

RECONeixEMENT AUTOMATIC D'INSTRUMENTS MUSICALS

Juan Castro Mayorgas

Memòria tècnica corresponent al
treball de fi de carrera de:

Enginyeria Tècnica de Telecomunicacions especialitat Telemàtica

Universitat Oberta de Catalunya

Segon semestre 2010-2011

Consultora: Helenca Duxans Barrobés

Per al Pol, el Quim i l'Alexandra

ABSTRACT

La identificació d'instruments musicals és una de les principals vies de treball per a donar solucions a la necessitat de disposar d'eines automàtiques de classificació i cerca a partir del contingut de les gravacions musicals. Aquest treball estudia el comportament d'un conjunt rellevant de característiques acústiques amb les tècniques de classificació Models de Mescles Gaussians, *Support Vector Machines* i k-veí-més-proper en el reconeixement automàtic de 18 instruments musicals. Pels experiments s'ha utilitzat un corpus de més de 115.000 mostres procedents de TunedIt i un altre de generació pròpia amb uns resultats d'encerts de més del 90 % en els dos casos. L'estudi posa de manifest com la rellevància dels paràmetres acústics així com la facilitat d'identificar un instrument en particular tenen dependència amb el classificador utilitzat.

Paraules clau: Reconeixement automàtic d'instruments musicals, identificació d'instruments musicals, reconeixement de patrons, indexació automàtica, processament del senyal musical, reconeixement de timbre musical.

The identification of musical instruments is one of the main areas of research to provide solutions to the need for automatic tools for indexing from the contents of recorded music. This paper studies the behavior of a set of relevant acoustic characteristics together with Gaussian Mixture Model, Support Vector Machines and k-Nearest Neighbor classification methods in the recognition of 18 musical instruments. Over 115,000 samples coming from TunedIt and another set of samples recorded by the author have been used for the experiments, reaching a success rate over 90 % in both cases. The study highlights that both the relevance of the acoustic parameters and the level of difficulty in identifying a particular instrument are related to the classifier.

Keywords: Automatic musical instruments recognition, musical instruments identification, pattern recognition, automatic indexing, musical signal processing, recognition of musical timbre.

La identificación de instrumentos musicales es una de las principales vías de trabajo para dar soluciones a la necesidad de disponer de herramientas automáticas de clasificación y búsqueda a partir del contenido de las grabaciones musicales. Este trabajo estudia el comportamiento de un conjunto relevante de características acústicas con las técnicas de clasificación Modelo de Mezclas Guassianas, *Support Vector Machine* y k-más-próximo-vecino en el reconocimiento de 18 instrumentos musicales. Para los experimentos se ha utilizado un corpus de más de 115.000 muestras procedentes de TunedIt y otro de generación propia con unos resultados de aciertos de más del 90 % en ambos casos. El estudio pone de manifiesto como la relevancia de los parámetros acústicos así como la facilidad de identificar un instrumento en particular están relacionados con el clasificador utilizado.

Palabras clave: Reconocimiento automático de instrumentos musicales, identificación de instrumentos musicales, reconocimiento de patrones, indexación automática, procesamiento de la señal musical, reconocimiento del timbre musical.

RESUM

Tenim accés a una enorme quantitat de música gravada en format digital. Els algorismes de compressió d'àudio i l'abaratiment dels suports digitals han propiciat la proliferació d'equips portàtils de reproducció molt petits i de gran capacitat, així com l'aparició de serveis en línia de difusió musical a la carta. L'usuari d'avui ja no té prou amb establir els criteris de cerca musical basant-se en el títol, autor o intèrpret. Ara, demana tindre l'opció d'escorcollar la música a través de paràmetres de tipus semàntic com ara l'estil musical o l'instrument que sona. Amb aquesta finalitat, la identificació d'instruments musicals és una de les principals vies de treball per a donar solucions a la necessitat de disposar d'eines automàtiques de classificació i cerca a partir del contingut de les gravacions musicals.

La teoria de sistemes sembla ser el camí natural per abordar aquesta problemàtica ja que existeix una clara e intuïtiva relació entre la música i els sistemes. Aquesta relació té el seu màxim representant en el timbre musical, el qual descriu tots els aspectes d'un so musical que no tenen res a veure amb la nota, la seva durada o el seu volum. És l'atribut que caracteritza un instrument d'un altre i que, per exemple, permet distingir entre el so de la nota 'mi' produïda per un clarinet de la mateixa nota 'mi' produïda per un oboè. El timbre s'identifica, en la teoria musical, amb la sèrie harmònica la qual, en teoria de sistemes, es coneix com espectre freqüencial, on el nombre i altura de cadascú dels harmònics són característics per a cada instrument. De l'espectre, també anomenat domini freqüencial, s'obtenen més dades d'interès, amb major o menor rellevància, per al reconeixement d'instruments musicals. El mateix passa amb en el domini temporal, d'on també es treuen característiques tímbriques associades als instruments.

Però el reconeixement d'instruments musicals no és una tasca senzilla. El timbre que el caracteritza pot variar sensiblement depenent de factors com la qualitat del propi instrument, l'expressió que l'intèrpret dóna al passatge musical o el nivell d'experiència del mateix intèrpret. Per altra banda, el so musical també pot estar barrejat amb soroll i, sobretot, amb més instruments sonant al uníson.

A més a més, s'ha d'elaborar un sistema d'entrenament que s'alimenti amb les característiques musicals de múltiples mostres degudament etiquetades i que faci la funció de model per a les posteriors consultes d'identificació. Les tècniques de classificació associades a aquest procés final d'identificació estan relacionades amb l'àmbit de la intel·ligència artificial.

Aquest treball estudia les etapes que tenen lloc en el procés de reconeixement automàtic d'instruments musicals i desenvolupa una sèrie de funcions d'alt nivell en Matlab que permeten, per un costat, l'extracció de característiques musicals d'arxius d'àudio i, per l'altre costat, la generació dels models d'entrenament i test a partir d'aquestes característiques. Amb un corpus de producció pròpia i un altre extern de més de 115.000 mostres, es desenvolupa un altre conjunt de funcions amb les que mitjançant experiments pràctics s'analitza el comportament de les característiques musicals i mètodes de classificació segons diferents criteris i escenaris.

Amb aquests resultats experimentals i, sota l'objectiu inicial d'identificar instruments aïllats en fraseig monofònic, s'implementa una funció d'alt nivell que, a partir d'una llista d'arxius d'àudio dona com a resultat la identificació de cada un d'ells. Finalment, aquesta funció s'aplica sobre arxius amb contingut polifònic d'un mateix instrument per a veure el seu comportament.

AGRAÏMENTS

Primer de tot voldria donar el meu agraïment en la realització d'aquest treball a la meva consultora de TFC Helenca Duxans, qui ha estat sempre vigilant perquè les coses sortissin com han de sortir.

I, com no, agraeixo a la Mayte el seu suport incondicional durant el transcurs de tot el projecte, en el que no només ha vist com molts caps de setmana compartia més temps amb l'ordinador que amb ella sinó que també m'ha ajudat en la producció de les gravacions d'àudio.

ÍNDEX DE CONTINGUTS

ABSTRACT	ii
RESUM	iv
AGRAÏMENTS	vi
ÍNDEX DE CONTINGUTS.....	vii
ÍNDEX DE FIGURES	ix
ÍNDEX DE TAULES	xi
CAPÍTOL 1: INTRODUCCIO.....	1
1.1. JUSTIFICACIO, PUNT DE PARTIDA I APORTACIO	1
1.2. OBJECTIUS	1
1.3. PRODUCTES OBTINGUTS	2
1.4. BREU DESCRIPCIO DELS ALTRES CAPÍTOLS.....	3
CAPÍTOL 2: ESTAT DE L'ART	4
2.1. INTRODUCCIO.....	4
2.2. PERCEPCIO DEL SO.....	4
2.2.1. Fisiologia del sistema auditiu humà	4
2.2.2. Propietats de percepció	6
2.2.3. El procés de reconeixement en les persones	11
2.3. LES CARACTERISTIQUES ACUSTIQUES.....	14
2.3.1. Domini freqüencial	14
2.3.2. Domini temporal.....	18
2.4. METODES DE CLASSIFICACIO	19
2.4.1. Model de Mescles de Gaussians (GMM)	20
2.4.2. k-NN (k-veí més proper).....	21
2.4.3. SVM (<i>Support Vector Machines</i>)	22
2.5. EL PROBLEMA DE LA IDENTIFICACIO	24
2.5.1. Tons aïllats.....	24
2.5.2. Fraseig monofònic.....	24
2.5.3. Polifonia.....	25
CAPÍTOL 3: DISSENY I IMPLEMENTACIO	27
3.1. INTRODUCCIO.....	27

3.2. ENTORN DE DESENVOLUPAMENT.....	28
3.3. DIAGRAMA DE BLOCS	28
3.3.1. Entrada d'àudio	30
3.3.2. Preprocessament i extracció de característiques	31
3.3.3. Guardar dades d'entrenament o test.....	32
3.3.4. Classificació	34
3.3.5. Presa de decisió i mostra de resultats	40
3.4. DESCRIPCIO DELS CORPUS	44
3.4.1. Corpus importat	44
3.4.2. Corpus de generació pròpia	46
3.4.3. Microsoft Musical Instruments	46
3.5. MODE D'OPERACIO	47
CAPITOL 4: PROVES I EXPERIMENTS.....	48
4.1. INTRODUCCIO.....	48
4.1. CORPUS IMPORTAT DE TUNEDIT	48
4.1.1. Rellevància de cada característica acústic	48
4.1.2. Combinació de paràmetres acústics	51
4.1.3. Variacions en el nombre de mostres.....	54
4.1.4. Variacions en el metode k-NN.....	55
4.1.5. Matrius de confusió.....	58
4.2. CORPUS DE GENERACIO PROPIA	62
4.2.1. Rellevància de cada característica acútica	62
4.2.2. Combinació de totes les característiques.....	64
4.2.3. Variacions en el mètode k-NN.....	65
4.2.4. Decisió del sistema a nivel d'arxius.....	66
CAPITOL 5: CONCLUSIONS.....	74
BIBLIOGRAFIA	77

