A mérésünkben a tengelyek közötti szög nem feltétlenül 90° . Ha több percig folytatjuk a felvételt, a kérdés marad: hogyan kapjuk meg a tetszőlegesen választott mérési rendszerünkből egy egyszerű x -től függő egyenletet?

Ha lehetőségünk lett volna rá, x tengely mentén is végezhetjük volna a felvételt. Azonban akár a valós életben, nem tudjuk melyik mérés reflektálja legjobban a kérdéses rendszert, így néhol több dimenziót rögzítünk, mint amennyire szükségünk van.

Valamint számolnunk kell a zajjal is. A labdás szemléltetésben ez a levegőt, a súrlódást, a kamerák helytelen pozícióját és a nem ideális rugót jelentheti.

3.3. Bázisváltás

A főkomponens elemzés célja a legjelentősegteljesebb bázis felismerése, hogy újraértelmezessük az adathalmazunkat. Azt reméljük, hogy ezzel az újraértelmezéssel ki tudjuk szűrni a zajt és rejtett struktúrákra bukkanhatunk. A rugós példánkban a főkomponens elemzés célja, hogy meghatározza: a mozgás x tengely mentén van. Más szavakkal a főkomponens elemzés célja, hogy meghatározza, hogy $\hat{x}$, x menti bázisvektor a lényeges dimenzió. Ez lehetőséget nyújt arra, hogy a vizsgáló felismerje, mely adatok fontosak, melyek redundánsak és melyik zaj.

3.3.1 Naiv bázis

Célunk pontosabb definiálása érdekében pontosabban kell meghatároznunk az adathalmazunkat is. Minden időmintát (vagy kísérleti próbánkat) egyéni mintaként értelmezünk. Minden egyes alkalommal rögzítünk több mérést (pl. feszültségérték, pozíció, szórás, stb.). Az adathalmazunkban, egy időpillanatban, az „A” kamera egy megfelelő labdapozíciót rögzít (x_A, y_A) . Egy minta tehát mindig hat dimenziós oszlopvektorral fejezhető ki

$$\vec{X} = \begin{bmatrix} x_A \\ y_A \\ x_B \\ y_B \\ x_C \\ y_C \end{bmatrix},$$

amelyben minden kamera egy két-dimenziós adatpárral járul hozzá az $\vec{X}$ vektorunkhoz. Ha tíz percig rögzítjük a labda pozícióját 120 Hz-es frekvenciával, akkor $10 \times 60 \times 120 = 72000$, ilyen vektorunk lesz.

Matematikai kifejezésekkel: minden minta ($\vec{X}$) egy m - dimenziós vektor, ahol „ m ” a mérés típusainak a száma. Minden minta egy vektor, amely egy m - dimenziós vektortérben fekszik, valamilyen ortonormált bázissal felépítve. Az ortonormált bázis definíciója: Az $e_1, \dots, e_n$ vektorrendszert V ortonormált bázisának nevezzük, ha minimális generátorrendszer V -ben, minden vektora egység hosszúságú és bármely két vektora egymásra merőleges. V egy vektortér, amelyen definiálva van egy skalársorozat.[5] Lineáris algebrából tudjuk, hogy minden mérésvektorunk lineáris kombinációja az egység bázisvektorunknak. A kérdés az, hogy mi az ortonormált bázisvektor? [3]

Tételezzük fel, hogy a labdás példánkban, csak az „A” kamerát használtuk. Mi az ortonormált bázisa (x_A, y_A) - nak? Egy triviális választás $\{(1,0),(0,1)\}$ lenne, de miért válasszuk ezt és miért nem ennek $\{(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), (\frac{-\sqrt{2}}{2}, \frac{-\sqrt{2}}{2})\}$ rotációját? Erre az a magyarázat, hogy a naiv bázis reflektálja a gyűjtött adatok módját.

Kétdimenziós esetben $\{(1,0),(0,1)\}$ újraértelmezhető két individuális sorvektorként. Ha mátrixként akarjuk kifejezni, akkor egy 2×2 -es identitásmátrixot kapunk (I). Az m - dimenziós esetben egy $m \times m$ -es identitásmátrixról beszélhetünk

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = I$$

ahol minden sor egy ortonormált bázis vektor (b_i) m darab komponenssel.

A naiv bázis hasznos kiindulópontnak számít. Minden adatunk ebben a bázisban lett rögzítve és triviálisan kifejezhető $\{b_i\}$ lineáris kombinációjaként.

3.3.2 A bázis megváltoztatása

A főkomponens elemzés kérdése a következő lehet: Létezik-e egy másik bázis, amely lineáris kombinációja az eredetinek, és ami a legjobban kifejezi az adathalmazunkat?

A főkomponens elemzés (PCA) szigorú előfeltétele: a linearitás. Ez csökkenti az esetleges bázisok számát, egyszerűsítve a problémát. Ezzel a kitétellel a PCA korlátozva van, mégpedig úgy, hogy az adathalmaz újraértelmezése a bázisvektorok lineáris kombinációja legyen.

Legyen X az eredeti adathalmazunk, ahol minden oszlop egyetlen mintája (vagy időpillanata) az adatnak (pl. $\vec{X}$). A labdás példában X egy $m \times n$ -es mátrix, ahol $m=6$ és $n=72000$ (6 darab pozíció adat, 10 percig rögzítve 120 Hz frekvencián, ami $10 \times 60 \times 120 = 72000$ darabot jelent). Legyen Y egy másik $m \times n$ -es mátrix, ami összefüggésben van P lineáris transzformációval. X az eredetileg rögzített adathalmaz és Y az adathalmaz új reprezentációja.

$$PX = Y$$

Vezessük be a következő mennyiségeket:

- p_i P sorai
- x_i X oszlopai
- y_i Y oszlopai

A $PX = Y$ egyenletnek több interpretációja van:

1. P az a mátrix, ami X -et Y -ba transzformálja
2. Geometriailag P egy rotáció és egy nyújtás, ami szintúgy X -et Y -ba transzformálja
3. P sorai, $\{p_1, \dots, p_m\}$, egy új bázisvektor készlet, amely kiterjeszti X oszlopait

A későbbi értelmezése még nem világos, de PX eredményeként felírható:

$$PX = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$$

$$Y = \begin{bmatrix} p_1 \cdot x_1 & \cdots & p_1 \cdot x_n \\ \vdots & \ddots & \vdots \\ p_m \cdot x_1 & \cdots & p_m \cdot x_n \end{bmatrix}.$$

Így Y oszlopai:

$$y_i = \begin{bmatrix} p_1 \cdot x_i \\ \vdots \\ p_m \cdot x_i \end{bmatrix}.$$

Felismerhetjük, hogy minden egyes y_i koefficiens egy x_i által, a hozzátartozó P sorral keltett pont-produktum. Más szavakkal, y_i j - edik tényezője, P j - edik sorának a vetülete. Ez tehát felírható egy egyenlettel, amelyben y_i a $\{p_1, \dots, p_m\}$ bázisra való vetülete. Következésképpen, P sorai az új bázisvektor-készlet, X oszlopainak a kiterjesztésére.

3.3.3 Fennmaradó kérdések

Linearitást előfeltételezve a probléma a helyes bázisváltásra csökken. Ebben a transzformációban a sorvektorok $\{p_1, \dots, p_m\}$ lesznek X főkomponensei.

Mi a legjobb mód X újraértelmezésére? Mi a P bázisk megválasztásának a legmegfelelőbb módja?

Előbb azt kell eldöntenünk, hogy Y milyen tulajdonságait szeretnénk előnyben részesíteni. A linearitáson kívül további kitételeket kell tennünk, hogy értelmes eredményhez jussunk. A következőkben ezeket fogom tárgyalni.

3.4. A cél és a variancia

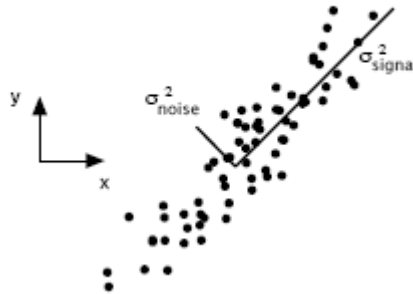
A legfontosabb kérdés: mit jelent az adathalmaz legjobb kifejezése? Erre a kérdésre próbálunk meg választ adni további feltételek bevonásával.

3.4.1 Zaj és rotáció

Bármely mérés zaja alacsony kell, hogy legyen, különben (az elemzés technikájától függetlenül) nem tudunk a jelből hasznos információt kinyerni. Nem létezik abszolútan kifejező skála a zajra, viszont annak összes típusa szoros összefüggésben van a jel erősségével. Egy gyakori mérési módszer a jel-zaj viszony mérése, angolul: signal-to-noise ratio (SNR), vagy kifejezhető a szórásnégyzetek (σ^2) hányadosaként:

$$SNR = \frac{\sigma_{signal}^2}{\sigma_{noise}^2}.$$

Magas SNR ($\gg 1$) nagy pontosságú mérésre utal, ennek függvényében az alacsony SNR zajos adatot feltételez.



2. ábra – „A” kamera szimulált adata. A jel és a zaj varianciáját ($\sigma_{signal}^2, \sigma_{noise}^2$) két egyenes ábrázolja grafikusan. Jegyezzük meg, hogy a variancia legnagyobb értéke nem a felvétel bázisa irányában helyezkedik el, mindinkább az adatfelhőre legjobban illeszkedő egyenes irányában.

Tekintsük meg a 2. ábrán lévő adatokat, amit az „A” kamerával rögzítettünk. A rugó egyenes vonalon mozog, így minden egyes kamera egyenes vonalú mozgást kell, hogy rögzítsen. Evégett zajnak számít minden az egyenes vonalú mozgástól való szélességi kitérés. A jel és a zaj szórásnégyzete (varianciája) fel van tüntetve az ábrán. A két mennyiség hányadosa a pontfelhő soványságát adja meg: az eredmény egy vékony vonal ha $SNR \gg 1$, egy kör ha $SNR=1$. Ha jó méréseket végzünk, feltételezhetjük, hogy az adatterünkben a legnagyobb varianciájú irány tartalmazza azon tulajdonságokat, amelyek érdekelnek bennünket. A 2. ábrában a legnagyobb varianciájú irány nem $\hat{x}_A = (1,0)$ vagy $\hat{y}_A = (0,1)$, hanem a felhő menti irány tengelye mentén van. Ily módon, feltételezhetjük, hogy a feltárni kívánt irányhoz tartozik a legnagyobb variancia és feltételezhetően a legnagyobb jel-zaj viszony érték is.

Ez a feltételezés azt sugallja, hogy a keresett bázis nem a naiv bázis, mert ennek irányaihoz (x_A, y_A) nem a legnagyobb szórásnégyzet tartozik. Ha maximalizáljuk a varianciát (és így az SNR-t is) megtalálhatjuk a megfelelő rotációját a naiv bázisnak. Ez

az intuíció egyezik azzal, hogy megtaláljuk a 2. ábrában σ_{signal}^2 irányát. A 2. ábrán lévő 2 dimenziós esetben a legnagyobb varianciájú irány megegyezik az adatfelhő vonalas lefedésével. Így a naiv bázis rotálása a legjobban illeszkedő irányhoz párhuzamosan felfedné a rugó mozgási irányát 2 dimenziós esetre. Hogyan lehet ezt az elgondolást általánosítani tetszőleges számú dimenziójú esetre? A kérdés megválaszolása előtt újabb szempontot kell vizsgálni.

3.4.2 Redundancia

A 2. ábra egy újabb befolyásoló tényezőt sugall: a redundanciát. A rugós szemléltetésünkben több szenzor rögzíti ugyanazt a dinamikai jelenséget. A 2. ábrát újra szemlélve megkérdezhetjük magunktól, hogy valóban szükséges-e 2 változót rögzíteni.



3. ábra – Két különálló r_1 és r_2 mérés lehetséges spektrumai. A bal szélén látható két mérés korrelálatlan, mivel az egyik nem jósolja meg a másik eredményét, ezzel szemben a jobb szélén látható két mérés erős korreláltságot mutat, magas redundanciára utalva.

A 3. ábra két tetszőleges mérési típus (r_1 és r_2) két lehetséges ábrázolási módjának távolságát, viszonyát adhatja meg. A baloldali panel két felvétele között nincs látszólagos összefüggés. Amiért r_2 nem jósolja r_1 -et, mondhatjuk, hogy r_1 és r_2 korrelálatlanok.

A szélső jobb oldali panelen két erősen korrelált adathalmazt láthatunk. Erre több magyarázat is lehet:

- (x_A, x_B) képe lehet, ha A és B kamera nagyon közel fekszenek egymás mellett

- $(x_A, \hat{x}_A)$, ha például x_A méterben adott, $\hat{x}_A$ pedig inchben

A szélső jobb oldali panelen lévő kép szerencsésebb lett volna ha, csak egy változót rögzítene, mivel r_2 -ből kiszámolható r_1 (vagy fordítva) a legjobban illeszkedő vonalat használva. Csak egyetlenegy változót rögzítve tömörebb adathalmazt eredményez és eggyel kevesebb szenzoros mérést igényel ($2 \rightarrow 1$). Ez a dimenziócsökkentés gondolatának az alapja.

3.4.3 Kovariancia mátrix

Kétdimenziós esetben könnyű felismerni a redundáns eseteket a legjobban illeszkedő vonal meredekségének meghatározásával és az illeszkedés minőségének megítélésével. Hogyan általánosítható ez a gondolat tetszőleges dimenziós esetekre? Nézzük meg két készlet mérési adatait melyeknek átlaga nulla

$$A = \{a_1, a_2, \dots, a_n\} \quad B = \{b_1, b_2, \dots, b_n\},$$

melyben az index jelzi az elemszámot. $\mathbf{A}$ és $\mathbf{B}$ varianciája definiálható

$$\sigma_A^2 = \frac{1}{n} \sum_i a_i^2$$

$$\sigma_B^2 = \frac{1}{n} \sum_i b_i^2,$$

$\mathbf{A}$ és $\mathbf{B}$ közötti kovariancia pedig:

$$\sigma_{AB}^2 = \frac{1}{n} \sum_i a_i b_i^2.$$

A kovariancia két változó lineáris kapcsolatának mértékét méri. Nagy, pozitív érték pozitívan korrelált adatokra utal. Ennek ellentéte is igaz, nagy, negatív érték negatívan korrelált adatokat feltételez. A kovariancia abszolút nagysága méri a redundancia mértékét. A kovariancia néhány tulajdonsága:

- Ha σ_{AB} nulla, akkor és csak akkor, ha **A** és **B** korrelálatlanok
- $\sigma_{AB}^2 = \sigma_A^2$, akkor, ha **A=B**

kifejezhetjük **A** és **B**-t nekik megfelelő sorvektorokként

$$a = \{a_1, a_2, \dots, a_n\}$$

$$b = \{b_1, b_2, \dots, b_n\}$$

így a fenti összefüggés alapján megadható a kovariancia

$$\sigma_{AB}^2 = \frac{1}{n} ab^T.$$

Gyakorlatban viszont a normalizált formát használják, ennek magyarázata meghaladja a dolgozat kereteit

$$\sigma_{AB}^2 = \frac{1}{n-1} \sum_i a_i b_i.$$

Végül általánosítani tudunk két darab vektorról tetszőleges számúra. Átnevezzük a -t és b -t x_1, x_2 -re, valamint további vektorokat adunk hozzá $x_3, \dots, x_m$ indexekkel. Definiálunk így egy $m \times n$ -es mátrixot: X (m sor és n oszlop)

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}.$$

X sorai egy mérési típushoz tartoznak, oszlopai pedig, egy adott indexű mérés eredményeit tartalmazza (összes típusból). Most definiálhatjuk a kovariancia mátrixot:

$$C_X \equiv \frac{1}{n} XX^T.$$

A C_X mátrix ij -edik eleme egy pont produktum az i -edik vektor mérési típusa és a j -edik vektor mérési típusa között. Néhány dolgot összefoglalhatunk:

- C_X négyzetes szimmetrikus $m \times m$ -es mátrix

- A mátrix átlójában különböző mérési típusok varianciája található
- Az átlón kívüli elemek különböző mérési típusok kovarianciája

C_X megragadja az összes mérési típus közötti kovarianciát. A kovariancia értékek tükrözik a redundanciát és a zajt a mérésünkben.