ÍNDIX DE FIGURES

Figura 1. Fisiologia de la orella (font: http://www.educa365.cat).....	5
Figura 2. Relació mels – Hz	9
Figura 3. Formes d'ona d'un La4 produïdes per un clarinet, un oboè i una trompeta... 10	
Figura 4. Etapes del processament auditiu en el reconeixement de sons (McAdams) ... 11	
Figura 5. Percentatge d'encerts d'identificació d'instruments musicals per persones (Tae Hong Park).....	13
Figura 6. Formes d'ona i espectre de to pur, clarinet, oboè i trompeta	15
Figura 7. Conjunt de filtres mel pel càlcul dels MFCCs.....	17
Figura 8. <i>Roll-Off</i> (font: MirToolbox).....	18
Figura 9. Procés d'entrenament i identificació en un reconeixedor automàtic	19
Figura 10. Esquematització del procés de classificació amb k-NN (font: Wikipedia)	22
Figura 11 . exemple de SVM amb un kernel no lineal i un altre lineal (font: wikipedia) 23	
Figura 12. Resum de percentatge d'encert de tons aïllats (Font: [7]).....	24
Figura 13. Diagrama general del sistema	29
Figura 14. Diagrama del mòdul de classificació	34
Figura 15. Diagrama del mòdul de presa de decisió i mostra de resultats	40
Figura 16. Percent. d'encerts de les característiques aïllades aplicant GMM.....	49
Figura 17. Percent. d'encerts de les característiques aïllades aplicant 1-NN	50
Figura 18. Percent. d'encerts de les característiques aïllades aplicant SVM.....	50
Figura 19. Percent. d'encerts amb combinació de característiques aplicant GMM.....	51
Figura 20. Percent. d'encerts amb combinació de característiques aplicant 1-NN	52
Figura 21. Percent. d'encerts amb combinació de característiques aplicant SVM.....	53
Figura 22. Percent. d'encert segons el nombre de mostres d'aprenentatge	54
Figura 23. Percentatge d'encerts en funció de k aplicant k-NN sobre totes les característiques	55
Figura 24. Percentatge d'encerts en funció de k aplicant k-NN sobre centroide	56
Figura 25. Percent. d'encerts en funció de k aplicant k-NN (totes les característiques) amb suport de SVM.....	57

Figura 26. Percent. d'encerts en funció de k aplicant k-NN (centroide) amb suport de SVM.....	58
Figura 27. Matriu de confusió (instruments) aplicant GMM	59
Figura 28. Matriu de confusió (família i subfamília) aplicant GMM	59
Figura 29. Matriu de confusió (instruments) aplicant 1-NN.....	60
Figura 30. Matriu de confusió (família i subfamília) aplicant 1-NN.....	60
Figura 31. Matriu de confusió (instruments) aplicant SVM	61
Figura 32. Matriu de confusió (família i subfamília) aplicant SVM	61
Figura 33. Percent. d'encerts aplicant GMM	63
Figura 34. Percent. d'encerts aplicant 1-NN.....	63
Figura 35. Percent. d'encerts aplicant SVM	64
Figura 36. Percent. d'encert segons mètode amb totes les característiques.....	65
Figura 37. Percentatge d'encerts en funció de k aplicant k-NN sobre totes les característiques	65
Figura 38. Percentatge d'encerts en funció de k aplicant k-NN sobre els 13 MFCCs....	66

ÍNDIX DE TAULES

Taula 1. Relacions freqüència-intensitat percebudes per igual en l'oïda (David Lapp) ...	7
Taula 2. Freqüència de les notes musicals en afinació temperada igual.....	8
Taula 3. Llistat dels instruments musicals que reconeixerà el sistema	30
Taula 4. Funcions TFCLoadAudioList i TFCTGetFeatures	32
Taula 5. Funcions TFCSaveFeatures i TFCCreateFeatures	33
Taula 6. Descripció del vector de selecció de característiques	35
Taula 7. Funcions TFCLoadFeatures, TFCCreateCheckMatrices i TFCScaleMatrices ..	36
Taula 8. Funcions TFCTGMMCheck i TFCTGMMClassify.....	37
Taula 9. Funcions TFCKNNClassify i TFCKNNClassify2.....	39
Taula 10. Funció TFCSVMClassify	39
Taula 11. Descripció dels camp de la matriu d'anàlisi.....	41
Taula 12. Funcions TFCCreateAnalytics, TFCSaveAnalytics i TFCLoadAnalytics.....	41
Taula 13. Funcions TFCTReportAnalytics i TFCSaveReportAnalytics	42
Taula 14. Descripció de la taula resum de decisió	43
Taula 15. Funcions TFCTReportDecision i TFCSaveReportDecision.....	43
Taula 16. Descripció dels camps del corpus importat de TunedIt.....	45
Taula 17. Funció TFCKNNClassify3.....	57
Taula 18. Resultats de classificació de l'instrument aplicant GMM.....	67
Taula 19. Resultats de classificació de l'instrument aplicant 1-NN	67
Taula 20. Resultats de classificació de l'instrument aplicant SVM.....	68
Taula 21. Resultats de classificació aplicant GMM.....	69
Taula 22. Resultats de classificació aplicant 1-NN	70
Taula 23. Resultats de classificació aplicant SVM.....	71
Taula 24. Resultats de classificació aplicant GMM.....	72
Taula 25. Resultats de classificació aplicant 1-NN	72
Taula 26. Resultats de classificació aplicant SVM.....	73

CAPÍTOL 1

INTRODUCCIO

1.1. JUSTIFICACIO, PUNT DE PARTIDA I APORTACIO

D'acord amb el pla docent de l'assignatura, “*El treball de Fi de Carrera (TFC) és una assignatura pensada per a realitzar un treball de síntesi dels coneixements adquirits al llarg de la carrera i que requereix posar-los en pràctica conjuntament amb un treball concret*”. Dintre de les diferents ofertes d'àrees de treball en què es pot cursar aquesta assignatura, l'elecció de l'àrea d'aplicacions multimèdia basades en processament de senyal i en concret de la especialitat de processament digital de l'àudio amb una proposta a mida, té la seva justificació personal en la confluència de tres camps de gran interès: música, teoria de sistemes i informàtica.

D'altra banda, el reconeixement automàtic d'instruments musicals no és un problema resolt en la seva totalitat i aquest TFC experimenta amb els diferents actors que hi participen per a finalment implementar una solució pròpia que assoleix els objectius inicialment plantejats.

Aquest treball aporta resultats experimentals a partir de dos corpus, un de generació pròpia i un altre de procedència externa amb un gran volum de mostres. A més a més, amb el conjunt de funcions desenvolupades també és possible definir nous escenaris (amb la generació de nous corpus si s'escau) i veure com afecta una nova parametrització en el resultat de la identificació.

1.2. OBJECTIUS

Els objectius fixats en aquest treball són els següents:

- Conèixer quines són les etapes que formen part d'un sistema de reconeixement de patrons.
- Conèixer un conjunt de característiques acústiques associades als sons, el seu significat i com es calculen.
- Conèixer les tècniques de classificació que sovint s'utilitzen en sistemes de reconeixement, els seus fonaments teòrics i les seves particularitats.
- Desenvolupar una sèrie de funcions que permetin l'extracció de característiques acústiques del contingut d'arxius àudio.
- Desenvolupar una sèrie de funcions que permetin classificar característiques acústiques i obtenir-ne un resultat percentual d'encerts.
- Definir i executar un conjunt de proves orientades a determinar el comportament dels sistemes de classificació en diferents condicions inicials.
- Desenvolupar una funció d'alt nivell que, a partir d'una llista d'arxius d'àudio, doni com a resultat l'instrument que sona en cada un d'ells.
- Veure el comportament del sistema amb una mostra d'arxius amb contingut polifònic.

1.3. PRODUCTES OBTINGUTS

Com a resultat d'aquest TFC es lliura tres productes:

- Una memòria tècnica que condensa tot el treball de recerca juntament amb el disseny del sistema de reconeixement i els resultats de les proves realitzades.
- Una presentació virtual que recapitula els aspectes més rellevants d'aquest treball de fi de carrera.

- Un desenvolupament en Matlab que inclou, per un costat, un conjunt de funcions que permeten l'extracció de característiques acústiques d'arxius d'àudio i classificació de mostres i, per l'altre, una funció d'alt nivell que identifica els instruments musicals que sonen en una llista d'arxius d'àudio.

1.4. BREU DESCRIPCIÓ DELS ALTRES CAPÍTOLS

Els següents capítols de la memòria estan estructurats segons l'ordre lògic de desenvolupament del treball.

El **capítol 2** correspon a la fase de recerca del treball, on s'adquireixen els coneixements necessaris per desenvolupar-lo. Inclou els conceptes que hi són presents i el seu estat de l'art a través de treballs ja desenvolupats sobre el mateix àmbit.

El **capítol 3** inclou el desenvolupament del conjunt de funcions de les eines d'extracció, classificació i decisió que formen els mòduls d'un sistema de reconeixement basat en patrons.

El **capítol 4** comprèn el conjunt de proves realitzades i els resultats obtinguts segons diverses configuracions de nombre de mostres, mètodes de classificació i nombre i tipus de característiques acústiques utilitzades amb la finalitat de buscar la configuració que doni el millor percentatge d'encerts. També s'inclouen proves amb fonts d'àudio polifòniques.

El **capítol 5** recull les conclusions tretes de les proves realitzades juntament amb les línies següent d'investigació.

CAPÍTOL 2

ESTAT DE L'ART

2.1. INTRODUCCIO

Aquest capítol consta de cinc apartats. En primer lloc, s'estudia com es produeix la percepció del so en els humans, les propietats que descriuen els sons i quines són les etapes que es desenvolupen en el cervell en el procés de reconeixement d'instruments. A continuació, s'analitza un conjunt de paràmetres acústics que es poden extreure dels sons i que, en major o menor grau de rellevància, els caracteritzen. Seguidament, es procedeix a l'estudi de les tècniques de classificació més habituals que es fan servir en el procés d'identificació de l'instrument. El capítol acaba amb els nivells de complexitat en què es pot desglossar un sistema automàtic de reconeixement d'instruments musicals i els resultats obtinguts en recerques ja realitzades.

2.2. PERCEPCIO DEL SO

El coneixement de com els humans perceben els sons és important per esbrinar què és rellevant i què no ho és a l'hora d'escollir els atributs acústics amb l'objectiu de dissenyar un sistema automàtic de reconeixement d'instruments musicals.

2.2.1. Fisiologia del sistema auditiu humà

La fisiologia del sistema auditiu humà ha estat estudiada en profunditat i existeix una gran quantitat d'informació sobre aquest tema [1,2]. El mecanisme que fa arribar l'audició al cervell és complex i el seu objectiu és codificar les ones acústiques en impuls elèctrics. La Figura 1 és una representació del sistema auditiu humà.

L'orella comprèn tres grans divisions: orella externa, orella mitjana i orella interna:

- **L'orella externa:** la seva funció és captar el so i, mitjançant el conducte auditiu extern que actua com a ressonador, amplifica les ones sonores. En aquest punt, les ones sonores encara estan en la seva forma acústica.
- **L'orella mitjana:** s'encarrega de transmetre el so. Per aquest fi, el timpà transmet les fluctuacions de pressió a través dels tres ossets (anomenats martell, esclusa i estrep) fins l'orella interna.
- **L'orella interna:** està composta per l'aparell vestibular, amb els tres conductes semicirculars de l'equilibri, i de la còclea, amb l'òrgan auditiu. La còclea conté dos conductes (plens de perilimfa): la rampa vestibular i la rampa timpànica. Dintre de la còclea es troben un conjunt d'unes 30.000 cèl·lules ciliades que transformen els canvis hidrodinàmics en polsos nerviosos fins al cervell a través del nervi auditiu, que pot arribar a transmetre fins a 1.500 diferències d'altura de so (on l'altura del so depèn de la posició de les cèl·lules majorment excitades) i 325 graus d'intensitat.

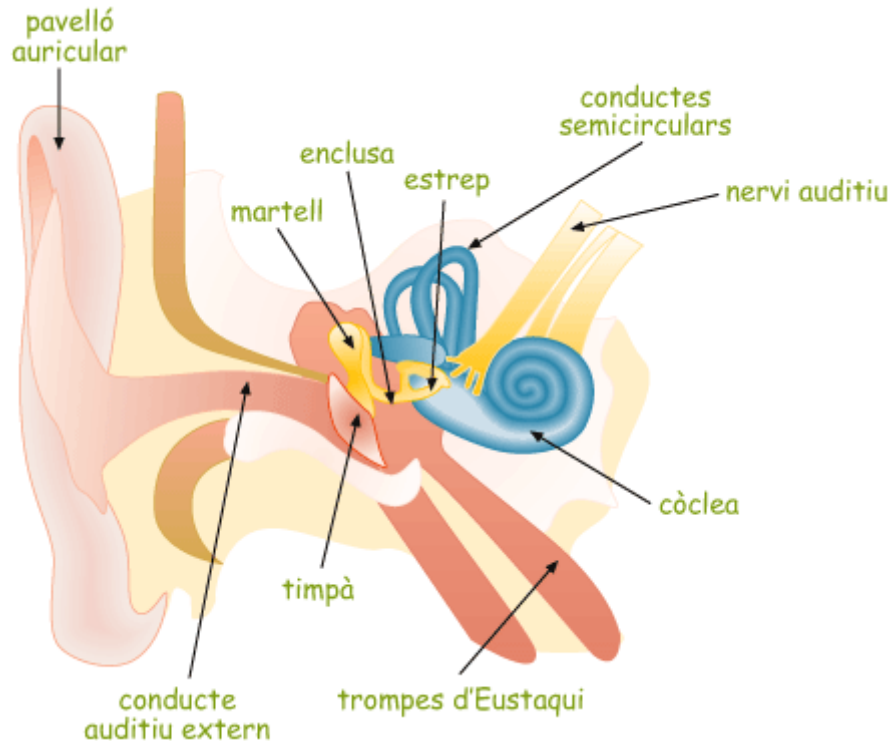


Figura 1. Fisiologia de la orella (font: <http://www.educa365.cat>)

2.2.2. Propietats de percepció

Els sons musicals tenen quatre atributs de percepció: intensitat i volum, altura, duració i timbre. Aquests quatre atributs són els que fan possible distingir dos sons musicals entre ells.

2.2.2.1. Intensitat i volum

Els atributs de volum i intensitat informen sobre la quantitat de so que es genera i que es percep. La intensitat és un valor absolut que pot mesurar-se quantitativament i s'expressa com energia per unitat de temps i unitat de superfície ($J/s \cdot m^2 = W/m^2$). El volum, per altra banda, és un valor subjectiu que depèn de la sensibilitat auditiva de l'oient.

L'oïda humana és sensible a intensitats que van des dels $10^{-12} W/m^2$ (el so més suau que es pot percebre a una freqüència de 1.000 Hz) fins a $1 W/m^2$ (llindar de dolor). Per treballar amb aquests intervals es fa servir l'escala logarítmica i es defineix el Bel (B) com el logarítmic del rati entre la intensitat considerada i la intensitat en el llindar d'audició. Matemàticament s'expressa com:

$$L_B = \log_{10} \left(\frac{I}{I_0} \right)$$

Aquesta relació d'intensitats s'acostuma a expressar en decibels, que és un submúltiple del Bel:

$$L_{dB} = 10 \log_{10} \left(\frac{I}{I_0} \right)$$

Com es tracta d'una mesura relativa, es pren com a conveni el valor de 0 dB per a la intensitat a la qual un so amb freqüència de 1.000 Hz deixa de ser audible.

La percepció de la intensitat sonora, per la seva banda, depèn de la freqüència del so [3]. La taula següent mostra, per a diferents freqüències, el nivell d'intensitat que ha de tenir per tal de percebre-les amb el mateix volum. Per exemple, l'oïda percep amb el mateix

volum una freqüència de 4.000 Hz amb 2 dB d'intensitat que una freqüència de 50 Hz amb 43 dB d'intensitat.

Freqüència (Hz)	Nivell d'intensitat (dB)
50	43
100	30
200	19
500	11
1000	10
2000	8
3000	3
4000	2
5000	7
6000	8
7000	11
8000	20
9000	22
14000	31

Taula 1. Relacions freqüència-intensitat percebudes per igual en l'oïda (David Lapp)

2.2.2.2. Altura

L'altura del so indica la freqüència de vibració de la font que l'origina. Els llimars d'audició en els éssers humans està entre els 16 i els 20.000 Hz i disminueixen significativament amb l'edat. Tot i això, per sobre dels 4.200 Hz, és molt difícil distingir altures de so.

En l'àmbit musical, l'altura dels sons no és arbitrària. Està definida segons el tipus d'afinació (relació d'altura que s'estableix entre les notes) aplicada a l'escala musical (successió ordenada de les notes en altura) i prenent una freqüència com a referència. Entre els diferents tipus d'afinació existents [4], la música occidental va adoptar des del segle XVIII el sistema d'afinació temperada igual. Aquest sistema estableix que la relació entre les freqüències de dos notes musicals consecutives és de $\sqrt[12]{2}$. Per altra banda, la freqüència de referència que es fa servir per a l'afinació és habitualment de 440 Hz i correspon a la tecla "La" central d'un piano.

Les notes musicals s'identifiquen internacionalment per les lletres A, B, C, D, E, F i G. A Espanya, França i Itàlia, es segueix utilitzant la nomenclatura tradicional (Do, Re, Mi, Fa, Sol, La i Si). L'equivalència amb el sistema internacional és la següent:

A = La	B = Si	C = Do	D = Re	E = Mi	F = Fa	G = Sol
--------	--------	--------	--------	--------	--------	---------

Nota	Freq. (Hz)	Nota	Freq. (Hz)	Nota	Freq. (Hz)	Nota	Freq. (Hz)
A ₀	27,50	G ₂	98,00	F ₄	349,23	D# ₆	1244,51
A# ₀	29,16	G# ₂	103,83	F# ₄	369,99	E ₆	1318,51
B ₀	30,89	A ₂	110,00	G ₄	392,00	F ₆	1396,91
C ₁	32,70	A# ₂	116,54	G# ₄	415,30	F# ₆	1479,98
C# ₁	34,65	B ₂	123,47	A₄	440,00	G ₆	1567,98
D ₁	36,71	C ₃	130,81	A# ₄	466,16	G# ₆	1661,22
D# ₁	38,89	C# ₃	138,59	B ₄	493,88	A ₆	1760,00
E ₁	41,20	D ₃	146,83	C ₅	523,25	A# ₆	1864,65
F ₁	43,65	D# ₃	155,56	C# ₅	554,37	B ₆	1975,53
F# ₁	46,25	E ₃	164,81	D ₅	587,33	C ₇	2093,00
G ₁	49,00	F ₃	174,61	D# ₅	622,25	C# ₇	2217,46
G# ₁	51,91	F# ₃	185,00	E ₅	659,26	D ₇	2349,32
A ₁	55,00	G ₃	196,00	F ₅	698,46	D# ₇	2489,02
A# ₁	58,27	G# ₃	207,65	F# ₅	739,99	E ₇	2637,02
B ₁	61,74	A ₃	220,00	G ₅	783,99	F ₇	2793,83
C ₂	65,41	A# ₃	233,08	G# ₅	830,61	F# ₇	2959,96
C# ₂	69,30	B ₃	246,94	A ₅	880,00	G ₇	3135,96
D ₂	73,42	C ₄	261,63	A# ₅	932,33	G# ₇	3322,44
D# ₂	77,78	C# ₄	277,18	B ₅	987,77	A ₇	3520,00
E ₂	82,41	D ₄	293,66	C ₆	1046,50	A# ₇	3729,31
F ₂	87,31	D# ₄	311,13	C# ₆	1108,73	B ₇	3951,06
F# ₂	92,50	E ₄	329,63	D ₆	1174,66	C ₈	4186,00

Taula 2. Freqüència de les notes musicals en afinació temperada igual

A l'igual que passa amb la intensitat, la percepció de l'altura del so no és lineal. Stevens, Volkman i Newmann van proposar en 1937 una escala per representar, d'una manera més fidedigna, la percepció que es té de la distància entre les altures dels sons [5]. Aquesta escala, denominada mel (de melodia), està basada en experiments amb tons purs en els que els oients ajustaven les freqüències que escoltaven al doble o meitat en altura en comparació amb una freqüència de referència.

L'escala mel equipara un to de 1000 Hz a 40 dB per sobre del llindar d'audició amb un to de 1000 mels. Fins els 500 Hz aproximadament, l'escala mel correspon linealment amb els Hz. Per sobre dels 500 Hz, els intervals de freqüència espaiats exponencialment són percebuts com si estiguessin espaiats linealment.

La fórmula que expressa la relació entre mels (m) i Hz (f) és la següent:

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

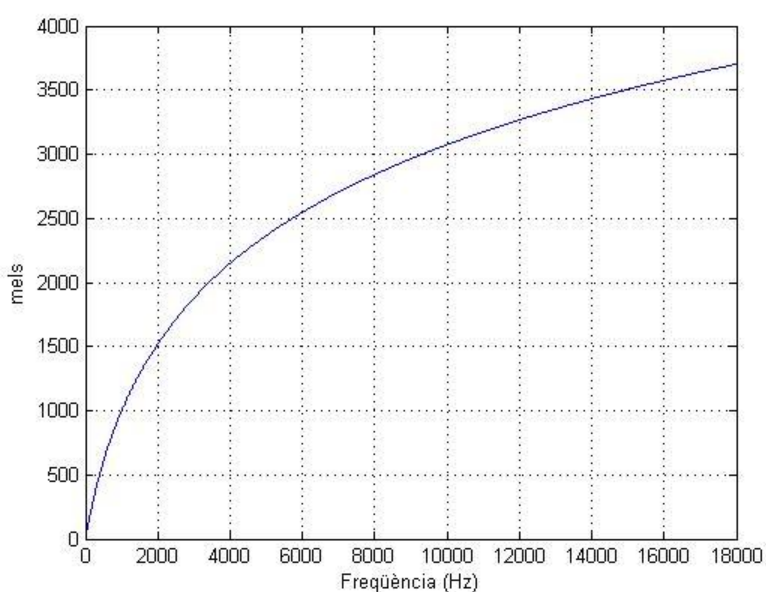


Figura 2. Relació mels – Hz

2.2.2.3. Duració

Tal com indica el propi nom, fa referència a la percepció del temps transcorregut des de què es rep el so fins què es deixa de percebre. Per a estimar la duració d'un so, l'oient acostuma a fer comparacions entre sons per així poder dir que un so és més llarg o curt que un altre.

Durant l'emissió del so en els instruments musicals es distingeixen tres fases:

- **Atac:** Correspon al temps transcorregut des del moment inicial de producció del so fins arribar a un punt d'intensitat constant.