- az átlóban, mint az előfeltétel értelmében, a nagy értékek érdekes struktúrának felelnek meg
- az átlón kívüli értékek nagy mennyiségei magas redundanciának felelnek meg

Tételezzük fel, hogy lehetőségünk van C_X manipulálására. Legyen a manipulált mátrixunk neve C_Y . Milyen tulajdonságait szeretnénk C_Y -nak optimalizálni?

3.4.4 Kovariancia mátrix diagonalizálása

Összefoglalhatjuk az előző két fejezetünket azzal, hogy kifejezzük a céljainkat: (1) a redundancia minimalizálása a kovariancia nagyságát mérve, és (2) a varianciát mérve a jel maximalizálása. Hogyan nézne ki az optimális C_Y mátrix?

- Minden átlón kívüli elem értéke nullával kellene, hogy egyenlő legyen, valamint C_Y -nak diagonális mátrixnak kell lennie. Másképp mondva, Y korrelálatlan
- Minden folytonos Y -béli dimenzió nagyságrendi sorrendbe kellene, hogy legyen, a variancia szerint

Sok módszer létezik C_Y diagonalizálására. Érdekességként megemlítem, hogy a PCA a módszerek legkönnyebb fajtáját választja: a főkomponens elemzés kiköti, hogy az összes bázis vektor ortonormált legyen $\{p_1, p_2, \dots, p_m\}$, ebből az következik, hogy P ortonormált mátrix.

Képzeld el a főkomponens elemzés működését. A 2. ábránban P egy általánosított rotációként funkcionál, hogy a bázisát a maximális varianciájú irányhoz igazítsa. Többdimenziós esetben:

1. válasszunk egy normalizált irányt m dimenziójú térben, amerre X varianciája maximális. Mentsük el ezt a vektort p_1 -ként
2. keressünk egy másik irányt, amely mellett a variancia maximális, azonban az ortonormált állapotot figyelembe véve szorítkozzunk azokra az irányokra, amelyek ortogonálisak az előbb kiválasztott irányokra. Mentsük el a vektort p_i -ként
3. addig ismételjük ezt a folyamatot, amíg m vektor ki lesz választva

Az eredményként kapott, majd sorba rendezett p elemkészletünkől lesznek a főkomponenseink.

Többször ez az algoritmus működik, bár ez nem az igazi oka, amiért az ortonormalitás előírt. Nagyság szerinti varianciák sorba rendezésével módszert kaptunk a főkomponensek fontosságának megítélésére. Névlegesen, minden p_i irányhoz tartozó variancia meghatározza mennyire dominánsak az egyes irányok, azáltal, hogy nagyságrendi sorrendbe rendezzük a p_i bázisvektorokat a hozzájuk tartozó varianciák szerint.

3.4.5 Kikötések összefoglalása

1. *Linearitás*

A linearitás bázisváltássá keretezi a problémát.

2. *A nagy változóknak fontos struktúrájuk van*

Ez a kikötés tartalmazza a gondolatot, miszerint az adatnak nagy az SNR-je. Innen következik, hogy a nagyobb főkomponenshez társult változók érdekes struktúrát, a kisebb változóúak viszont zajt jelentenek.

3. *A főkomponensek ortogonálisak*

Ez a kikötés lehetőséget nyújt arra, hogy a főkomponens elemzést lineáris algebrában ismert technikákkal kezelhessük. A következő fejezetekben ezeket fogjuk tárgyalni.

3.5. Főkomponens elemzés megoldása a sajátvektorokra bontás módszerével

Az adathalmazunk X , ami egy $m \times n$ -es mátrix, m a mérési típusok száma és n a mérések száma (m oszlop, n sor). A célunk a következő:

Találjunk valamilyen ortonormált mátrixot (P) amely eleget tesz a $Y = PX$ egyenletnek úgy, hogy $C_Y = \frac{1}{n} YY^T$ diagonális mátrix legyen. A P sorai X főkomponensei.

Azzal kezdjük, hogy kifejezzük C_Y - t az ismeretlennel.

$$\begin{aligned} C_Y &= \frac{1}{n} YY^T \\ &= \frac{1}{n} (PX)(PX)^T \\ &= \frac{1}{n} PXX^T P^T \\ &= P\left(\frac{1}{n} XX^T\right)P^T \\ C_Y &= PC_X P^T \end{aligned}$$

Jegyezzük itt meg, hogy az előző részben X kovariancia mátrixa meg lett határozva.

Tervünk annak felismerése, hogy bármely A szimmetrikus mátrix egy sajátvektorokból álló ortogonális mátrix által van diagonalizálva (lásd a Függelék 3-as és 4-es tételét). Egy szimmetrikus mátrixra a 4-es tétel kimondja, hogy $A = EDE^T$, ahol D egy diagonális mátrix és E egy sajátvektorokból álló mátrix, A rendezett oszlopaiból.

Most jön a trükk. P mátrixot úgy választjuk ki, hogy minden p_i sora $\frac{1}{n}XX^T$ sajátvektora legyen. Ezzel a választással $P \equiv E^T$. Ezzel a relációval és a Függelék 1. tételével ki tudjuk fejezni C_y -t.

$$\begin{aligned} C_y &= PC_xP^T \\ &= P(E^TDE)P^T \\ &= P(P^TDP)P^T \\ &= (PP^T)D(PP^T) \\ &= (PP^{-1})D(PP^{-1}) \\ C_y &= D \end{aligned}$$

Nyilvánvaló, hogy P választása diagonalizálja C_y -t. Ez volt a főkomponens elemzés célja. Összegezhetjük a PCA eredményét P és C_y mátrixban.

- X főkomponensei $C_x = \frac{1}{n}XX^T$ sajátvektorai.
- C_y i -edik diagonális értéke p_i menti varianciája X -nek

Számítógépes környezetben egy X adathalmazon végzett főkomponens elemzés velejárója: (1) minden méréstípusból az átlagérték levonása és (2) C_x sajátértékeinek a kiszámítása. Ezen megoldás Matlab kód formájában mellékelve a Függelékben található.

4. Érzelmi tartalmú hanganyag adatbázis

Ebben a fejezetben azon paramétereket mutatom be, amelyeken a főkomponens elemzést akarjuk elvégezni. Tehát a PCA fent említett X mátrixának elemeit tárjuk fel, mely inputként szolgál. A főkomponens elemzés célja a dimenziócsökkentés és tömörítés, azaz meg akarjuk fejteni, hogy mely paraméterek azok, amelyek a legnagyobb súllyal szólnak bele az adathalmazunk információtartalmának leírásába, valamint melyek a leghasználatóbbak a gépi érzelem-osztályozás során. Ha ezeket a bemeneteket megtaláljuk, akkor optimalizálni tudjuk a mérésünket, ki tudjuk szűrni a zajokat, a feladatot egyszerűsíteni tudjuk, ezáltal jobb eredményekre számíthatunk a jövőben.

Öt fontos paramétert kell figyelembe venni egy érzelmi tartalmú hanganyag adatbázis során, ezek az alábbiak: érvényességi tartomány (scope), spontaneitás (naturalness), kontextus (context), leírók (descriptors), hozzáférés (accessibility) [13].

E dolgozat tárgya az akusztikai paraméterek vizsgálata. Sokáig kérdés volt, hogy az ember képes-e csupán hang alapján az érzelem felismerésére. A nyolcvanas évek végére bebizonyították, hogy az ember 60%-os pontossággal képes hang alapján felismerni a beszélő érzelmeit [13][14]. Ez az arány javul, ha a hallgató ismeri a beszélőt, és ha előre megadott kategóriák közül kell választania. Így az emberi érzelem felismerésének pontosságát 60-80%-os pontosságra teszik [15][16].

A kérdés tehát, hogy mely akusztikai jellemzők tükrözik leginkább a hang érzelmi töltését, szakdolgozatomra vonatkoztatva a kérdést, milyen mérések paramétereire akarjuk a főkomponens elemzést végezni?

A cikkek alapján két csoportra oszthatjuk a jellemzőket:

Alacsony szintű jellemzők:

Ide tartozik az alaphang (F_0), a hangintenzitás és a beszéd tempója, rövid idejű keretenként számolt akusztikai jellemzők. Köztudott, hogy a férfiak alaphangjának értéke 50-250 Hz közé tehető, a nőké 150-300 Hz intervallumba, a gyermekek esetében pedig 350-500 Hz között változik. Fontos megemlíteni, hogy személyenként és időben

is változik az értéke. A hang időtartama a tempóra van hatással, illetve az intenzitás, és annak időbeli változása a nyomatékkal van kapcsolatban.

Magas szintű jellemzők:

Ezek az alacsony szintű jellemzőkből képzett értékek. Eredményezik, például a prozódiai sajátosságot, információt hordoznak az intonációról, tempóról, hangosságról, színekpi információkról [17].

Bense és Scherer három jellemző mentén vizsgálta a beszédhang érzelmi töltését, ezek: az energia, az alaphang és a beszéd tempója [13][14].

Ezekből kiindulva képezhetünk magas szintű jellemzőket: az alaphang változásának átlaga, maximuma, illetve minimuma, alaphang deriváltjának változása, vagy a formánsok pozíciójának átlaga (F_1 és F_2 formánsoké) is ilyen lehet. Továbbá: intenzitás átlagának változása, maximális, minimális értéke, intenzitás deriváltjának változása.

Néhány érzelmekategóriának a jellemzője: [13] [18]

Harag:

Az alaphang és az intenzitás értéke emelkedik, beszédtempó gyorsul, az alaphang széles tartományának felölelése jellemzi ezt a kategóriát (minimális és maximális értéke alapján).

Félelem:

Az alaphang átlagának értéke emelkedik, a magas frekvenciás értékek növekednek, a beszédtempó gyorsul, és ugyancsak az alaphang széles tartományának a felölelése a jellemző.

Bánat:

Az alaphang átlaga, tartománya és intenzitása egyaránt csökken. A magas frekvenciás együtthatók is csökkennek, a beszédtempó pedig lassul.

Öröm:

Alaphang átlaga emelkedik, változása és tartománya is növekedik, valamint az intenzitás átlaga is nő. Magas frekvenciás együtthatók emelkednek és gyorsul a beszédtempó.

Minden egyes kategóriánál kiemelendő, hogy bizonyos mérési eredmények csak részben reprezentálják a fent leírtakat. Vannak esetek, amikor teljesen ellentmondanak a fent leírtaknak. Ez abból adódik, hogy a kategóriáknak léteznek alkategóriái, amelyek nagyon távol esnek az akusztikai paraméterek terén (például harag esetén a „hot-anger” és a „cold-anger”, vagy félelem esetén a kétségbeesés és az aggodalom).

Az irodalom ismertetésének végéhez értünk. Tisztáztuk a főkomponens elemzés elméleti alapjait és megnéztük milyen természetű a bemeneti adatunk. A következőkben a PCA számítógépes környezetben való megvalósítását taglalom.

5. PCA Matlab-os megvalósítása

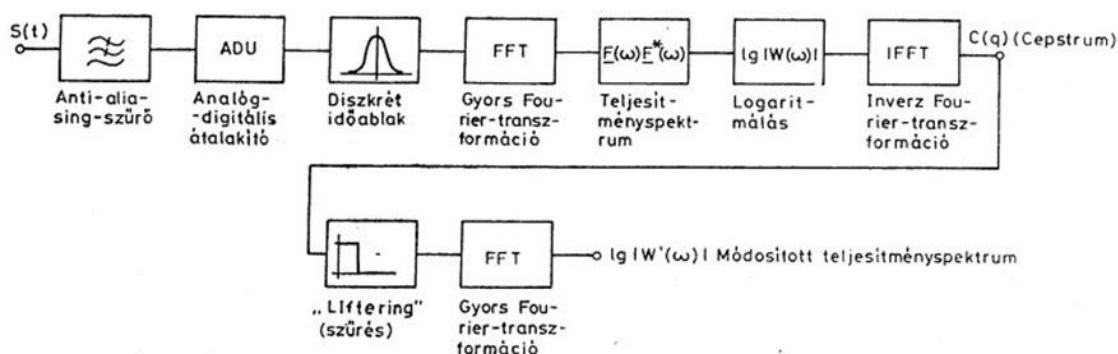
5.1. Bemenő adatok

A bemeneti vektor, a fentebb leírtak alapján, elég sok fajta paramétert tartalmazhat. A főkomponens elemzést a bemeneti vektorokra Matlabbal fogom elvégezni, a már korábban is hivatkozott, a függelékben szereplő kód segítségével.

Az analízist a négy alapérzelemre végzem el, külön-külön minden érzelemre, viszont külön kategóriát képez a beszéd közben nevetést tartalmazó minták. Tehát összesen öt érzelmi kategóriára fogom elvégezni a PCA-t. A négy alapérzelem: semleges (Neutral, N), öröm (Joy, J), düh (Angry, A), bánat (Sad, S).

A mérési típusokat, vagy akusztikai paramétereket, egy szöveges fájl tartalmazza. Ezek intonációs frázisonként, időegységenként számolt, származtatott (statisztikai) értékek, magas szintű (statisztikai függvényértékek) jellemzők. Szám szerint 210 akusztikai jellemzőt tartalmaz, azaz a négy alapérzelemre ennyi fajta paraméter áll a rendelkezésemre.

A 210 darab paraméter a következőképpen áll össze. Az érzelmek osztályozására a következő mennyiségeket használtam. Színképi jellemzőkként tizenkét darab MFC koefficiens számoltam. Ezen értékeket a teljesítményspektrum logaritmusának inverz Fourier transzformáltjából nyertem ki [19]. Ezt nevezzük Kepsztrum-analízisnek. Ennek előállítását egy ábrán szemléltetem:



4. ábra - A kepsztrum, valamint a spektrális burkoló görbe meghatározása

A következő mennyiség a Harmonics-to-Noise Ratio (HNR) [12]. A HNR az akusztikus periodicitás mértékét mutatja. Mértékegysége a dB . Ha a jel energiatartalma 99%-ban a periodikus részben található és 1% a zaj, akkor a HNR a következő képlettel számítható:

$$HNR = 10 \cdot \log_{10} \cdot \left(\frac{99}{1} \right) = 20 dB.$$

Ha az értéke nulla, akkor az energia egyenlően oszlik el a harmonikus részben és a zajban. A HNR jel-zaj viszony meghatározására, vagy beszédminőség leírására alkalmazható.

A HNR és a tizenkét darab MFC koefficiensen kívül rendelkezésemre áll még az energia és az alaphang, mint akusztikai paraméter. E tizenöt mennyiségünkön túl adva vannak a mennyiségek deriváltjai is. Továbbá ezekből hét statisztikai értéket képezünk.