- **Estat d'equilibri:** Període de duració variable segons les limitacions interpretatives i les físiques pròpies de l'instrument en el qual es manté una intensitat constant.
- **Decaïment:** Període d'atenuació del so fins la seva desaparició.

2.2.2.4. Timbre

El timbre, també denominat “color del so” [6], descriu tots els aspectes d'un so musical que no tenen relació amb els atributs mencionats amb anterioritat (intensitat, altura i duració). Així doncs, el timbre permet diferenciar dos instruments musicals que estan tocant la mateixa nota amb la mateixa intensitat i duració.

El timbre es degut al fet de què les notes generades pels instruments musicals són ones complexes formades per més d'una freqüència: la freqüència fonamental (l'altura del so que l'oient percep) i els seus sobretons harmònics. Els harmònics són freqüències múltiples sencers de la freqüència fonamental. El nombre i amplitud dels diferents sobretons caracteritzen l'instrument.

En el següent exemple es mostren les formes d'ona d'un clarinet, un oboè i una trompeta amb una freqüència fonamental de 440 Hz (La₄) en comparació al to pur:

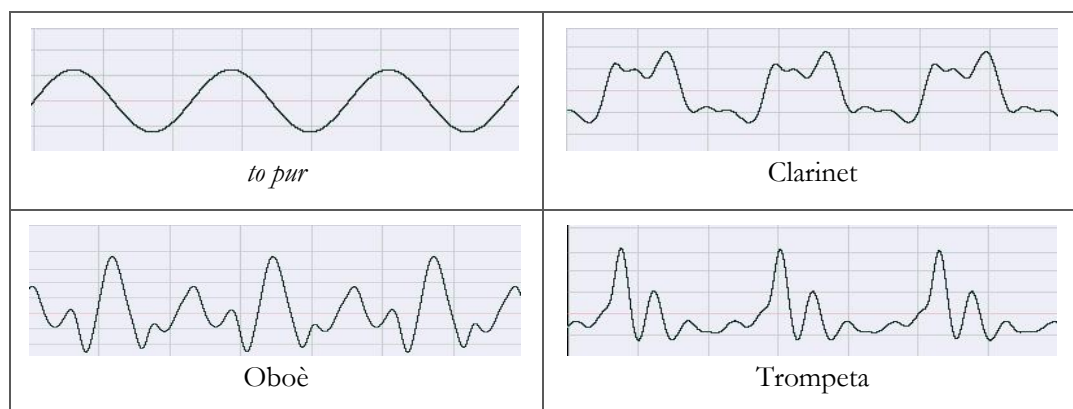


Figura 3. Formes d'ona d'un La₄ produïdes per un clarinet, un oboè i una trompeta

L'ésser humà és capaç d'apreciar molt petites diferències en el timbre. Així doncs, no només pot distingir entre dos instruments molts similars sinó que també pot distingir entre dos instruments iguals o el mateix instrument tocat per diferents persones.

2.2.3. El procés de reconeixement en les persones

Reconèixer vol dir que el que es sent coincideix d'alguna manera amb quelcom que s'ha escoltat amb anterioritat. Les recerques en l'àmbit del reconeixement de veu i d'instruments musicals [7,8] prenen com referència l'article de Stephen McAdams "*Thinking in sound: The Cognitive Psychology of Human Auditory*" (1993) que representa el procés humà de reconeixement segons el següent diagrama:

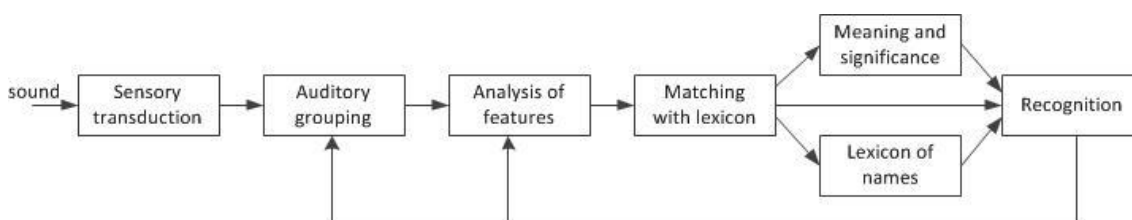


Figura 4. Etapes del processament auditiu en el reconeixement de sons (McAdams)

La primera etapa (*sensory transduction*) correspon al procés explicat en l'apartat 2.2.1. d'aquest capítol, en el qual la informació arriba al cervell a través del nervi auditiu.

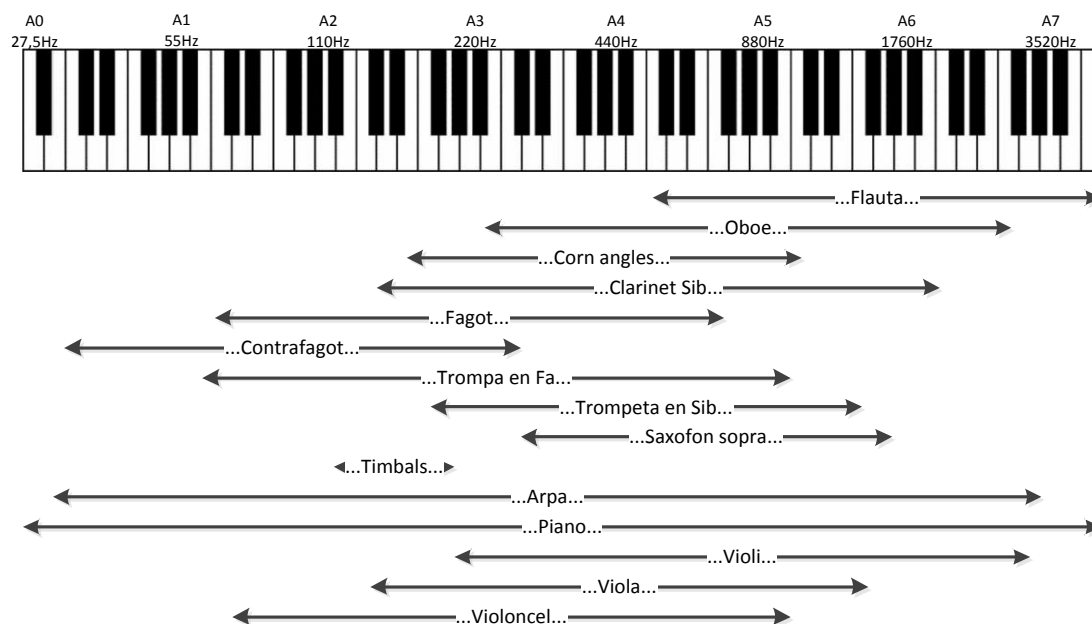
En l'etapa següent, "*auditory grouping*", el cervell processa la informació que li arriba, la descompon i classifica els seus components en grups per a poder fer seguidament el seu anàlisi ("*Analysis of features*").

Tot seguit, busca coincidències en la seva memòria per arribar a la possible identificació.

El sistema es retro alimenta, de manera que va adquirint experiència i fa possible identificacions progressives en entorns més complicats (soroll, etc.).

L'esquema de McAdams reflexa el que a vegades succeeix en el procés de reconeixement quan no es produeix una coincidència clara entre el resultat de l'anàlisi del so que es percep i la informació que el cervell té memoritzada: no és possible identificar l'instrument però sí que es pot assignar dintre d'una categoria.

Per una banda, l'altura dels sons que es perceben serveix per discriminar o seleccionar possibles candidats. La figura següent mostra l'extensió d'altures d'una sèrie d'instruments musicals:



Per l'altre costat, el tipus de mecanisme que un instrument fa servir per generar els sons és una forma natural e intuïtiva d'agrupar-los en famílies amb certes característiques tímbriques semblants.

Segons aquest mecanisme, els instruments musicals es poden classificar seguint un ordre jeràrquic tal com es mostra a continuació:

- **Puntejats:** com el baix, el banjo o el violí tocat en pizzicato.
- **Rascats:** com la guitarra i el llaüt.
- **Fregats:** com el violí, la viola i el violoncel.
- **Bufats**
 - **De bisell:** com tot tipus de flautes.
 - **De Llengüeta**
 - **Lliure:** com l'acordió.
 - **Simple:** com el clarinet i el saxofon.
 - **Doble:** com l'oboè.
 - **Per vibració dels llavis:** com la trompeta, el trombó i la tuba.
- **Colpejats:** com els tambors.

Aquest ordre jeràrquic explica perquè una persona que no hagi escoltat mai una zurna persa, li semblarà més similar a un oboè que a un clarinet: la zurna és un instrument de vent de doble llengüeta a l'igual que l'oboè quan el clarinet és de llengüeta simple.

Experiments d'audició d'instruments musicals orquestrals realitzats amb persones [9,10] revelen que:

- El rati d'encerts varia en funció del nombre d'instruments que formen part dels experiments.
- Hi ha instruments que es presten més a confusió (per exemple violí – viola – violoncel, trompa - saxofon).
- Músics amb experiència presenten certa confusió en la identificació correcta quan els sons estan fora de context.
- Si es suprimeix el moment d'atac del so en l'experiment, el rati d'encerts disminueix en un 30% de mitjana.
- A mesura que s'incrementa el nombre d'instruments, el rati d'encerts disminueix inclús pels músics experimentats.

La següent taula mostra un resum dels diferents experiments amb els instruments implicats i percentatge d'encerts:

Experiment	Data	Oboe	Sax.	Clar.	Flauta	General	Nombre d'instruments
Eagleson/Eagleson	1947		59	45	20	56	9
Saldhana/Corso	1964	75		84	61	41	10
Berger	1964					59	10
Clark/Milner	1964					90	3 (flauta, clar., oboe)
Strong/Clark	1967					85	8
Campbell/Heller	1978					72	6 (2 notes lligades)
Kendall	1986					84	3 (trumpeta, clar., violí)
Brown	1999	85	92			89	2 (oboe, sax.)
Martin	1999					46	27 (tons aïllats)
						67	27 (fragments de 10s)
Houix/McAdams/Brown	2001	87	87	71	93	85	4 (oboe, sax., clar., flauta)

Figura 5. Percentatge d'encerts d'identificació d'instruments musicals per persones (Tae Hong Park)

2.3. LES CARACTERISTIQUES ACUSTIQUES

D'un senyal acústic es pot treure informació tant del domini freqüencial com del temporal. Entre les múltiples característiques existents, (en [11] fan referència a un projecte que defineix 72 característiques), es descriuen aquelles que són més rellevants pel reconeixement d'instruments musicals.

2.3.1. Domini freqüencial

L'anàlisi de les característiques que es poden extreure dintre del domini freqüencial estan basades en la transformada discreta de Fourier (DFT) ja que la informació a analitzar ja es troba en forma discreta després d'un procés de mostreig. La fórmula de la DFT és la següent:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i\frac{2\pi k}{N}n} \quad k = 0, \dots, N-1$$

on:

- $x[n]$ correspon al senyal en forma discreta del senyal original $x(t)$ abans de ser digitalitzada.
- $X[k]$ és la transformada discreta de Fourier de $x[n]$

És important remarcar que per a no perdre informació del senyal original, segons el teorema de Nyquist, la velocitat de mostreig de $x(t)$ per generar $x[n]$ haurà de ser del doble de l'amplada de banda original, en aquest cas 44.000 Hz (2 x 20.000 Hz).

El càlcul de la DFT és complex i per reduir el temps de computació s'utilitza la denominada FFT (*Fast Fourier Transformation*). La FFT consisteix en un algorisme molt eficient per calcular la DFT.

La DFT dona l'espectre de freqüència del senyal d'entrada, que en el cas de sons musicals, correspon a la fonamental i els seus harmònics.

En el següent exemple es mostren les formes d'ona d'un clarinet, un oboè i una trompeta amb una freqüència fonamental de 440 Hz (L_{a4}) i els seus espectres freqüencials:

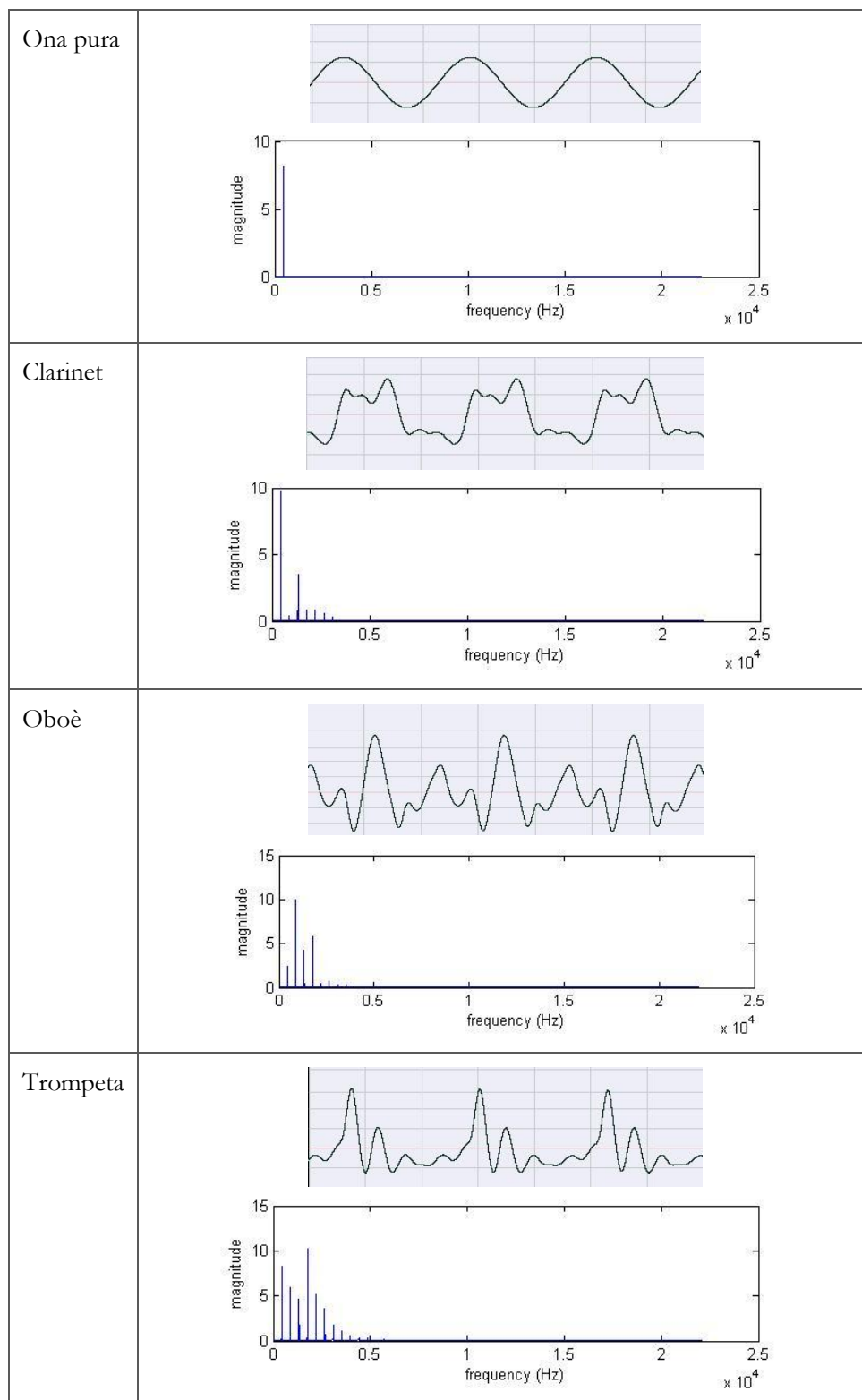


Figura 6. Formes d'ona i espectre de to pur, clarinet, oboè i trompeta

2.3.1.1. Centroide espectral (*Spectral Centroid*)

Aquest valor és l'equivalent al “centre de gravetat” de l'espectre freqüencial. Aquesta característica ha estat usada en [7,10,11,12,13,14,15]. Es relaciona molt directament amb la percepció de brillantor del so. A partir de l'espectre calculat amb la FFT, $X[0..N-1]$, el valor del centroide espectral es calcula com:

$$SC = \left(\sum_{k=0}^{N-1} k \cdot X[k] \right) / \left(\sum_{k=0}^{N-1} X[k] \right)$$

2.3.1.2. Extensió (*Spread*)

També referit com l'amplada del centroide. Aquesta característica ha estat usada en [12,13,14,15]. Mostra el rang de freqüències del senyal ponderat pel seu espectre. La seva expressió és la següent:

$$Spread = \left(\sum_{k=0}^{N-1} |SC - k| \cdot X[k] \right) / \left(\sum_{k=0}^{N-1} X[k] \right)$$

2.3.1.3. Coeficients espectrals de freqüència mel (MFCCs)

Aquests coeficients són molt comuns tant en el reconeixement d'instruments musicals com en el reconeixement de veu. Aquesta característica ha estat usada en [7,15,14,9,12,16]. Constitueixen una característica bàsica en la cerca de patrons en senyals d'àudio i estan basats en l'escala mel (veure apartat 2.2.2.2.).

L'obtenció dels MFCCs prové de l'espectre del so, del qual es calcula la magnitud de la energia de cada una de les bandes (normalment es fan servir 40) segons l'escala mel. Els coeficients MFCC són cadascú de la transformada discreta cosinus del logaritme de l'energia de cada banda. L'expressió matemàtica és la següent:

$$c_i = \frac{1}{N} \sum_{j=1}^N \log(m_j) \cos \left[\frac{\pi i}{N} \left(j - \frac{1}{2} \right) \right], \quad i = 0, \dots, N - 1$$

on:

- c_i correspon al coeficient i MFCC
- N és el nombre de bandes en què es divideix l'espectre
- m_j és l'energia de la banda (en escala de mel) j

La figura següent mostra un banc de filtres mel. Si la freqüència estigués en escala mel, l'amplada dels triangles seria uniforme.

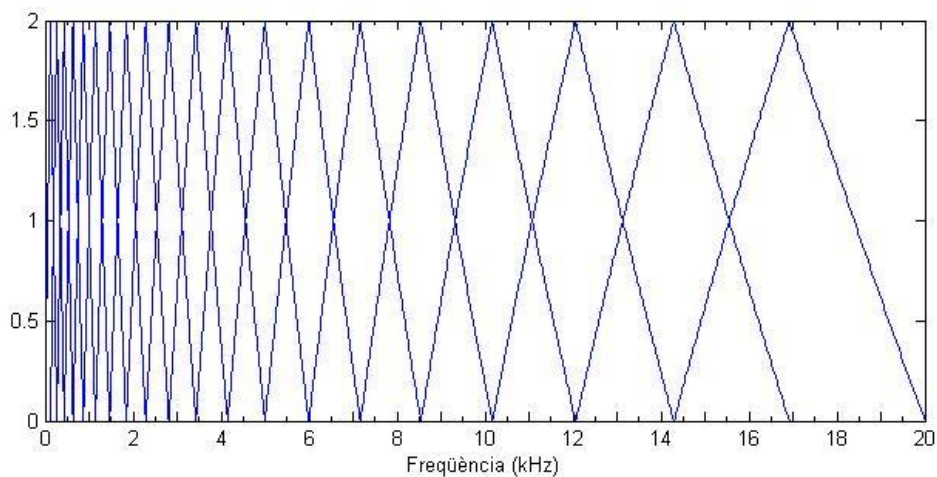


Figura 7. Conjunt de filtres mel pel càlcul dels MFCCs

Descartant el primer coeficient que només representa l'energia mitjana, es tenen en compte els 12-16 primers coeficients.

2.3.1.4. *Roll-Off*

Indica quins són els límits freqüencials on resideix el 85% de l'energia total de l'espectre. Aquesta característica s'ha utilitzat en [11,10]. És útil per a diferenciar entre sons constants d'altres amb més estats transitoris. S'expressa com:

$$\sum_{k=0}^R X[k] = 0.85 \cdot \sum_{k=0}^N X[k]$$

on:

- R és la freqüència de *roll-off*
- $X[k]$ és la magnitud de l'espectre en la posició k .

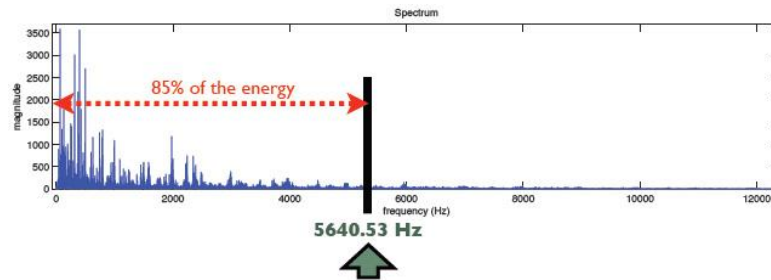


Figura 8. *Roll-Off* (font: MirToolbox)

2.3.2. Domini temporal

2.3.2.1. Taxa de pas per zero

Aquest és un indicador temporal molt simple que informa de la taxa de canvi de signe que es produeix en el senyal. Aquesta característica s'ha utilitzat en [15,10,14].

Pel tipus de contingut freqüencial que tenen la majoria de classes de soroll, el senyal canvia molt en el domini temporal, fent freqüents canvis de valors positius a valors negatius i, per tant, la taxa de pas per zero és un indicador que es pot utilitzar per veure el contingut “sorollós” del senyal que s'està analitzant.

Per altra banda, certs instruments, com els de percussió presenten cert “soroll” associat al seu timbre.

L'expressió de la taxa de pas per zero (ZCR) és la següent:

$$ZCR = \left(\sum_{n=1}^N |sign(x[n]) - sign(x[n-1])| \right) / (N - 1)$$

2.3.2.2. Energia mitjana

Aquest paràmetre, també conegut com la mitjana quadràtica, equival a la energia mitjana del senyal i és un indicador de la intensitat del senyal. Aquesta característica s'ha utilitza en [15,10]. La seva expressió és la següent:

$$RMS = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^N x[n]^2}$$

2.3.2.3. Temps d'atac

Descrit en [10], mesura el temps entre un punt inicial (determinat per un llindar) i el punt on l'amplitud arriba al seu màxim.