A deriváltakat a matematikából jól ismert definíció alapján képezzük. Legyen f egy egyváltozós valós függvény és x_0 az értelmezési tartomány egy belső pontja [22]. Ebben az esetben f függvény x_0 -beli deriváltja alatt a következő határértéket értjük:

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

Nézzük, hogy egy akusztikai paraméternek mely hét statisztikai értékével dolgozunk:

- NORMMAX – maximum
- NORMMIN – minimum
- NORMMEAN – átlag
- NORMMEDIAN – medián
- NORMSKEW – ferdeség (aszimmetriát jelöl, nem-normális eloszlású elemeket)
- NORMRANGE – fekvés (maximum és minimum különbsége)

- NORMSTD – szórás

Ahogy előbbiekben írtam, az akusztikai paraméterek deriváltjai is adottak, ezek jelölése és jelentése:

- NORMDEVMAX – derivált maximuma
- NORMDEVMIN – derivált minimuma
- NORMDEVMEAN – derivált átlaga
- NORMDEVMEDIAN – derivált mediánja
- NORMDEVSKREW – derivált ferdesége
- NORMDEVVRANGE – derivált fekvése
- NORMDEVSTD – derivált szórása

A deriváltak értékei ugyancsak intonációs frázisokra, időegységenként (ablakonként) értendők.

A ferdeség pontosabban azt fogalmazza meg, hogy mennyire aszimmetrikus, vagy nem szimmetrikus egy valószínűségi változó eloszlása. Az m várható értékű X valószínűségi változó ferdesége az

$$\frac{E[(X-m)^3]}{\left(E[(X-m)^2]\right)^{\frac{3}{2}}}$$

képlettel adható meg, ahol $E[\]$ a várható értéket jelöli. Mondhatjuk, hogy a ferdeség a harmadik centrális momentum és a szórás köbének a hányadosa [21].

Az alaphang, az energia, a tizenkét darab MFC koefficiens, a HNR, valamint ezeknek deriváltjai által összesen harminc paraméterünk adott. Ezeken fejenként elvégezzük a fent említett hét statisztikai származtatást, akkor $((15) \times 2) \times 7 = 210$ akusztikai paraméter.

Az akusztikai jellemzők értékeit összeállítva egy olyan mátrixot kapunk, amelynek egy sora egy hangmintához tartozó jellemzővektor, az oszlopai pedig az

akusztikai jellemzők. A mátrixból egy X_{ij} mező tehát megadja i . hangmintához tartozó j . akusztikai jellemzőt.

Az mátrixot tartalmazó fájl soronkénti első karaktere (0,1,2,3,4) arra utal, hogy melyik alapérzelemnek a mérési adatairól van szó.

0: düh/ideges

1: vidám/öröm

2: beszéd közben nevetés

3: semleges

4: szomorú/bánatos

A főkomponens elemzést először az öt külön érzelemre, majd az összes adatunkon egyszerre végzem el. A rendelkezésemre álló adatbázisban az érzelmeken belül az intonációs frázisok száma eltérő (1. táblázat).

<i>Osztálynév</i>	<i>Frázisok száma</i>
Dühös/Ideges (A - Angry)	304
Vidám/Öröm (J - Joy)	34
Beszéd közbeni nevetés	63
Semleges (N - Neutral)	647
Szomorú/Bánatos (S - Sad)	62

1. táblázat - A kiadott érzelembázis osztályai és a kategóriák mérete

5.2. A főkomponens elemzés Matlab-beli előkészítése

Matlab – ban ha be akarjuk olvasni egy fájl tartalmát, akkor azt szöveggént kezeljük. Emiatt minden egyes sort *float*, azaz valós típusra kell konvertálni. A

beolvasáshoz használt *fgets* függvény soronként olvas, ezért azt is meg kell oldani, hogy folyamatosan olvassa be egymás alá az adatokat. Ezt úgy lehet elérni, hogy először definiálunk egy üres mátrixot, majd beolvasási soronként alámásoljuk az értékeket. A beolvasást elvégző kódrészlet:

```
file = 'G:/0.txt'

fh=fopen(file, 'r');

A=fgets(fh) ;

data=[];

while ischar(A)

    disp(A);

    A=sscanf(A,'%f');

    data = [ data ; A];

    A=fgets(fh);

end
```

Mivel az érzelem sorszáma nem tartozik a kiértékelni kívánt adathalmazunkba így azt eltávolítjuk:

```
temp = size(data);

data = data(:,2:temp(2));

data=data';

data
```

A beolvasás után az értékeket normalizálni kell egy általunk meghatározott skálázási intervallumra. Erre azért van szükség, mert a PCA bemenetén, az adatmátrixban, azonos súlyú értékeket feltételez, különben az eltérő nagyságrendű adatok eltérő súllyal járulnak hozzá az eredményhez. Ezt úgy tehetjük meg, hogy a bemenő adatunk minden oszlopában megkeressük a legnagyobb értéket, majd az oszlop minden tagját elosztjuk ezzel a legnagyobb értékkel, majd megszorozzuk az általunk meghatározott skálázási értékünkkel. Mivel a legnagyobb kiterjedés negatív is lehet, így az oszlop maximumának az abszolút értékét képezzük.

$$\frac{oszlop(i)}{\max(oszlop)} \times skálázás$$

Ha ezt elvégezzük az egész mátrixra egy módosított és normalizált mátrixot kapunk.

```
Xmax=10;

i=1;

[rowData,columbData]=size(data);

while(i<=columbData)

    columbMax=max(abs(data(:,i)));

    data(:,i)=data(:,i)/columbMax*Xmax;

    i=i+1;

end
```

A teljes főkomponens elemzést megvalósító kódrészlet a függelékben található.

A kiértékeléshez megvizsgáltam, hogy az elemzés az egyes paraméterek esetén milyen értékeket ad. Ehhez legelőször a *PC* mátrix elemeit rendezzük csökkenő sorrendbe és egy mátrixban szemléltetjük, az indexével együtt:

```
[value, index] = sort(abs(PC(110,:)), 'descend')

VI=[value; index]; %index value matrix
```

Végül szükséges tudni nem csak a paraméterek sorszámát, de nevét is. Ezt az akusztikai mennyiségeinket leíró fájlunkkal tudjuk elvégezni.

```

file = 'G:/forPCA.set_vector_leiras.txt' ;
fd=fopen(file, 'r');
B=fgets(fd) ;
parameter=[];
while ischar(B)
    disp(B);
    B=sscanf(B,'%16c');
    parameter = [parameter ; cellstr(B) ];
    B=fgets(fd);
end

```

Ahhoz, hogy az eredményeimet jobban szemléltethessem, százalékos arányban jelenítem meg a főkomponensek súlyát. Az ezt megvalósító kódrészlet pedig:

```

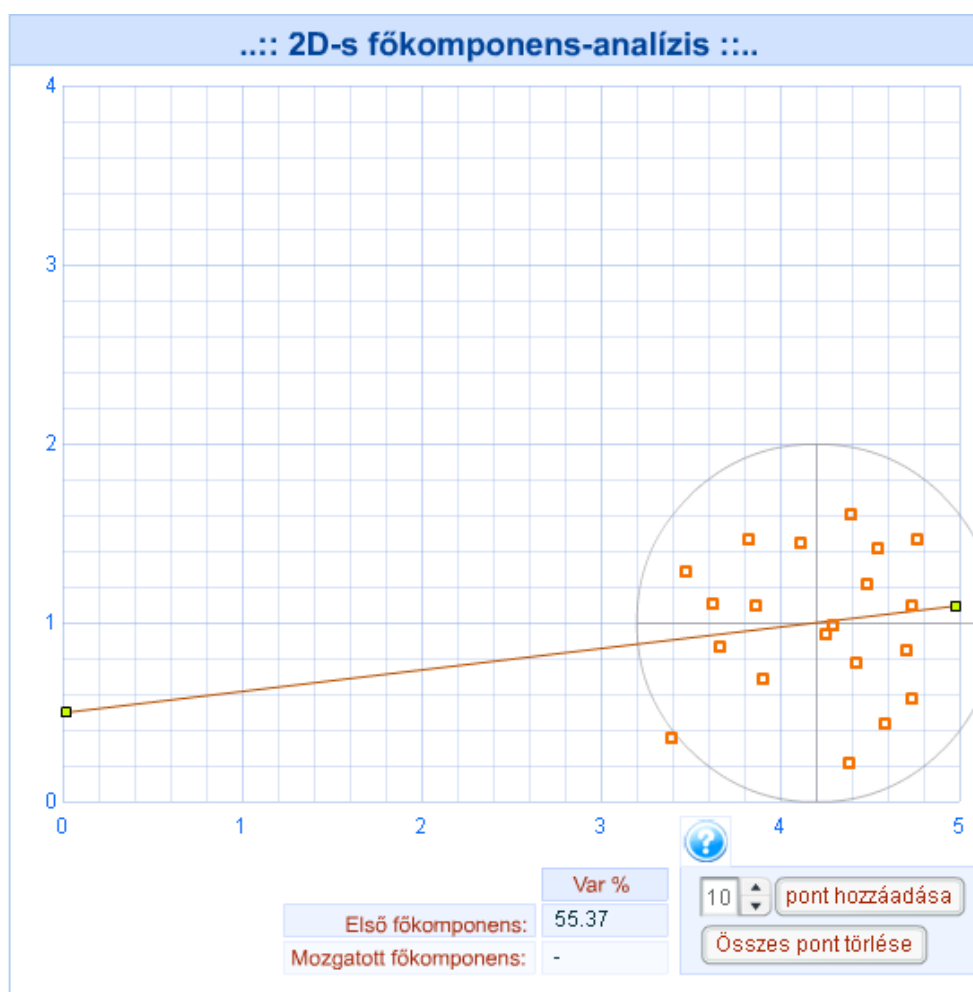
for i=1:size(V)
    V2(i,:)=V(i)/sum(V)*100;
end
temp=size(VI);
for i=1:temp(2)
    VI2(:,i)=VI(1,i)/sum(VI(1,:))*100;
end

```

5.3. Az eredmények ismertetése

A kiértékelésnél több szempontot kell figyelembe venni. Elsősorban szem előtt kell tartani, hogy az adathalmaz többdimenziós. A mátrix nagyságát a frázisok mintaszáma és a 210 akusztikai paraméter határozza meg. Ez merőben más, mint a szemléltetésben használt kétdimenziós rugómozgás. Kétdimenziós esetben a főkomponens egy egyenes illesztésével határozható meg, ez többdimenziós esetben egy hipersík illesztése lehet.

Képzeljük el a következő gondolati kísérletet: tegyük fel, hogy kétdimenziós eset áll előttünk, a vízszintes tengely legyen az alaphang értéke, a függőleges tengely az energiaszint. Egy olyan érzelmek adatfelhője rajzolódik ki a koordináta rendszerünkben, amelyre magas alaphang és alacsony energiaszint a jellemző. Az 5. ábra mutatja, hogy a főkomponensünk mely egyenes illesztésével határozható meg.



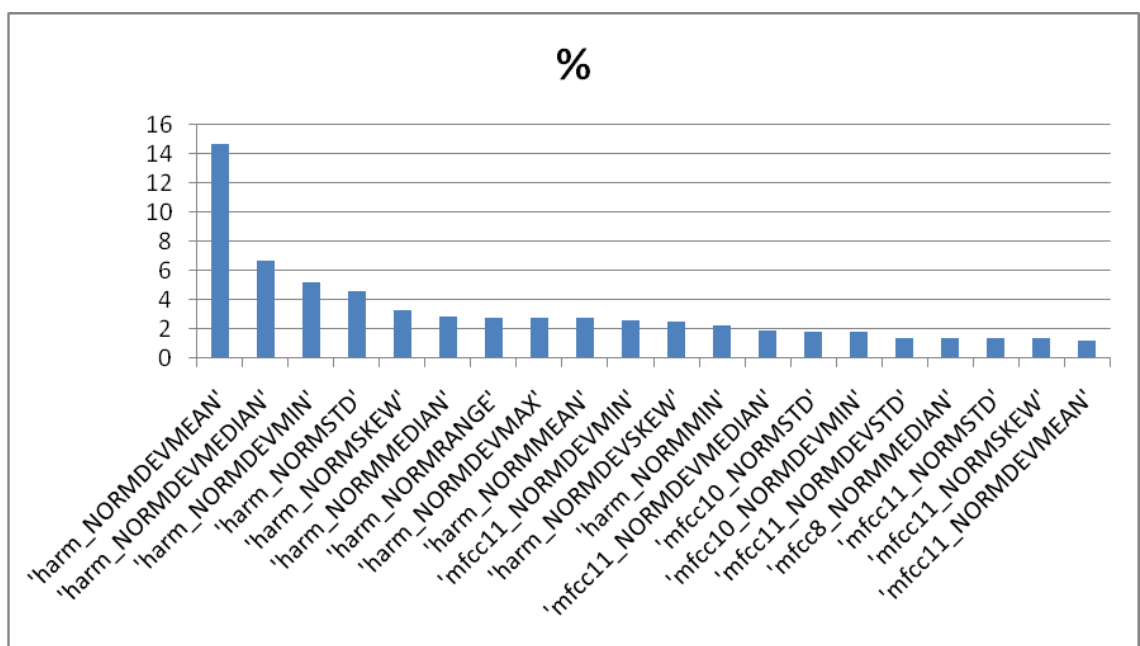
5. ábra – Kétdimenziós gondolati kísérlet: egy érzelmek magas alaphang és alacsony energiaszint jellemez [20]

Viszont gyakorlati megvalósításban sokkal célszerűbb lenne egy egyenes helyett egy körrel meghatározni ezt, ez az ábrán is jól látszik. A 3.4.5. fejezetben leírtam a főkomponens elemzésre vonatkoztatott kikötéseket és az első kikötés a linearitás volt. Tehát az eredmény használhatóságának feltétele, hogy az adathalmaz tartalmazott struktúrája lineárisan elemezhető legyen.

A következőkben a kapott főkomponenseim elemzését mutatom be az adott érzelmek esetén. Az analízis minden érzelemnél több főkomponenst eredményez. Minden komponens a hozzá tartozó súlyérték szerint járul hozzá az adathalmaz teljes variációjának leírásához. Bemutatom a legfontosabb főkomponenseket, valamint az azokban szereplő akusztikai jellemzők variációjának mintázatát, amennyiben ilyen létezik. Felsorolom a legnagyobb súllyal, tehát a legnagyobb variációval rendelkező akusztikai paramétereket, valamint, ha meghatározható, akkor a legkisebb súlyúakat, amelyek értékei a legállandóbbak egy adott osztály során.

5.3.1 Düh

A főkomponens elemeket tartalmazó PC mátrix legnagyobb súlyú elemeiből a legnagyobb értékeket mutatja be a 6. ábra, százalékos arányban.