S'ha de considerar que aquest valor s'ha de calcular a partir del fragment de senyal que contingui tota la duració del so generat.

2.4. METODES DE CLASSIFICACIO

La finalitat d'extreure característiques acústiques dels sons és que serveixen per construir models matemàtics que permetin classificar aquestes característiques per a, posteriorment, ser capaç de classificar o identificar instruments musicals.

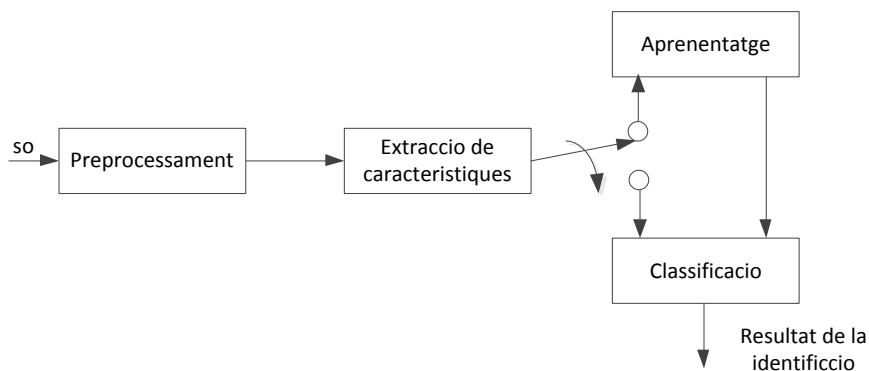


Figura 9. Procés d'entrenament i identificació en un reconeixedor automàtic

En l'esquema es mostra com el sistema genera la base de dades de patrons. De la qualitat i nombre de mostres enregistrades en la fase d'entrenament dependrà molt la fiabilitat en el procés posterior d'identificació d'instruments.

En aquest apartat s'analitzen tres mètodes que són molt habituals de trobar en l'àmbit del reconeixement de patrons.

2.4.1. Model de Mescles de Gaussians (GMM)

El model de Mescles de Gaussians (*Gaussian Mixture Model*) és un mètode de classificació de tipus probabilístic i sense supervisió. Aquest mètode classificació ha estat utilitzat en [7,11,9,13].

Un model de mescles és un model probabilístic en el que les dades pertanyen a una distribució de mescles. La seva funció de densitat de probabilitat és una combinació lineal de les altres funcions de densitat de probabilitat:

$$p(x) = \omega_1 p_1(x) + \omega_2 p_2(x) + \dots + \omega_n p_n(x)$$

Cada un dels $p_i(x)$ s'anomena component de mescla i els corresponents ω_i defineixen els pesos de cada mescla.

En el cas d'un model de mescles de Gaussians, la funció de densitat $p_i(x)$ correspon a la funció Normal $N(\mu, \Sigma)$ d'una distribució Gaussiana caracteritzada per la seva mitjana i variància:

$$\begin{aligned} p(x) &= \omega_1 N(x | \mu_1, \Sigma_1) + \omega_2 N(x | \mu_2, \Sigma_2) + \dots + \omega_M N(x | \mu_M, \Sigma_M) = \\ &= \sum_{k=1}^M \omega_k N(x | \mu_k, \Sigma_k) \end{aligned}$$

on $x = [x_1 x_2 \dots x_p]^T$ és un vector aleatori de dimensió p , $N(x | \mu_k, \Sigma_k)$ són les components de densitat gaussiana, ω_k els pesos de les mescles i M el nombre de gaussians.

Cada component és una funció Normal Gaussiana de dimensió p d'acord a la següent expressió:

$$N(x | \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right\}$$

amb μ_k com vector de mitjana i Σ_k com matriu de covariància.

El model queda finalment parametritzat amb els vectors de mitjana, matrius de covariàncies i pesos de les mescles:

$$\Lambda = \{ \mu_k, \Sigma_k, \omega_k \}, \quad k = 1 \dots M$$

L'estimació d'aquests paràmetres hauria de ser tal que maximitzi el nombre d'identificacions en la fase d'entrenament. L'algorisme EM (*Expectation Maximization*) calcula mitjançant un nombre d'iteracions fixades en un llinard els millors valors.

GMM té una gran capacitat per a classificar mostres d'un gran nombre de classes. Per altra banda, el fet de què en el procés d'entrenament no hi hagi etiquetatge de les mostres, habilita el mètode GMM per a descobrir classes que puguin haver quedat ocultes.

2.4.2. k-NN (k-veí més proper)

El mètode k-NN (*k-Nearest Neighbors*) és un dels mètodes més senzills de classificació: és simple, potent i no requereix temps per construir el model d'entrenament. En canvi, el seu temps de computació per classificar o identificar és alt. Aquest mètode ha estat utilitzat en [10,14].

És un mètode basat en la distància que emmagatzema totes les mostres d'entrenament i posteriorment calcula la distància entre la mostra de test i cada una de les mostres d'entrenament. La classe de la mostra en la que la distància és mínima correspon al resultat de la classificació.

El mètode de classificació k-NN permet fixar k i la fórmula per calcular la distància entre la mostra de test i cada una les mostres d'entrenament.

El paràmetre k indica el nombre de “veïns més pròxims” que es demana com resultat de la classificació. L'elecció de k es realitza de forma experimental escollint aquell valor que doni millors resultats.

El mètode de càlcul de distància que s'acostuma a aplicar és la distància Euclidiana:

$$dist(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

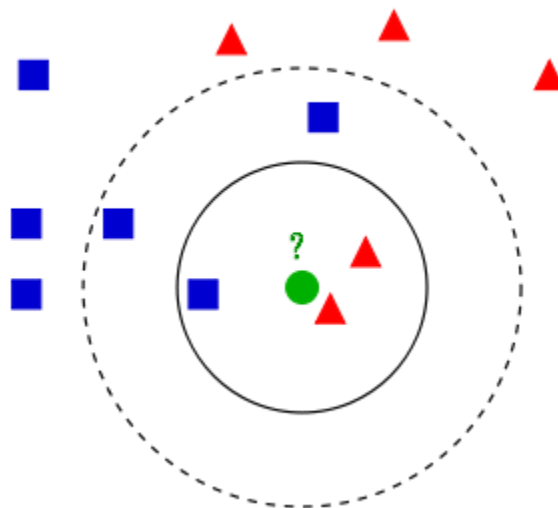


Figura 10. Esquematzació del procés de classificació amb k-NN (font: Wikipedia)

2.4.3. SVM (*Support Vector Machines*)

El mètode SVM es basa en la construcció d'un hiperplà n-dimensional en el que s'optimitzi la separació entre dades de dues classes. Aquest mètode ha estat utilitzat en [9,14,13].

Donat un conjunt de dades d'entrenament x_k , $k = 1, \dots, m$ i un vector de classes y tal que $y_k \in \{-1, +1\}$, $k = 1, \dots, m$, el mètode SVM busca un hiperplà $\omega \cdot x + b = 0$ que verifiqui

$$y_k(x_k \cdot \omega + b) - 1 \geq 0, \quad k = 1, \dots, m$$

La funció de decisió és:

$$f(x) = \text{sgn}(x \cdot \omega + b)$$

El mètode SVM fa servir unes funcions denominades *kernel* ($K(x_i, x_j)$) que permeten generalitzar l'aplicació del mètode per a quan les dades no son separables linealment. Alguns dels *kernels* més comuns són:

- lineal: $K(x_i, x_j) = x_i^T x_j$
- polinòmic: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- funció de base radial (RBF): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
- sigmoide: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

γ , r i d són els paràmetres del *kernel*. En el cas de la funció base radial, al paràmetre γ s'assigna el valor de $1/I$, on I equival al nombre de característiques.

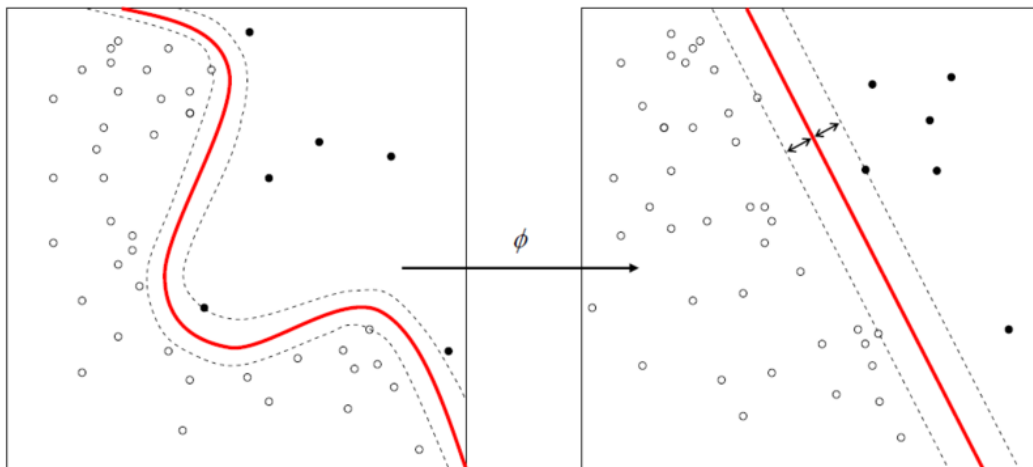


Figura 11 . exemple de SVM amb un kernel no lineal i un altre lineal (font: wikipedia)

SVM està pensat per a discernir entre observacions de dues classes. En un escenari de més de dues classes, existeixen implementacions multi-classe.

El procés d'ajustament per a obtenir els millors resultats en el mètode SVM no és simple. L'elecció del *kernel* i paràmetres més apropiats en cada cas és una tasca complexa que requereix molts assajos.

2.5. EL PROBLEMA DE LA IDENTIFICACIO

El reconeixement automàtic es planteja des de tres nivells de complexitat diferents: tons aïllats, fraseig monofònic i polifonia.

2.5.1. Tons aïllats

Al tractar-se del cas més simple, és sobre el que més treballs de recerca i experiments s'han realitzat arribant, alguns d'ells a nivells d'encert pròxims, en alguns casos als dels humans.

Estudi / Any	Percentatge d'encerts	Nombre d'instruments
Kaminsky / 1995	98	4
Kaminsky / 2000	82	19
Fujinaga / 1998	50	23
Frsser / 1999	64	23
Fujinaga / 2000	68	23
Martin / 1998	93	14
Kostek / 1999	97	4
	81	20
Kostek / 2001	93	4
	90	18

Figura 12. Resum de percentatge d'encert de tons aïllats (Font: [7])

2.5.2. Fraseig monofònic

Aquest tipus de problema es pot replantejar com múltiples casos de tons aïllats obtinguts a partir de finestres temporals del fraseig monofònic.

Experiments amb 16-18 instruments orquestrals en [7,8] van donar una taxa d'encerts de més del 90% per a família i d'un 70% per a instruments aplicant el mètode de classificació k-NN en els nodes d'una taxonomia semblant a la presentada en aquest treball.

Per l'altra costat, la classificació d'instruments utilitzant GMM en [7] no arriba a una taxa del 46% en el millors dels casos. En [13], utilitzant SVM en un estudi amb 10 instruments, la taxa d'encerts oscil·la entre el 62 i el 92 % en funció de l'instrument.

Així mateix, en [15] es va arribar, en un estudi amb 4 instruments, a més del 94% d'encerts en la identificació de l'instrument utilitzant SVM com mètode de classificació i els coeficients MFCC.

En [9] es va enfrontar el problema entrenant el sistema amb característiques melòdiques juntament amb SVM assolint un nivell d'encerts entre el 66 i el 83 % amb un total de 6 instruments orquestrals.

Els millors resultats s'obtenen en [14] on s'arriba una taxa d'encerts de més del 92 % amb 27 instruments i k-NN com a mètode de classificació.

2.5.3. Polifonia

Es tracta del cas més complicat i sobre el que encara està molt per fer. El problema principal associat al reconeixement d'instruments a partir d'una informació polifònica és que els harmònics i la resta de característiques espectrals i temporals associades a cada un dels instrument es barregen i són molt difícils (si possible) aïllar. Els estudis per a resoldre aquest repte estan, de moment, basats en un nombre molt reduït d'instruments simultanis.

Un enfocament del problema és a través de l'anàlisi harmònic del senyal polifònic per trobar els tons fonamentals dels instruments que sonen en simultani i fer un anàlisi per separat de cada un d'ells. Experiments amb quatre instruments assolien una taxa d'encert prop del 70% [12].

Un altre enfocament és utilitzar les mateixes eines d'anàlisi que es fan servir per a tons aïllats. En [11] aconseguixen un percentatge d'encerts entre el 67% i 84% amb 3 instruments utilitzant GMM com a mètode de classificació. No obstant, en la fase d'entrenament es va afegir també les característiques de gravacions on sonaven els instruments en parelles.

CAPÍTOL 3

DISSENY I IMPLEMENTACIO

3.1. INTRODUCCIO

L'objectiu fixat en aquest TFC és estudiar el comportament d'un conjunt de paràmetres acústics i tècniques de classificació en el reconeixement automàtic d'instruments musicals. Aquest objectiu, posa de manifest que el sistema ha de permetre, per un costat, extreure les característiques acústiques de fonts d'àudio i, per l'altra costat, capacitar la parametrització dels arguments que entren en joc, tant en el procés d'entrenament com en el procés de reconeixement, per a, finalment, fer avaluacions comparatives dels resultats.

Per una altra part, com que el corpus creat per a aquest TFC a partir d'arxius d'àudio és molt limitat en el nombre de mostres, s'ha optat per incorporar de forma addicional un corpus ja existent i molt extens de paràmetres acústics extrets d'un total de 19 instruments musicals. Aquest corpus, encara que no es pot barrejar amb el de creació pròpia ni fer-ho servir com a model en el procés de reconeixement d'aquests, és útil per fer avaluacions dins del mateix corpus amb l'objectiu d'obtenir resultats estadístics sobre mostres molt grans.

El capítol està organitzat en tres gran apartats. Primerament, es fa una presentació de l'entorn de desenvolupament, llenguatge de programació i llibreries escollides per a la implementació del sistema. En segon lloc, es mostra el disseny general del sistema per a continuació detallar les funcionalitats i característiques de cadascú dels mòduls que el componen. Posteriorment, es descriu el contingut i procedència dels dos corpus amb els que es treballa. El capítol finalitza amb una descripció del mode d'operació del sistema.

3.2. ENTORN DE DESENVOLUPAMENT

La plataforma de desenvolupament del sistema és MATLAB. Concretament, s'ha fet servir la versió 7.12 (R2011a) “*Student Version*”. La raó fonamental de triar MATLAB és que es tracta d'un entorn idoni pel tractament de dades en forma de matrius i que incorpora un gran nombre de funcions matemàtiques orientades a molts variats àrees d'aplicació juntament amb eines de representació gràfica. Un altra factor molt important, és que permet ampliar la seva capacitat a través de llibreries denominades *toolboxes* (caixes d'eines) amb funcions específiques en el desenvolupament en temes concrets. En aquest sentit, es fan servir els següents *toolboxes*:

- “Signal processing Toolbox”: *Toolbox* opcional de MATLAB que aporta algorismes pel tractament digital i analògic del senyal, com visualitzar senyals en els dominis espacial i temporal, calcular FFTs, dissenyar filtres, etc.
- “Statistics Toolbox”: *Toolbox* que s'utilitza pel tractament estadístic, del qual es fa servir les funcions per realitzar classificacions segons els mètodes k-NN i GMM.
- “MIRtoolbox”: *Toolbox* de lliure distribució desenvolupat pel Centre Interdisciplinari d'investigació de música de la Universitat de Jyväskylä (Finlàndia). Integra un conjunt de funcions dedicades a l'extracció de característiques musicals d'arxius d'àudio.
- “Auditory Toolbox”: *Toolbox* de lliure distribució desenvolupat per Malcom Slaney (1998) que MIRToolbox fa servir internament per a, entre altres finalitats, calcular els coeficients MFCC.
- “LIBSVM”: Mòdul externs desenvolupat en llenguatge C per realitzar la classificació segons el mètode SVM multi-classe.

3.3. DIAGRAMA DE BLOCS

El sistema de reconeixement segueix el següent esquema:

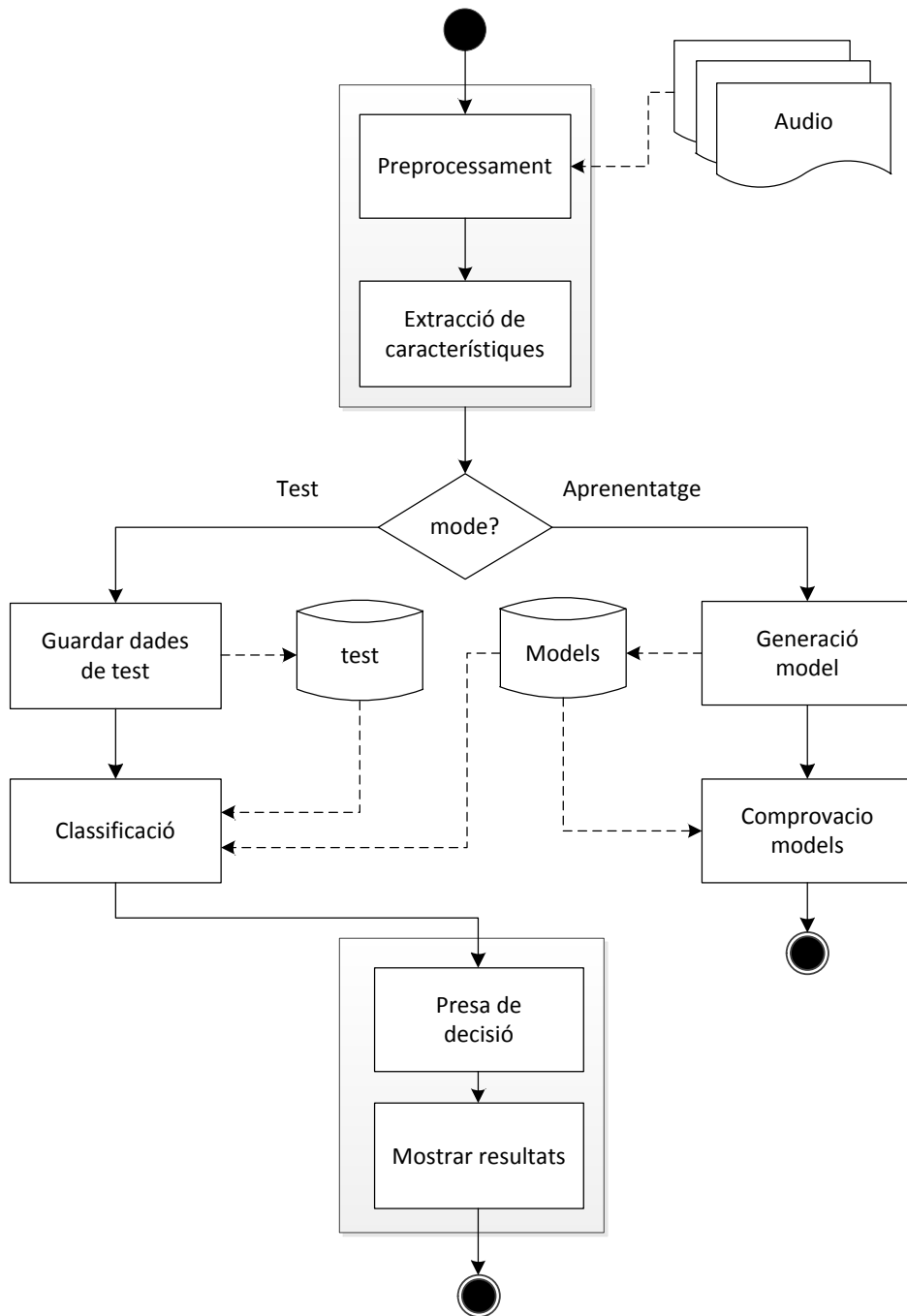


Figura 13. Diagrama general del sistema

L'esquema mostra com l'àudio entrant passa por un procés previ abans de treure les seves característiques. Depenent del mode d'operació escollit, aquestes característiques es fan servir per a la creació de model o bé per a la classificació o prova. Els mòduls de preprocessament i de classificació permeten la seva parametrització per tal de poder analitzar els resultats en funció d'aquests.

3.3.1. Entrada d'àudio

La informació d'àudio d'entrada s'introdueix al sistema a través d'un arxiu de text en el que figuren els arxius d'àudio a tractar juntament amb unes etiquetes de classificació segons descrit en el capítol 2 estat de l'art.