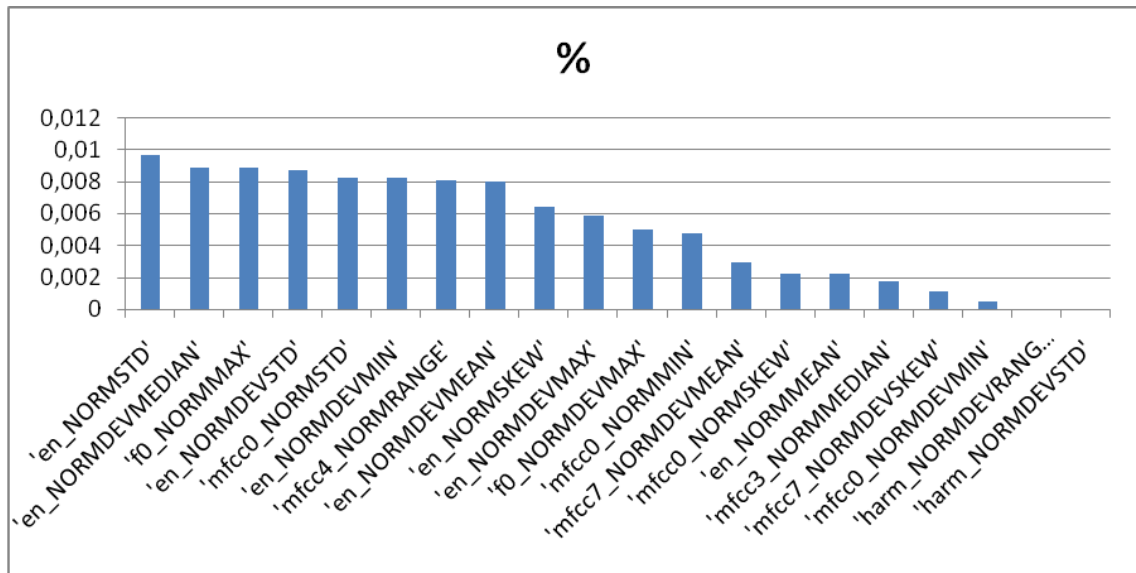


6. ábra – A düh alapérzelem első húsz akusztikai jellemzője százalékos ábrázolásban

Az első húsz értékről készítettem diagramot csökkenő sorrendben, a további eredmények kis mértékben térnek el egymástól. Az értékek exponenciálisan csökkennek. Észrevehető, hogy a tizennégy HNR-ből származtatott akusztikai paraméter közül tizenegy benne van az első húszban, ezek közül is a HNR deriváltjának átlaga az, amely legjellemzőbb az adathalmazra. 14,7% - ban

illeszthető olyan hipersík, amely mentén legnagyobb a variancia. Tehát a düh érzelmen belül a 'harm_NORMDEVMEAN' a legváltozékonyabb.

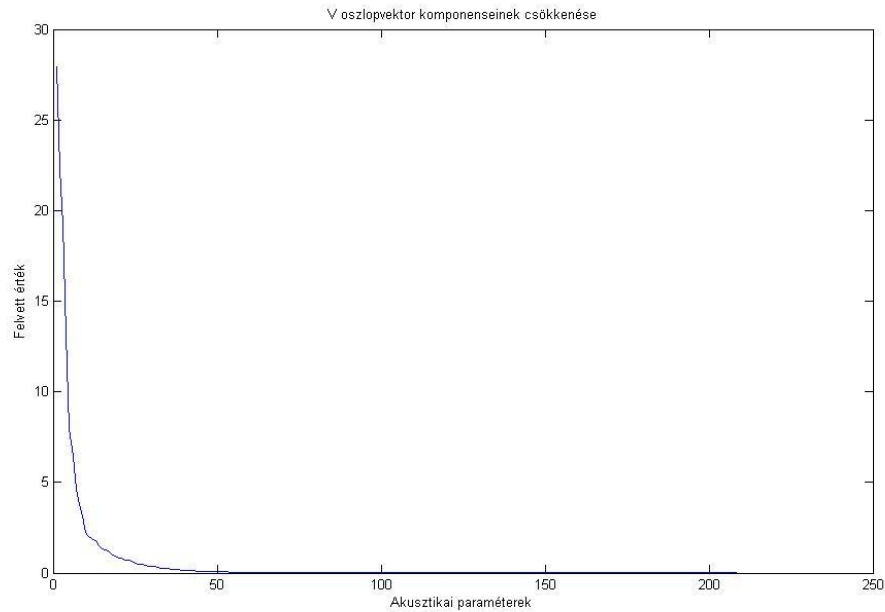
A 7. ábra a düh érzelem utolsó húsz akusztikai paraméterét mutatja be, ugyancsak százalékos arányban.



7. ábra - A düh alapérzelem utolsó húsz akusztikai jellemzője százalékos ábrázolásban

Fontos megemlíteni, hogy bár a végén az értékek tényleg nagyon alacsonyak, mégsem tekinthetők nullának. Érdekes módon a HNR két származtatott értéke helyezkedik el az utolsó két helyen, ezek változékonysága ezen érzelem mentén a legkisebb.

Mivel az érzelemben van mintázat (kirívóan nagy értékű akusztikai paraméter százalékos értéke) ezért érdemes a komponensek csökkenését illusztrálni. A 8. ábra a kódban szereplő *V* oszlopvektor értékeit tartalmazza (nem százalékos alakban). *V* a kovariancia mátrix sajátvektorait tartalmazza diagonizálva és sorba rendezve, az ábrán látszik ezek egymáshoz képesti csökkenése. A legnagyobb érték 27,9. A törés az *x* tengely mentén körülbelül 10-nél következik be, utána az értékek drasztikusan közelítenek a nullához. A *x* tengely menti paramétereket ki lehet olvasni a Matlab kód *kimenet* változójából, az *y* tengely menti értékeket pedig *V*-ből.



8. ábra - „V” oszlopvektor komponenseinek csökkenése dűh alapérzelem esetén

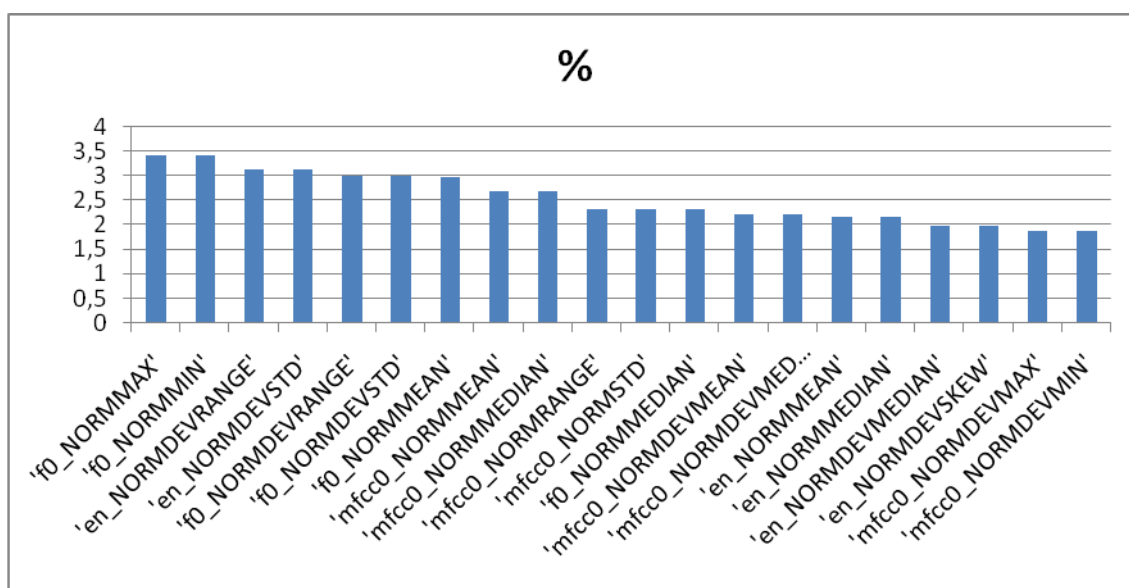
Az eredmény a legnagyobb súlyú főkomponens analízisre vonatkozik, amely a legnagyobb varianciájú irány mentén illeszt hipersíkot és ebben a transzformált térben keresi a rejtett mintákat. Mivel a nagyságszerinti varianciákat sorba rendeztem, módszert kaptunk a főkomponensek fontosságának megítélésére. A főkomponenseket tartalmazó (*PC*) mátrix 210×210 -es nagyságú, a fenti eredmény a kétszázötödik sorára vonatkozik. Minden irányhoz tartozó variancia meghatározza mennyire meghatározók az egyes irányok, azáltal, hogy nagyságrendi sorrendbe rendezzük a bázisvektorokat a hozzájuk tartozó varianciák szerint. Egy főkomponens súlya megegyezik a kódban szereplő *PC* mátrix sorának sorszámaival. Kisebb súlyú (vagy többrendű) főkomponensek esetében a mintázat kialakulása nem egyértelmű. Az értékek jóval alacsonyabbak és többnyire színképi információk dominálnak. A kétszázhetedik súlyú főkomponens (ahol a sorszám a súly szerint sorba rendezett fontosság) amely még elég nagy súllyal szerepel, ugyancsak, a legnagyobb súlyú főkomponensre emlékeztető mintázatot mutat (9. ábra).

Mint az ábrán látszik, a 'harm_NORMDEVSKEW' akusztikai paraméter kimagaslik a többi közül.

További többrendű főkomponensben nem találtam mintázatot, bár egyikét paraméter nagyobb, mint a másik, az értékek közel egyformák, az összvarianciához annyira kevésbé járulnak hozzá, hogy nem érdemes őket elemezni.

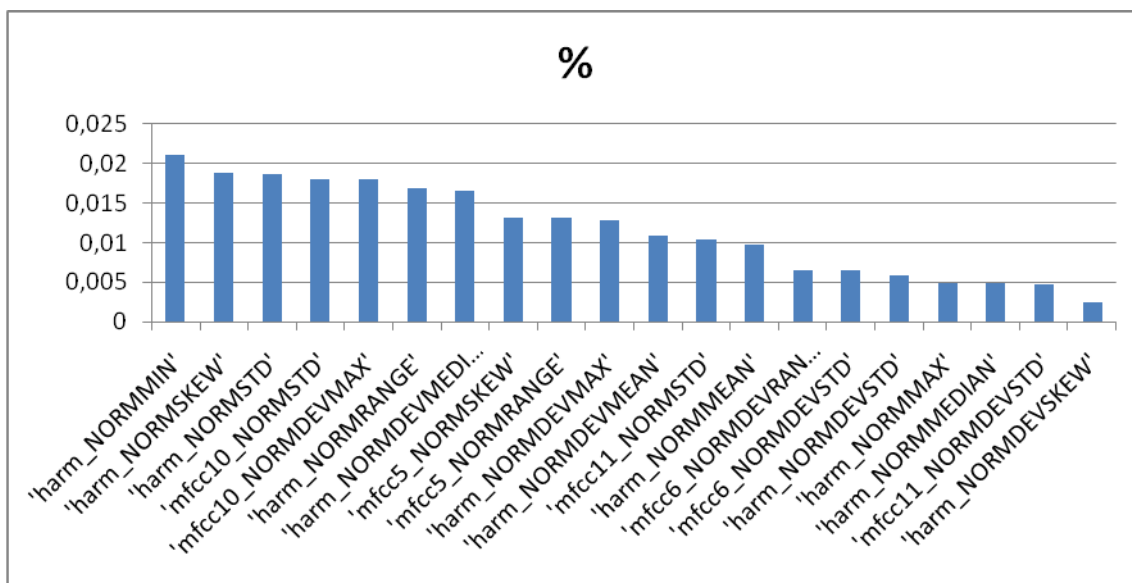
5.3.2 Öröm

A 11. ábra mutatja a főkomponensek százalékos súlyát az öröm érzelmre. Itt nincs jellegzetes minta, azaz nincs kirívó nagyságú elem. Megfigyelhető, hogy az energiából származtatott akusztikai paraméterek itt szembetűnőek, hat helyen találhatóak meg, viszont az alaphang mennyiségei is szép számban megjelennek. Úgyszintén hat alaphangból származtatott statisztikai paraméter foglal helyet az első húsz legnagyobb értékből.



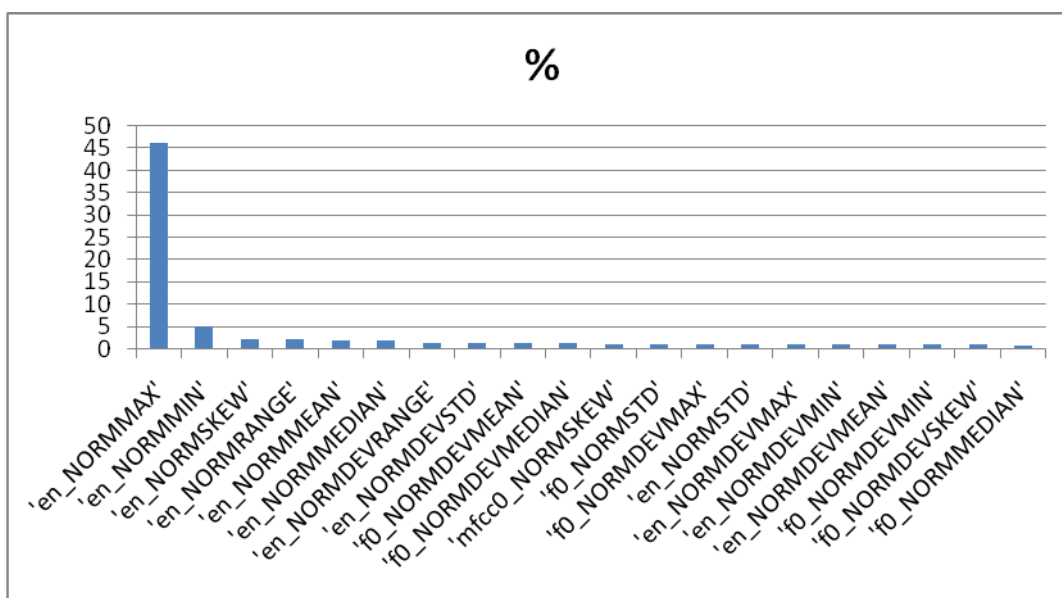
11. ábra – Az öröm érzelm első húsz akusztikai jellemzője százalékos arányban

A 12. ábrán pedig az alsó húsz akusztikai jellemző van ábrázolva. A paraméterek megjelenéséről nem lehet megállapítani semmi szabályszerűséget.



12. ábra – Az öröm érzelem utolsó húsz akusztikai jellemzője százalékos arányban

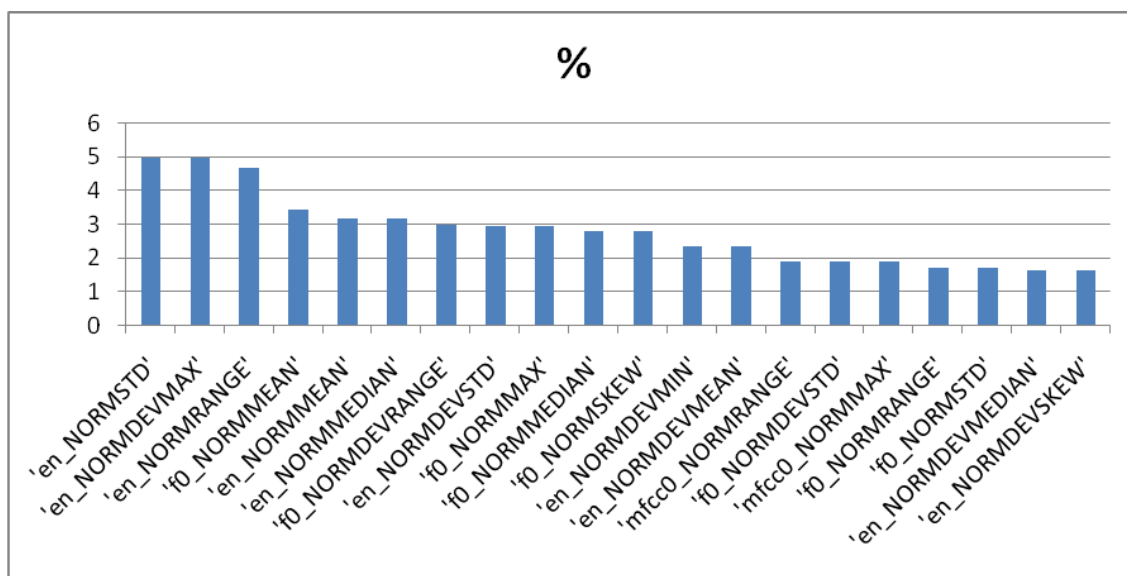
A többedrendű főkomponensek vizsgálatánál igazán meglepő eredmény született. Egyre kisebb sorszámú főkomponenseket nézve az akusztikai paraméterek százalékos értéke maximum fél százalékban tértek el egymástól, viszont a kétszázhatos főkomponens tartalmaz egy igazán kirívó elemet (13. ábra)



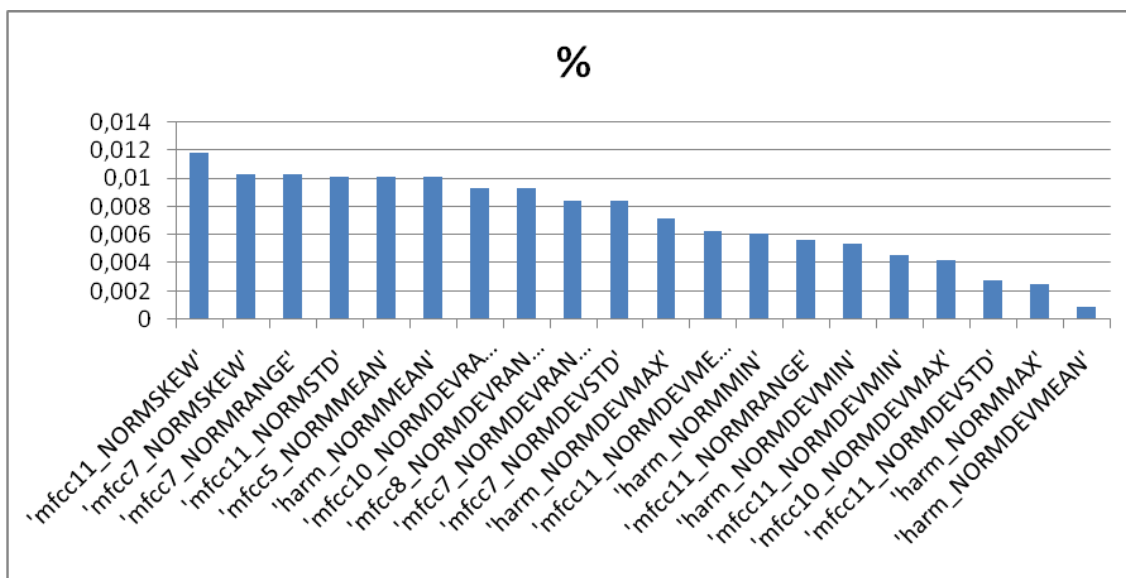
13. ábra - Kétszázhatodik súlyú főkomponens akusztikai paramétereinek százalékos eloszlása öröm érzelem esetén

További többedrendű főkomponens nem mutatott mintát.

Következő ábrán (14. ábra) sem találunk olyan akusztikai paramétert, amely valamilyen mintázatra utalna. Közel egyforma az értékek súlyozása. Tíz darab energiából származtatott mennyiség található és hét darab alaphangbéli.

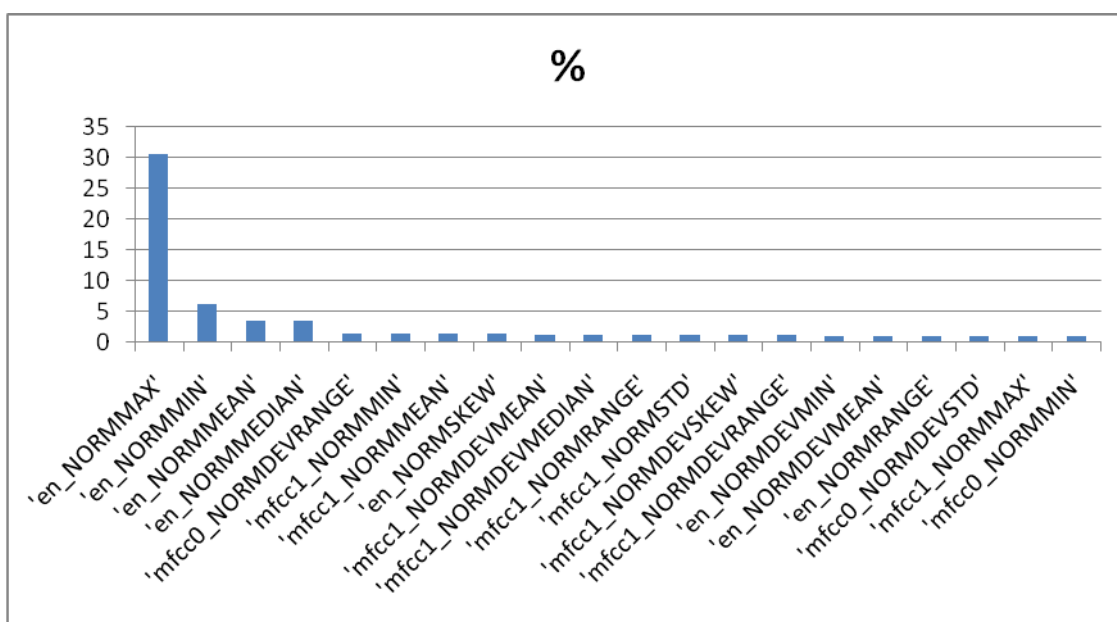


Az alsó szektorban ugyancsak HNR és MFC koefficienseket találunk (15. ábra).



15. ábra - Nevetve beszél érzelem utolsó húsz akusztikai jellemzője százalékos arányban

További főkomponenseket vizsgálva hasonló eredményre jutottam, mint az öröm érzelménél. A kétszázhatodik sorszámu többrendű főkomponens kirívó mintát mutat, ez a 16. ábrán látható.



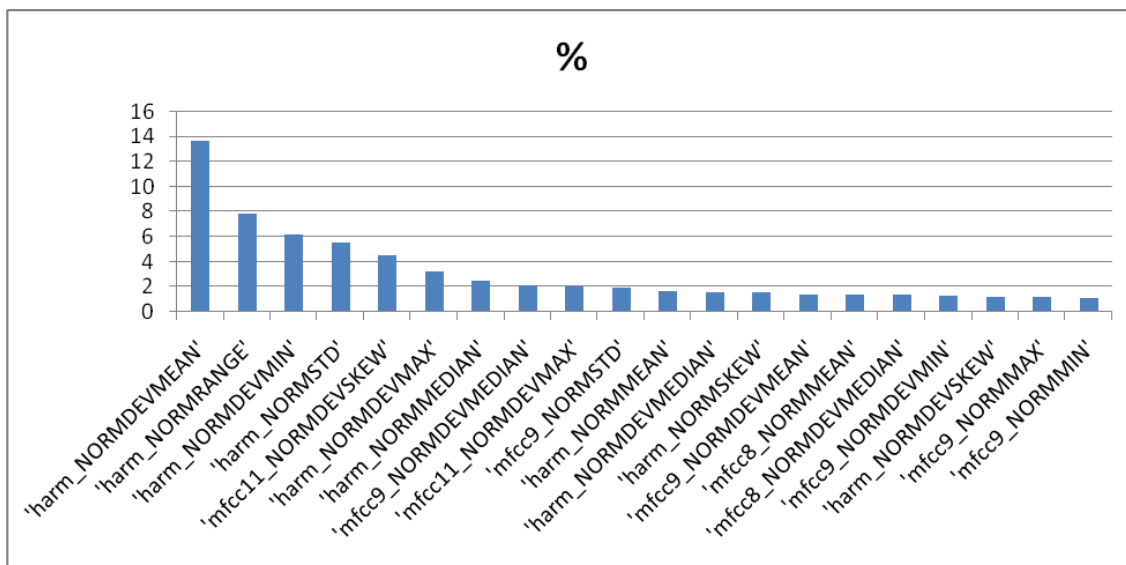
16. ábra - Kétszázhatodik súlyú főkomponens akusztikai paramétereinek százalékos eloszlása nevetve beszél érzelem esetén

Ebben az esetben is az energia maximának van a legnagyobb értéke 30,5%-os értékkel.

További esetekben az értékek közel megegyezők voltak, a színeképi származtatott paraméterek domináltak mintázatot nem mutatva.

5.3.4 Semleges

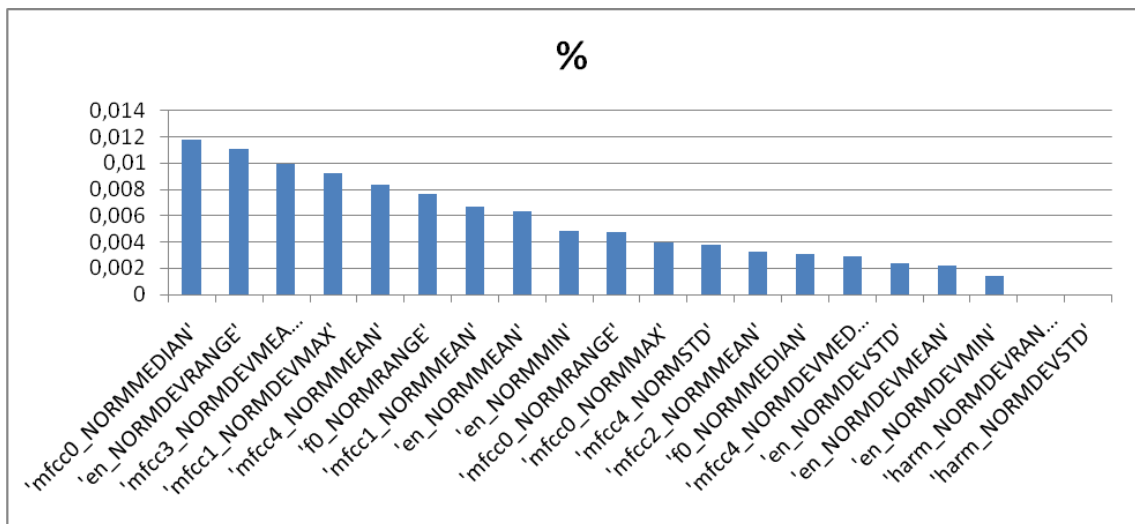
Ebből az emócióból állt rendelkezésemre a legtöbb intonációs frázis. A százalékos eloszlásban, a düh érzelemhez hasonlóan, találunk mintázatot, azaz van kirívó nagyságú komponens (17. ábra).



17. ábra - A semleges alapérzelem első húsz akusztikai jellemzője százalékos arányban

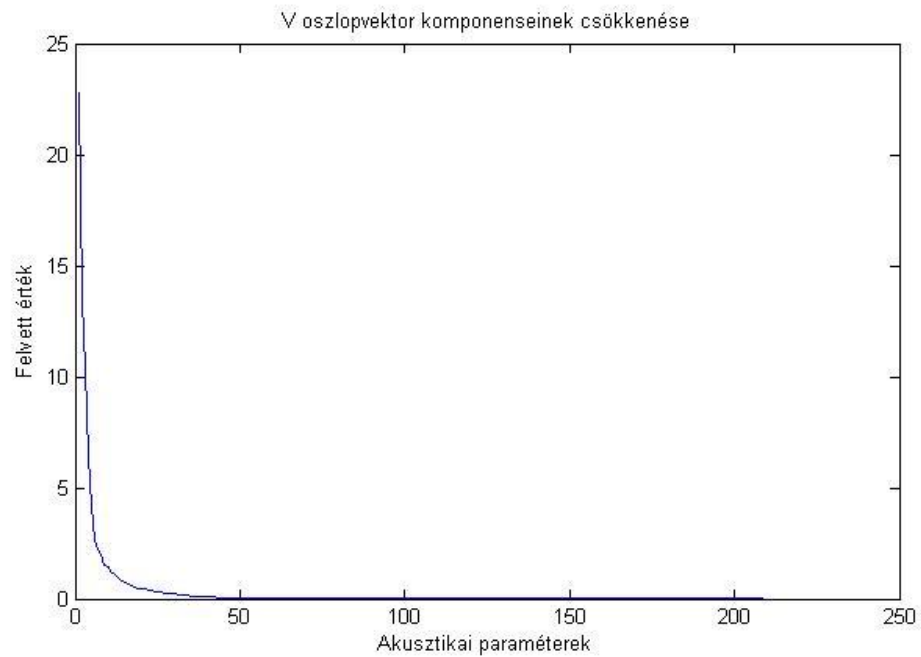
A dühös érzelemhez hasonlóan a HNR derivált átlagán van a legnagyobb százalékos súlya. Közel 14%-ban (13,7) illeszthető hipsík a legnagyobb varianciájú irányra. Ez a paraméter veszi fel az értékét a legnagyobb határokon belül. További három HNR tag szerepel közel hat és nyolc százalékos súllyal, majd exponenciálisan csökkennek.

Az utolsó két pozíciót ugyancsak két HNR komponens foglalja el (18. ábra). Mivel ezek közel nullával egyenlők, akár a hipsík menti sűrűsödésként is felfoghatók.



18. ábra - A semleges alapérzelem utolsó húsz akusztikai jellemzője százalékos arányban

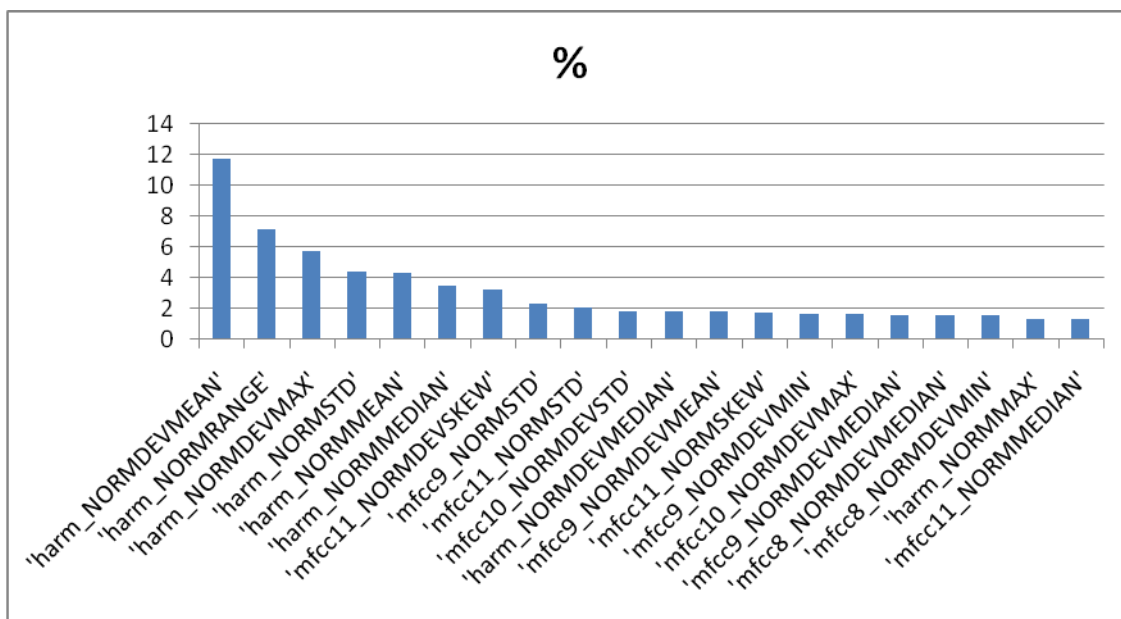
Ugyancsak érdemes kirajzoltatni a főkomponensek törésének grafikonját (19. ábra). A V oszlopvektor értékei ismételten nem százalékos arányban szerepelnek, és a paramétereket kiolvashatjuk a dühös érzelemnél említett Matlab paraméterek segítségével. A törés 10 előtt elkezdődik.



19. ábra - „V” oszlopvektor komponenseinek csökkenése semleges alapérzelem esetén

Így érdemes vizsgálni ebben az esetben is a többedrendű főkomponenseket további mintázat reményében.

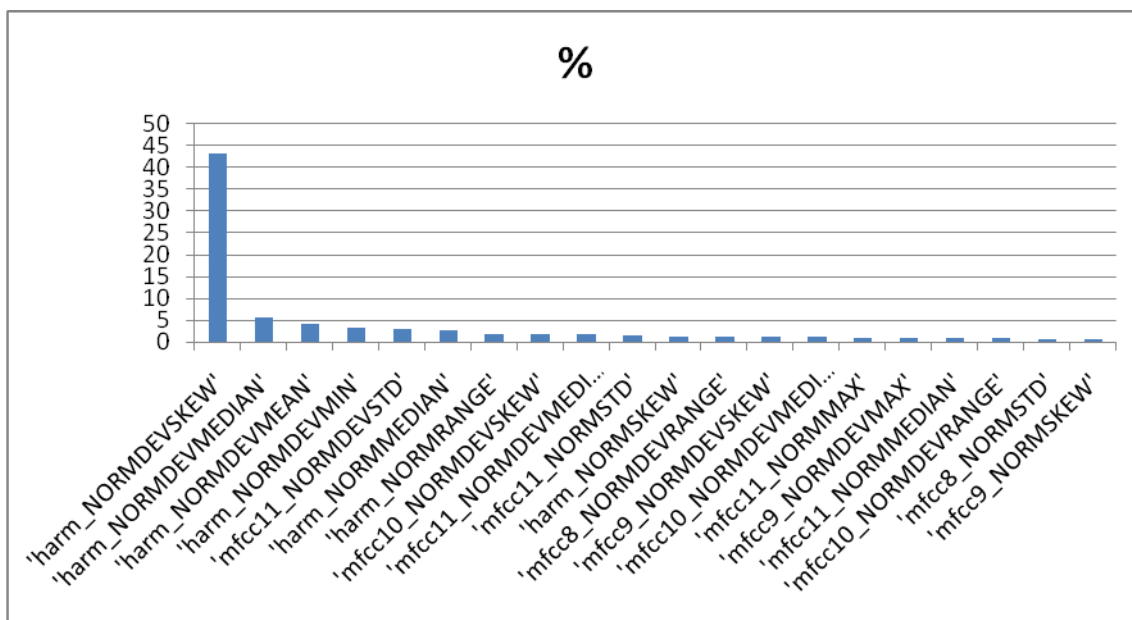
Kezdetben, ebben az esetben is az MFC származtatott értékei a dominánsak, kis eltérésekkel, majd a kétszázhetedik főkomponenshez érve újra mintázat rajzolódik ki (20. ábra).



20. ábra - Kétszázhetedik súlyú főkomponens akusztikai paramétereinek százalékos eloszlása semleges érzelem esetén

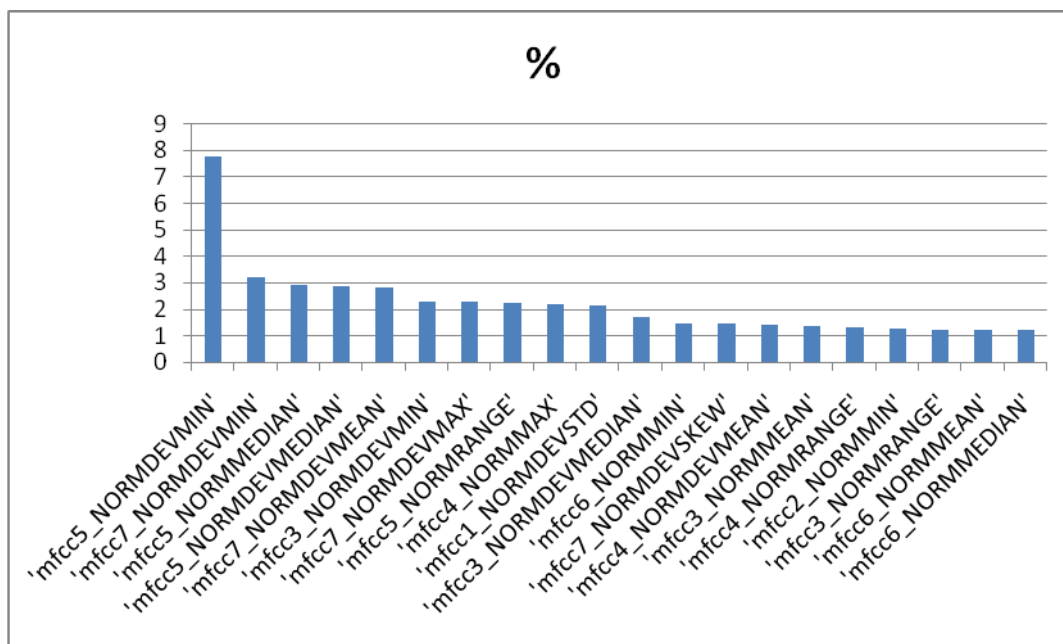
Mint a legnagyobb súlyú főkomponens esetében, itt is a HNR derivált átlaga veszi fel a legnagyobb százalékos értéket.

Egyel kisebb sorszámú főkomponensnél egy újabb HNR mennyiség kap főszerepet, ezúttal a derivált ferdesége, egy meglepő 43,1%-os súllyal (21. ábra)



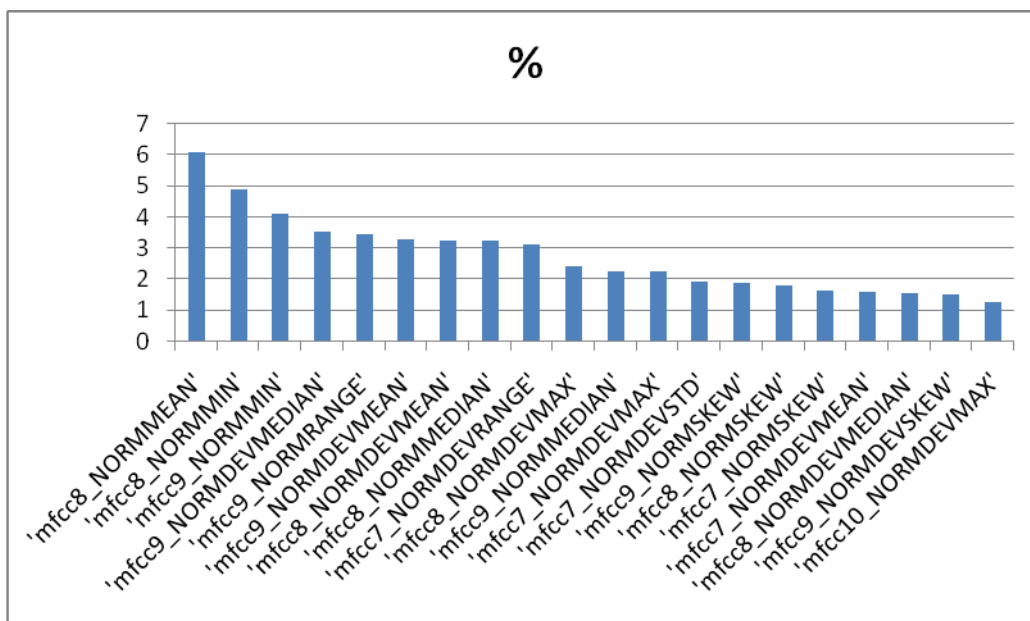
21. ábra - Kétszázhatodik súlyú főkomponens akusztikai paramétereinek százalékos eloszlása semleges érzelem esetén

A többi főkomponensnél az értékek közel egyformák voltak és a színeképi mennyiségek voltak többségben, egy esetben viszont egy színeképi paraméter kimagaslott (22. ábra).



22. ábra - Kétszáznegyedik súlyú főkomponens akusztikai paramétereinek százalékos eloszlása semleges érzelem esetén

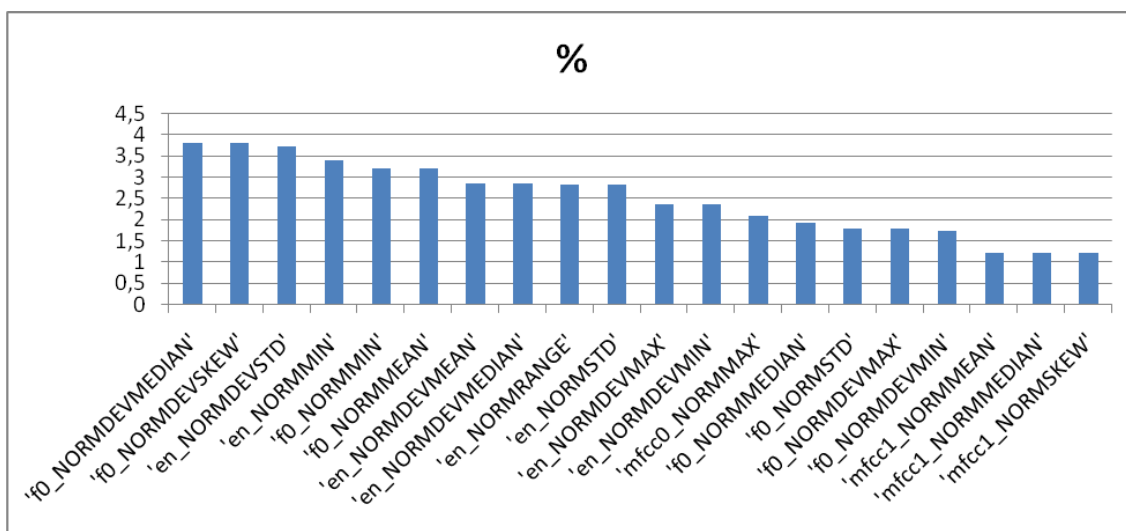
További nagyobb színeképi értékeket mutatott a kétszázharmadik többrendű főkomponens (23. ábra).



23. ábra - Kétszázharminc súlyú főkomponens akusztikai paramétereinek százalékos eloszlása semleges érzelem esetén

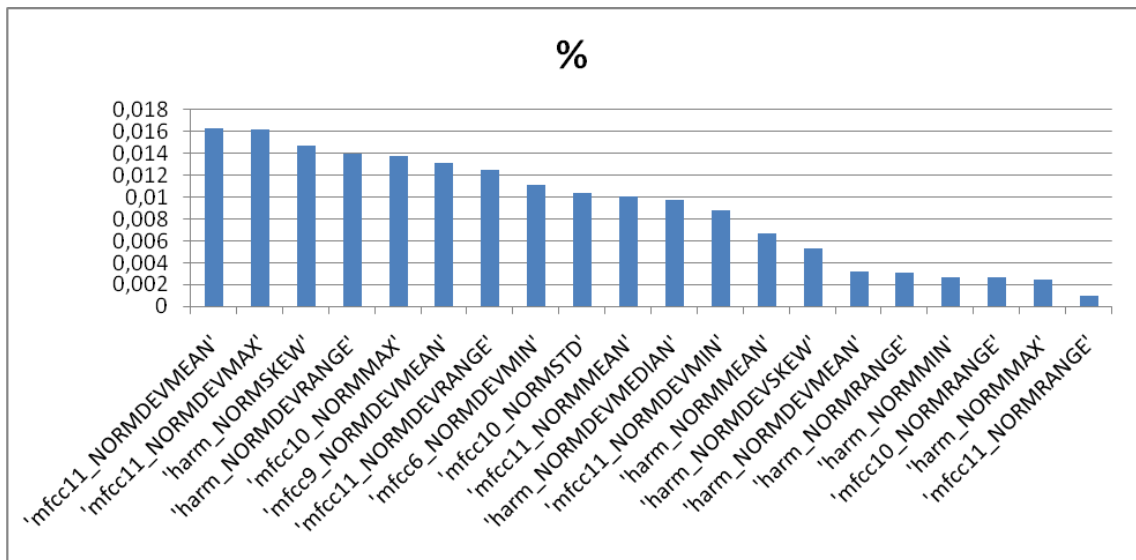
5.3.5 Szomorú

Az utolsó érzelem főkomponenseinek táblázatában nem lelhető fel mintázat (24. ábra). Az arányok közel egyenlők. Az alaphang származtatott értékei nagy számban fellelhetők (nyolc elemszámmal), az energia is szerepel nyolc származékával.



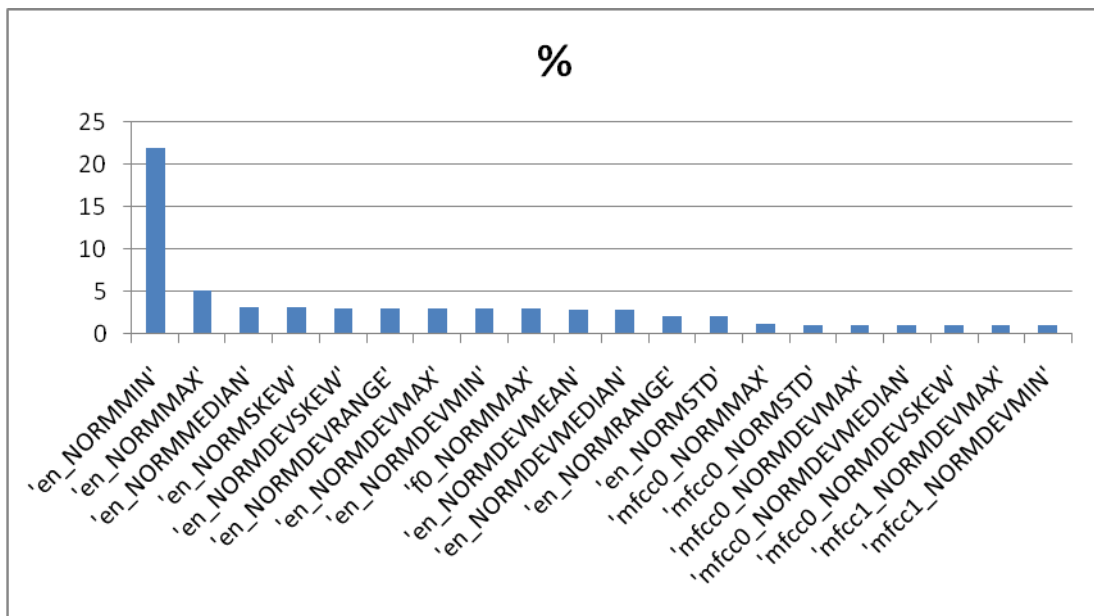
24. ábra - A szomorú alapérzelem első húsz akusztikai jellemzője százalékos arányban

Az utolsó húsz paraméter súlya a 25. ábrán található.



25. ábra - A szomorú alapérzelem utolsó húsz akusztikai jellemzője százalékos arányban

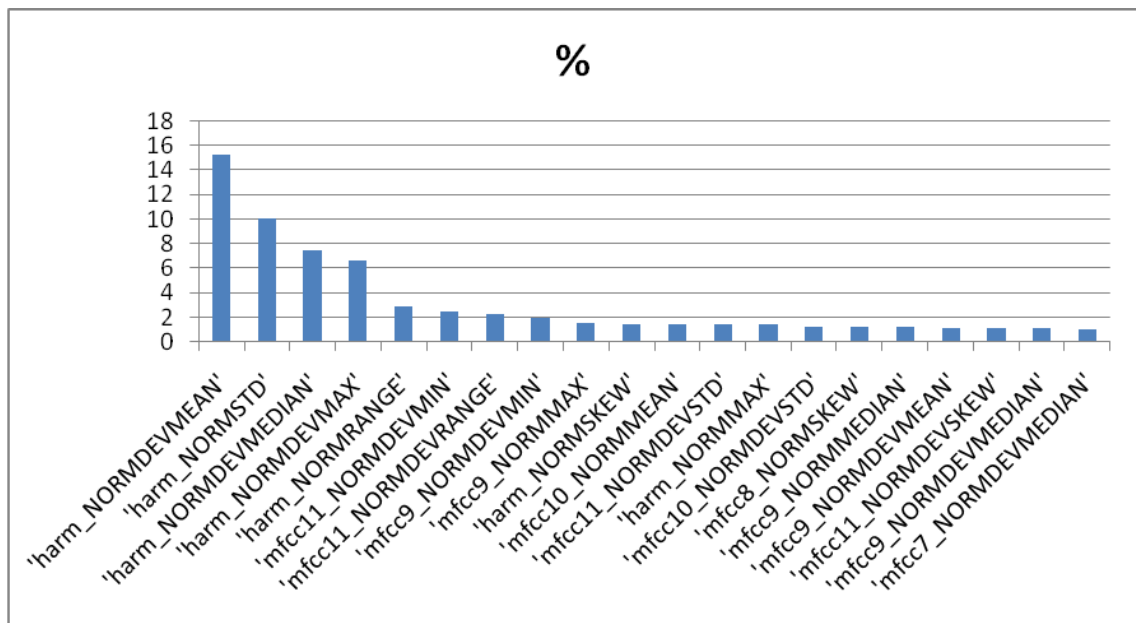
Többrendű főkomponensek közül csak egy van, ami az összvarianciához még elég nagy súllyal járul hozzá és ez a kétszázhatodik főkomponens (26. ábra).



26. ábra - Kétszázhatodik súlyú főkomponens akusztikai paramétereinek százalékos eloszlása szomorú érzelem esetén

5.3.6 Az összes érzelemre egyszerre végzett analízis

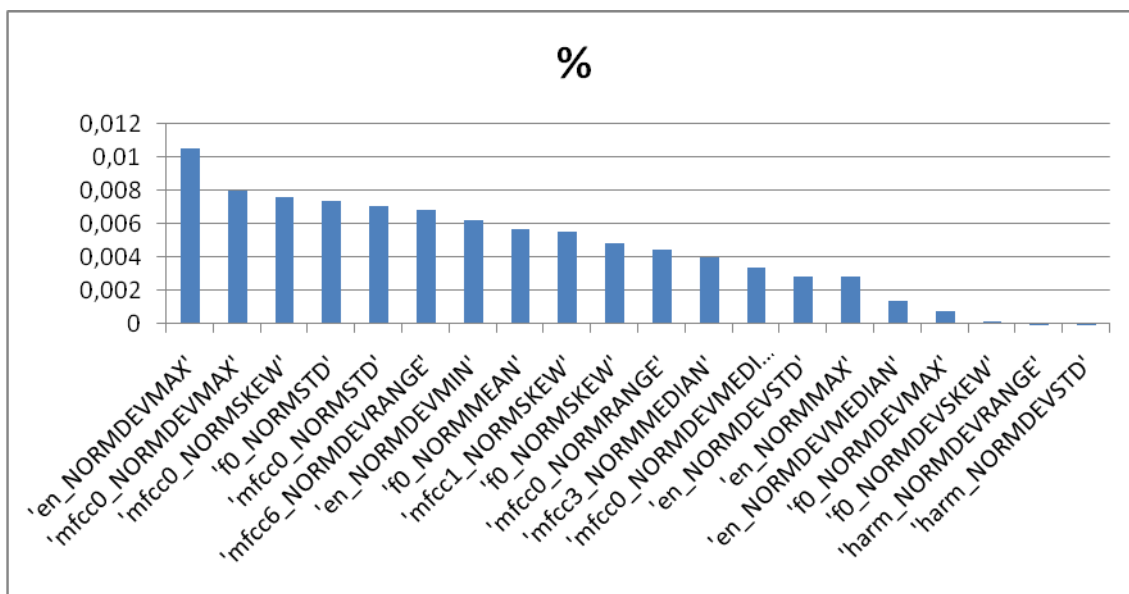
Ha az összes érzelmeket egyszerre vizsgáljuk, akkor az intonációs frázisok száma érzelmek szerint nem arányos. Tehát a képzeletbeli többdimenziós térünkben több „pontunk” lesz semleges emócióból és dühből, mint például öröm alapérzelemből. A kapott eredmény tehát adatbázisfüggő, ám ez az adatbázison végrehajtani kívánt osztályozási feladat szempontjától független. A PCA elemzés az osztályozás előtt arra szolgál, hogy a legnagyobb varianciájú, tehát az osztályozás szempontjából leglényegesebb dimenziókat válasszuk ki. A 27. ábra szemlélteti az eredményt.



27. ábra - Az összes érzelem első húsz akusztikai jellemzője százalékos aránya

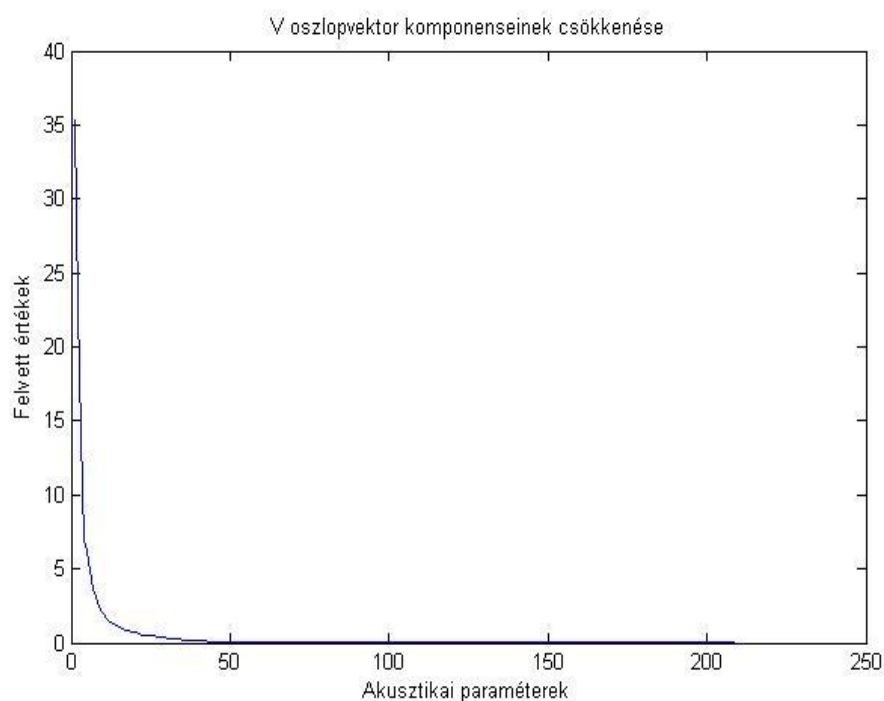
Ami a dühös és a semleges érzelem eredményeképp, 14% körüli mintázatoknak minősíthető (HNR deriváltjának az átlaga), az itt is kiemelkedik. Ebben az esetben a 15%-ot is eléri (15,3%). Ahogy az említett érzelmek eredményénél, további HNR értékek szerepelnek, bár nem ugyanolyan sorrendben és nem ugyanolyan aránnyal.

Az energia és alaphang komponensek csak az utolsó húsz főkomponensnél találhatók meg (28. ábra).



28. ábra Az összes érzelem utolsó húsz akusztikai jellemzője százalékos aránya

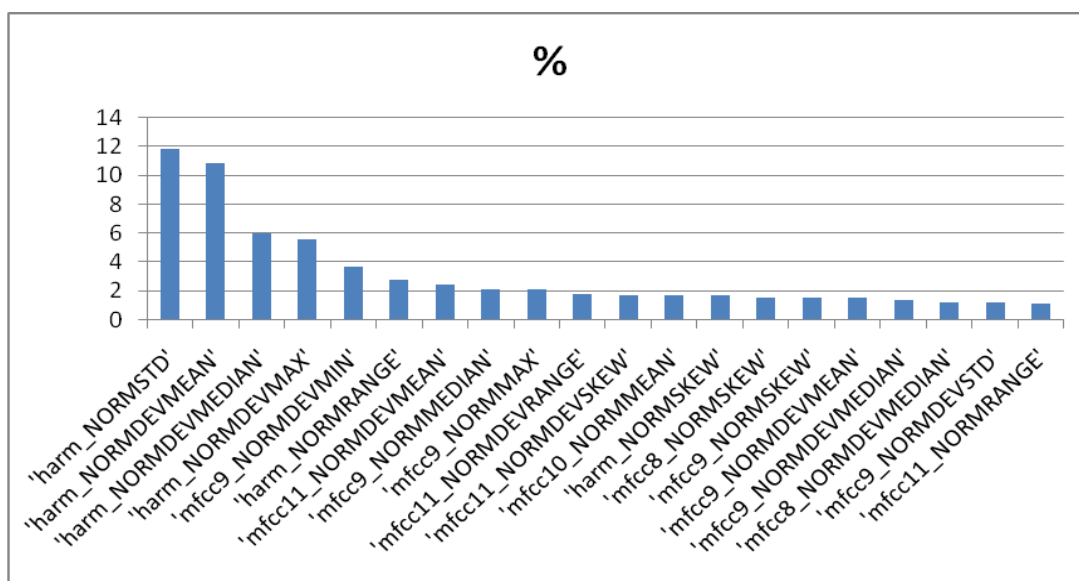
A 29. ábra mutatja az összes érzelemre számított V oszlopvektort és értékeinek csökkenését.



29. ábra – Az összes érzelemre vett „V” oszlopvektor komponenseinek csökkenése

A görbe külalakja hasonlít az előző kettőéhez, viszont az első főkomponens nagysága sokkal nagyobb, pontosabban 35,3- os értéket vesz fel. Mivel a töréspont közel ugyanott van, mint a düh és a semleges emóció esetében (10 körül), az esés meredekebb.

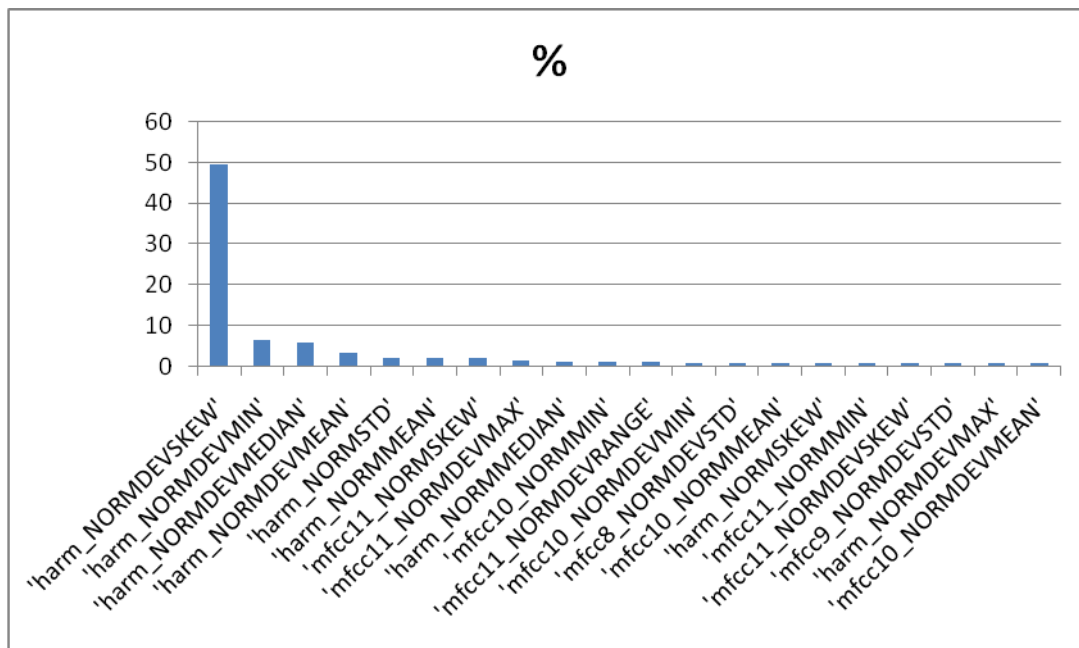
A többrendű főkomponensek esetében hasonló a helyzet, mint külön érzelmekre vett analízisnél, MFC származtatott paraméterei dominálnak, az értékek közel egyforma nagyságúak. A kétszázhetedik sorszámú főkomponens mintát mutat (30. ábra).



30. ábra - Kétszázhetedik súlyú főkomponens akusztikai paramétereinek százalékos eloszlása az összes érzelem esetén

Ismét a HNR paraméterek mutatnak kiemelkedő értéket. Ennek magyarázata lehet, hogy a düh és semleges érzelem esetében is jellemzőek voltak e paraméter magas értékei, különösen a 'harm_NORMDEVMEAN' és a 'harm_NORMDEVMSKEW' származtatott mennyiség. A düh és semleges érzelemből van a legtöbb intonációs frázis, így az összes érzelemnél is inkább ezek az értékek lehetnek nagyobb súlyúak.

A kétszázhatodik főkomponens tovább erősíti ezt a gondolatot (31. ábra).



31. ábra - Kétszázhatodik súlyú főkomponens akusztikai paramétereinek százalékos eloszlása az összes érzelem esetén

A HNR deriváltjának ferdesége 49,4%-os súlyt mutat. Bár kisebb súlyú főkomponens, mégis hasznos mintát képez.

5.4. Összefoglalás

Az általam elvégzett főkomponens elemzés várt és kevésbé várt kimenetet eredményezett. A mintázat kialakulása a legnagyobb varianciájú hipersík mentén egyáltalán nem meglepő, elsősorban ezt érdemes figyelembe venni további feladatok elvégzése esetén. A HNR -ből származtatott paraméterek mennyiségei nagy értéktartományok között mozoghatnak, ez a düh és semleges érzelemre érvényes, valamint az összes érzelmet, ha együtt nézzük.

További irányokat és mintázatokat vizsgálva meglepő eredmények születtek. Személy szerint azt vártam, hogy a legnagyobb súlyú főkomponens elemzés körülbelül definiálja az adott érzelemre legjobban jellemző akusztikai paramétereket, de ez csak részben valósult meg. A többedrendű főkomponensekben egyes színeképi információk is kirívóan magas súllyal szerepeltek. Itt a százalékos értékek is sokkal magasabbak voltak a mintázatok során, mint a legnagyobb varianciájú irányban. Volt eset, amikor több mint 40%-os volt egy adott származtatott mennyiség százalékos aránya.

Öröm és nevetve beszél érzelmekre az energiából származtatott akusztikai paraméterek formáltak mintázatot, bár nem a legnagyobb súlyú főkomponensben jelentek ezek meg. Ezt a két érzelmet közel ugyanazon akusztikai paraméterek jellemezték.

Szomorú érzelem esetén viszont csak a többedrendű főkomponensben rajzolódott ki mintázat (energia minimumából) és tudva azt, hogy ez kisebb súllyal szerepel, kérdéses, hogy további osztályozási feladatban mennyire használható.

Az összes érzelemre vett analízis során inkább a legnagyobb intonációs frázisszámú érzelmek (düh és semleges emóció) jellemzői jelentek meg, ami nem meglepő, viszont a többedrendű főkomponensben, közel 50%-os értéket elérő 'harm_NORMDEVSKEW' származtatott akusztikai paraméter igazán figyelemre méltó.

6. Felmerülő kérdések, további célok, egyéb tapasztalatok

Felmerülő kérdések:

A fenti gondolatmenetet követve felmerül a kérdés, hogy mennyire befolyásolja az eredményt az, hogy a rendelkezésre álló adatbázison belül egy érzelemből hány darab intonációs frázisunk van. Spontán beszédben a leggyakoribb a semleges érzelem, de kérdéses, hogy a fent kapott eredmény valóban hűen tükrözi-e a valóságot vagy inkább az adatbázisra ad-e leírást. Mondhatjuk-e, hogy az emberi érzelmek ezen akusztikai paraméterek mentén változatosak?

További célok:

Véleményem szerint érdemes lenne elvégezni a főkomponens analízist egy optimalizált érzelmi tartalmú hang adatbázison, amiben az érzelmek egyforma súllyal szerepelnek, intonációs frázisszámukat tekintve. Ez az eredmény viszont mindenképpen jó alapot szolgál az osztályozás javításában.

Ezen kívül, a főkomponens elemzés alkalmazható további nagy adathalmazok analízisére és lineáris strukturális szabályszerűségek feltárására is. Ilyen terület lehet például a depressziósok hangmintáját tartalmazó adatbázis.

Egyéb tapasztalatok:

Elsősorban meglepő volt számomra, hogy milyen nehéz volt irodalomhoz jutni a témakörben, ezért nehezen is indult be a munkám. Természetesen létezik magyar nyelvű szakiromány, viszont az nem a PCA gépi megvalósítására összpontosított, inkább egy haladó szintű elméleti leírás, amely számomra nem volt kielégítő. A megfelelő angol nyelvű irodalom birtokba vétele után tudtam effektíven haladni.

A főkomponens elemzés Matlab-os megvalósítására létezik egy másik módszer is, amely az SVD-t használja (Singular Value Decomposition), viszont az nem tartalmazza a kovariancia mátrix szemléletességét.

Az eredmények megleptek, személy szerint munkám kezdetekor arra tippeltem, hogy az alaphangból és energiából származtatott értékek lesznek a kirívó nagyságú akusztikai paraméterek, főleg a düh érzelemre. Tehát a kapott eredmények igazán izgalmasak.

Úgy érzem, hogy a félév során hasznos munkát végeztem. Sikerült elmélyülnöm az anyagban és megérintett a szépsége. Nem utolsó sorban, Matlab környezetben is otthonosabban mozgom. Remélem, hogy szakdolgozatom megfelelő alapot nyújt azoknak, akik főkomponens elemzésre szánják magukat. Szívesen foglalkoznék még a témával!

7. Köszönetnyilvánítás

Elsősorban köszönetet szeretnék mondani konzulensemnek, Dr. Vicsi Klára tanárnőnek, aki lehetőséget adott arra, hogy a beszédakusztika és érzelemfelismerés problémakörével foglalkozhassak. Figyelmembe ajánlotta a Beszédkommunikáció című tárgyat és készségesen megválaszolt minden kérdésemet, segített ahol tudott.

Köszönettel tartozom a Beszédakusztika Laborban dolgozóknak: Sztahó Dávidnak a sok türelméért, útbaigazításért, közös konzultálásért és tanácsaiért. Kiss Gábornak, aki az érzelem tartalmú hangadatbázis jellemzőit magyarázta el nekem.

Végül, de nem utolsó sorban, szeretném megköszönni barátaimnak, Damsa Andreinak, hogy az evolúciós pszichológia mai állásáról informált, Dávid Violettának és Varga Zsuzsának, hogy átolvasták a dolgozatomat és segítettek a formai és tartalmi hibák javításában.

Irodalomjegyzék

- [1] Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of the septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, 47, 419–427
- [2] Jonathon Shlens: A Tutorial on Principal Component Analysis, Center for Natural Science, New York University and Systems Neurobiology Laboratory, Salk Institute for Biological Studies; April 22, 2009; Version 3.01
- [3] Damsa Andrei, Damsa Dániel, Grünsperger Balázs, Hancz Helga Tímea: Érzelmek és motiváció evolúciós szemszögből, Erdélyi Tudományos Diákköri Konferencia, 2009 Kolozsvár

http://etdk.adatbank.transindex.ro/pdf/pszi_damsa_damsa_grunsperger_hancz.pdf
- [4] Darwin, Ch. (1963): Az ember és az állat érzelmeinek kifejezése. Budapest, Gondolat.
- [5] Freud Róbert, Lineáris algebra, ELTE Eötvös Kiadó, 1998.
- [6] <http://www.mathworks.com/>
- [7] Izard, E. (1977) Human emotions. New York, Plenum
- [8] Lang, P. J., Bradley, M. M., & Curthbert, B. N. (1990). Emotion, attention and the startle reflex. *Psychological Review*, Vol. 97
- [9] Tooby, J., & Cosmides, L. (1990). The past explains the present. Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, Vol. 11
- [10] Frijda, N. H. (1986). The Emotion. Cambridge University Press, New York
- [11] Oatley, K. és Jenkins, J. (2001). Érzelmek. Osiris kiadó, Budapest
- [12] Praat manual: Harmonicity

<http://www.fon.hum.uva.nl/praat/manual/Harmonicity.html>
- [13] Kiss Gábor: Érzelmeket kifejező beszédatbázis gyűjtése, és az érzelmeket kifejező akusztikai paraméterek vizsgálata, BME szakdolgozat, 2010 Bp.
- [14] Rainer Banse, Klaus R. Scherer: Acoustic Profiles in Vocal Emotion Expression *Journal of Personality and Social Psychology* 1996, Vol. 70, No. 3, p. 614-636
- [15] Tóth Sz. Sztahó D. Vicsi K.: Speech Emotion Perception by Human and Machine Proceeding of COST Action 2102 International Conference; Papers in Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction Springer

2007, (p. 213-224);
http://berber.tmit.bme.hu/final/docs/cikkek/Emotion_Patras.pdf

- [16] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, Shrikanth Narayanan: Analysis of Emotion Recognition using Facial Expressions; Speech and Multimodal Information International Conference on Multimodal Interfaces, Proceedings of the 6th international conference on Multimodal interfaces (2004), p. 205 – 211
- [17] Sztahó Dávid: Beszéd érzelmi tartalmát kifejező fizikai paraméterek vizsgálata és az érzelem gépi felismerése; BME diplomamunka 2008 Bp.
- [18] Ellen Douglas-Cowie, Nick Campbell, Roddy Cowie, Peter Roach: Emotional speech: Towards a new generation of databases; Speech Communication 40 (2003), p. 33–60; www.elsevier.com/locate/specom .
- [19] Vicsi Klára: Beszédkommunikáció
<http://alpha.tmit.bme.hu/speech/education.php>
- [20] <http://psycho.unideb.hu/statisztika/pages/interaktiv.html#ill2>
- [21] Fazekas I. (2000): Bevezetés a matematikai statisztikába. Kossuth Egyetemi Kiadó, Debrecen
- [22] Thomas, George B., Maurice D. Weir, Joel Hass, Frank R. Giordano. Thomas-féle Kalkulus I; Typotex: Budapest (2006)

Függelék

Lineáris Algebra

Ez a szakasz a lineáris algebra néhány tételét tartalmazza, amely elengedhetetlen a főkomponens elemzés matematikájának megértéséhez.[2]

1. Az ortogonális mátrix inverze egyben transzponáltja is.

Legyen A egy $m \times n$ – es ortogonális mátrix ahol a_i az i -edik oszlopvektor. Az $A^T A$ ij – edik eleme

$$(A^T A)_{ij} = a_i^T a_j = \begin{cases} 1 & \text{ha } i = j \\ 0 & \text{egyébként} \end{cases}$$

Tehát, mivel $A^T A = I$, ebből következik, hogy $A^{-1} = A^T$.

2. Minden A mátrixra igaz, hogy $A^T A$ és AA^T szimmetrikus.

$$(AA^T)^T = A^{TT} A^T = AA^T$$

$$(A^T A)^T = A^T A^{TT} = A^T A$$

3. Egy mátrix szimmetrikus akkor és csak akkor, ha ortogonálisan diagonalizálható.

Mivel ez a tétel kétirányú, az is igaz, hogy ha egy mátrix ortogonálisan diagonalizálható, akkor szimmetrikus, kétirányú „akkor és csak akkor” jellegű bizonyítást igényel.

Előbb nézzük az egyenes irányt: ha A ortogonálisan diagonalizálható, akkor A szimmetrikus mátrix. Hipotézis alapján az ortogonálisan diagonalizálhatóság azt jelenti, hogy létezik E melyre igaz, hogy $A = EDE^T$, ahol D egy diagonális mátrix és E egy speciális mátrix mely diagonalizálja A -t. Számítsuk ki A^T -t.

$$A^T = (EDE^T)^T = E^{TT}D^TE^T = EDE^T = A$$

Tehát, ha A ortogonálisan diagonalizálható, szimmetrikusnak kell lennie.

A bizonyítás ellentétes iránya bonyolultabb és zavarosabb, ennek ismertetése meghaladja a dolgozat keretét.

4. Egy szimmetrikus mátrix az ortonormált sajátvektoraiból felépített mátrixával diagonalizálható.

Legyen A egy négyzetes $n \times n$ - es szimmetrikus mátrix, sajátvektoraival $\{e_1, e_2, \dots, e_n\}$. Legyen $E = [e_1 \ e_2 \ \dots \ e_n]$ ahol az E i -edik oszlopa e_i sajátvektora. Ez feltételezi azt, hogy létezik D diagonális mátrix, melyre igaz, hogy $A = EDE^T$.

A bizonyítás két részből áll. Először belátjuk, hogy bármely mátrix ortogonálisan diagonalizálható akkor és csak akkor, ha a mátrix sajátvektorai lineárisan függetlenek. Majd belátjuk, hogy a szimmetrikus mátrix rendelkezik egy speciális tulajdonsággal, mely szerint a sajátvektorok nem csak lineárisan függetlenek egymástól, de ortogonálisak is, ezzel befejezván a bizonyításunkat.

A bizonyítás első részében, legyen A egy tetszőleges mátrix, nem feltétlenül szimmetrikus, és legyenek független sajátvektorai. Továbbá, legyen $E = [e_1 \ e_2 \ \dots \ e_n]$ sajátvektorokból álló mátrix az oszlopokba helyezve. Legyen D egy diagonális mátrix, amelyben az i -edik sajátérték legyen ii -edik pozícióba helyezve.

Most ezekkel a kikötésekkel megmutatjuk, hogy $AE = ED$. Vizsgáljuk meg az egyenlet bal és jobb oldalát.

$$\text{Bal oldala: } AE = [Ae_1 \ Ae_2 \ \dots \ Ae_n]$$

$$\text{Jobb oldala: } ED = [\lambda_1 e_1 \ \lambda_2 e_2 \ \dots \ \lambda_n e_n]$$

Tehát, ha $AE = ED$, akkor $Ae_i = \lambda_i e_i$, minden i -re. Ez az egyenlet a definíciója a sajátérték egyenletnek. Ezáltal, $AE = ED$ -nak igaznak kell lennie. Az átrendezés $A = EDE^{-1}$ -et eredményez, ezzel befejeztük a bizonyításunk első részét.

A bizonyítás második felében pedig meg kell mutatnunk, hogy a szimmetrikus mátrixnak mindig ortogonális sajátvektorai vannak. Valamilyen szimmetrikus mátrixnak legyen λ_1 és λ_2 két különböző sajátértéke e_1 és e_2 sajátvektorból.

$$\begin{aligned} \lambda_1 e_1 \cdot e_2 &= (\lambda_1 e_1)^T e_2 \\ &= (Ae_1)^T e_2 \\ &= e_1^T A^T e_2 \\ &= e_1^T A e_2 \\ &= e_1^T (\lambda_2 e_2) \\ \lambda_1 e_1 e_2 &= \lambda_2 e_1 e_2 \end{aligned}$$

Az utolsó összefüggésből átrendezéssel megkapjuk, hogy $(\lambda_1 - \lambda_2)e_1 \cdot e_2 = 0$.

Mivel kikötöttük, hogy a sajátvektorok különbözőek, így a kivonás eredménye nem lehet nulla, ebből következik, hogy $e_1 \cdot e_2 = 0$. Tehát, a szimmetrikus mátrix sajátvektorai ortogonálisak.

Térjünk vissza az eredeti posztulátumunkhoz, miszerint A egy szimmetrikus mátrix. A bizonyítás második feléből tudjuk, hogy A sajátvektorai ortonormáltak (úgy választjuk meg őket, hogy normalizáltak legyenek). Ez azt jelenti, hogy E egy ortogonális mátrix, így az 1. tétel alapján, $E^T = E^{-1}$ és felírhatjuk a végső megoldást.

$$A = EDE^T$$

Tehát, egy szimmetrikus mátrix sajátvektoraiból felépített mátrix által diagonalizálható.

5. Bármely $m \times n$ -es X mátrix, az $X^T X$ szimmetrikus mátrix rendelkezik egy készlet ortonormált sajátvektorral $\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n\}$ és egy készlet hozzájuk tartozó sajátértékkel $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$. A vektorkészlet így egy ortonormált bázist alkot $\{X\hat{v}_1, X\hat{v}_2, \dots, X\hat{v}_n\}$, ahol minden $X\hat{v}_i$ elem $\sqrt{\lambda_i}$ hosszúságú.

Az összes ilyen tulajdonság a két készlet pontszorzatából jön létre.

$$(X\hat{v}_i) \cdot (X\hat{v}_j) = (X\hat{v}_i)^T (X\hat{v}_j)$$

$$= \hat{v}_i^T X^T X \hat{v}_j$$

$$= \hat{v}_i^T (\lambda_j \hat{v}_j)$$

$$= \lambda_j \hat{v}_i \cdot \hat{v}_j$$

$$(X\hat{v}_i) \cdot (X\hat{v}_j) = \lambda_j \delta_{ij}$$

Az utolsó egyenlet X sajátvektorainak ortonormáltságából adódik, ez juttat minket a Kronecker delta függvényhez

$$(X\hat{v}_i) \cdot (X\hat{v}_j) = \begin{cases} \lambda_j & i = j \\ 0 & i \neq j \end{cases}$$

Ez az egyenlet azt mondja ki, hogy bármely két vektor ebből a készletből ortogonális.

A második tulajdonság is a fenti egyenletből származik abból a felismerésből, hogy a vektorok hosszának négyzete kifejezhető:

$$\|X\hat{v}_i\|^2 = (X\hat{v}_i) \cdot (X\hat{v}_i) = \lambda_i$$

Matlab Kód

Ez a kód Matlab 6.5-ra íródott és a Mathworks-től származik.[2][6]

A kód szemléltetés jellegű.

```
function [signals,PC,V] = pca1(data)

% PCA1: Perform PCA using covariance.
% data - MxN matrix of input data
% (M dimensions, N trials)
% signals - MxN matrix of projected data
% PC - each column is a PC
% V - Mx1 matrix of variances

[M,N] = size(data);
% subtract off the mean for each dimension
mn = mean(data,2);
data = data - repmat(mn,1,N);

% calculate the covariance matrix
covariance = 1 / (N-1) * data * data';

% find the eigenvectors and eigenvalues

[PC, V] = eig(covariance);

% extract diagonal of matrix as vector
V = diag(V);

% sort the variances in decreasing order
[junk, rindices] = sort(-1*V);
V = V(rindices);
PC = PC(:,rindices);

% project the original data set
signals = PC' * data;
```