Els instruments que es tenen en compte són els següents:

Instrument	Família	Subfamília
acordio	bufat	lliure_llengueta
clarinet	bufat	llengueta_simple
corn_angles	bufat	doble_llengueta
contrabaix	corda	fregat
fagot	bufat	doble_llengueta
flauta	bufat	bisell
guitarra	corda	puntejat
oboe	bufat	doble_llengueta
piano	corda	colpejat
piccolo	bufat	bisell
saxofon	bufat	llengueta_simple
trombo	bufat	vibracio_llavis
trompa	bufat	vibracio_llavis
trompeta	bufat	vibracio_llavis
tuba	bufat	vibracio_llavis
viola	corda	fregat
violí	corda	fregat
violoncel	corda	fregat

Taula 3. Llistat dels instruments musicals que reconeixerà el sistema

Un exemple d'entrada al sistema seria el contingut del següent arxiu anomenat input.txt:

arxiu	familia	subfamilia	instrument
arxiu01.wav	corda	puntejat	guitarra
arxiu02.wav	corda	fregat	violí
arxiu03.wav	bufat	bisell	flauta
arxiu04.wav	bufat	vibracio_llavis	trompeta
arxiu05.wav	bufat	vibracio_llavis	trombo

3.3.2. Preprocessament i extracció de característiques

En aquesta fase es preparen els arxius d'àudio i, tot seguit, s'extreuen les característiques musicals (totes elles descrites en el capítol 2.3.) següents:

- Centroide espectral
- Extensió
- Coeficients 1-13 MFCC
- *Roll-off*
- *Zero-crossing*
- Energia

L'algorisme que segueix aquesta primera etapa del sistema és el següent:

```
Llegir_arxiu_amb_llistat_arxius_audio_i_etiquetes
PER cada arxiu de la llista FER
    SI arxiu NO es monocanal
        convertir a monocanal
    FSI
    frames = obtenir_posicion_frames[arxiu, windowSize]
    PER cada frame FER
        SI energia[frame] > 0.06 LLAVORS
            Feature[1] = posicio_en_llista_arxiu
            Feature[2] = centroide(frame)
            Feature[3] = extensio(frame)
            Feature[4..16] = mfcc1..mfcc13(frame)
            Feature[17] = rolloff(frame)
```

```

Feature[18] = zerocrossing(frame)
Feature[19] = enrgia(frame)
Feature[20] = familia (arxiu)
Feature[21] = subfamilia (arxiu)
Feature[22] = instrument (arxiu)

```

FSI

FPER

FPER

En el procés de fragmentació de l'àudio, la FFT es calcula amb una finestra de Hamming i amb una amplada, per defecte, de 40 ms amb un solapament del 50%.

A l'acabament d'aquest procés, els arxius d'àudio estan fragmentats en *frames* i cada *frame* amb els valors de les seves característiques acústiques.

La següent taula llista les funcions desenvolupades per aquesta fase:

Funció	Descripció
cAudioList = TFCLoadAudioList (filename)	Llegeix l'arxiu passat per paràmetre i posa el seu contingut en la matriu de cel·les cAudioList
cFeatures = TFCEXtractFeatures (cAudioList, windowSize)	Retorna una matriu de cel·les amb totes les característiques extretes del arxiu d'àudio (separats per <i>frames</i>) continguts en cAudioList amb una grandària de finestra <i>windowSize</i> .

Taula 4. Funcions TFCLoadAudioList i TFCGetFeatures

3.3.3. Guardar dades d'entrenament o test

Les característiques que s'han obtingut en la fase anterior es guarden en un arxiu per a no haver de tornar a processar de nou els arxius originals.

Els arxius que mantenen les dades d'entrenament o les dades de test tenen el mateix format. Són uns arxius de text, on els camps estan separats per tabuladors. La primera fila conté els títols dels camps i la resta de files els valors per a cada camp. La primera columna (índex) es una referència a la fila en l'arxiu original de llista d'arxius d'àudio.

Un exemple corresponent als dos primers *frames* processats del primer arxiu que hi consta en la llista mostra de l'apartat 3.3.1. és el següent:

Columnes 1-6:

Index	centroide	extensio	mfcc1	mfcc2	mfcc3
1	994.4960	52.9342	4.0048	0.1137	0.7588
1	1072.9300	34.7494	3.7364	-0.0585	0.6528

Columnes 7-12:

mfcc4	mfcc5	mfcc6	mfcc7	mfcc8	mfcc9
-0.2043	0.2832	0.1976	0.3577	0.0027	0.0263
-0.2178	0.2312	0.0394	0.5600	0.1511	0.1322

Columnes 13-18:

mfcc10	mfcc11	mfcc12	mfcc13	rolloff	zerocross
-0.9378	-0.7971	-0.2765	0.0644	1735.12	570.26
-0.9428	-1.0336	-0.5030	-0.2653	1767.45	770.34

Columnes 19-22:

energia	familia	subfamilia	instrument
0,1474	corda	puntejat	guitarra
0.2328	corda	puntejat	guitarra

La següent taula llista les funcions desenvolupades per aquesta fase:

Funció	Descripció
TFCSaveFeatures (filename, cFeatures)	Guarda en l'arxiu indicat per <i>filename</i> la matriu de cel·les (<i>cFeatures</i>) amb les característiques.
TFCEXportFeatures (audioListFile, featuresFile, windowSize)	A partir de l'arxiu <i>audioListFile</i> amb una llista d'arxius d'àudio, crea l'arxiu <i>featuresFile</i> amb les seves característiques. La grandària del <i>frame</i> està determinat pel paràmetre <i>windowSize</i> .

Taula 5. Funcions TFCSaveFeatures i TFCCreateFeatures

3.3.4. Classificació

Aquest és el mòdul que ofereix més alternatives. S'encarrega de classificar un conjunt de característiques de les mostres de test a partir de les dades d'entrenament i segons diferents mètodes de classificació. El següent esquema mostra les fases que tenen lloc en aquest mòdul:

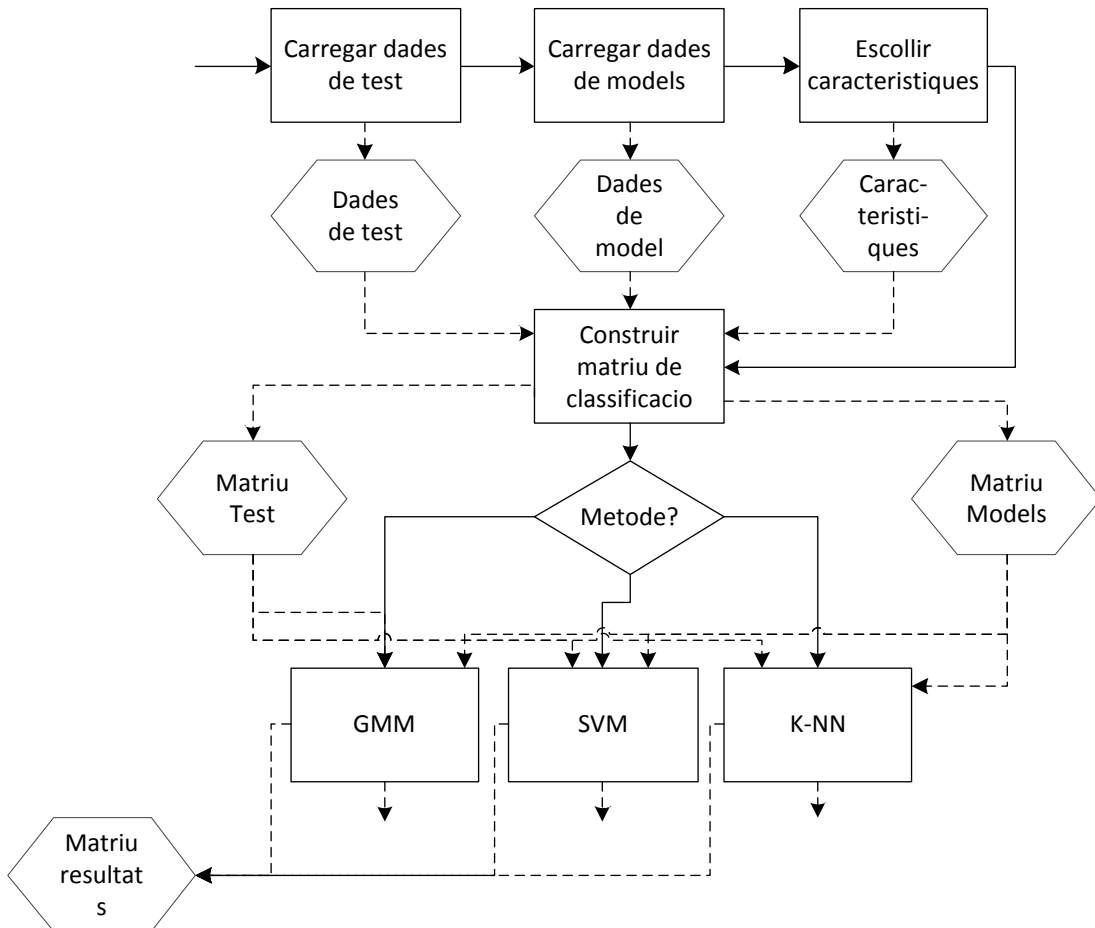


Figura 14. Diagrama del mòdul de classificació

En primer lloc, cal carregar les característiques d'àudio tant del model com del conjunt que es fa servir com test. Les dades carregades tenen el format descrit en l'apartat anterior. Posteriorment, s'ha d'escollir el subconjunt del total de les 18 característiques possibles que es vol fer servir per l'estudi. Amb aquesta finalitat, es construeix un vector de dimensió 1x18 en el que figura un '1' o un '0' segons es vulgui considerar la característica corresponent d'acord a la següent taula:

Posició	Característica
1	centroide espectral
2	extensio
3	MFCC1
4	MFCC2
5	MFCC3
6	MFCC4
7	MFCC5
8	MFCC6
9	MFCC7
10	MFCC8
11	MFCC9
12	MFCC10
13	MFCC11
14	MFCC12
15	MFCC13
16	<i>Roll-off</i>
17	<i>Zero-crossing</i>
18	energia

Taula 6. Descripció del vector de selecció de característiques

A partir de la informació proporcionada per aquest vector i les matrius de característiques model i test, es creen dos noves matrius que contenen exclusivament aquells valors escollits per a aplicar qualsevol dels tres mètodes classificació següents: *Gaussian Mixture Model (GMM)*, *k-Nearest Neighbor (k-NN)* o *Support Vector Machines (SVM)*.

Amb els tres casos, es parametriza:

- de manera opcional, però aconsellada, un escalat dels valors de les característiques en el rang $[-1, +1]$ per tal d'evitar els efectes que característiques mesurades en escales diferents tindrien sobre el resultat.
- Tipus d'atribut que es vol classificar: instrument, família o subfamília.

Les funcions encarregades de preparar les dades per la classificació són les següents:

Funció	Descripció
<code>cFeatures = TFCLoadFeatures (file]name)</code>	Carrega les característiques de l'arxiu a memòria.
<code>[mTraining mTest] = TFCCreateCheckMatrix(cFeaturesTraining, cFeaturesTest, vWhat)</code>	A partir de les matrius amb totes les característiques d'entrenament i de test, crea un nou parell de matrius amb els subconjunt de característiques determinades pel vector <i>vWhat</i> .
<code>[mA1 mB1] = TFCScaleMatrices (mA, mB)</code>	Donat dos matrius, <i>mA</i> i <i>mB</i> , genera un parell de noves matrius <i>mA1</i> i <i>mB1</i> en el que el rang dels valors de totes les seves cel·les estan entre -1 i +1.

Taula 7. Funcions `TFCLoadFeatures`, `TFCCreateCheckMatrices` i `TFCScaleMatrices`

3.3.4.1. Classificació per Model de Mescles Gaussians (GMM)

Per l'aplicació d'aquest mecanisme de classificació, es fa servir les funcions que incorpora MATLAB amb la creació inicial del model amb l'algorisme de maximització de la expectativa (EM, *Expectation Maximization*) i amb els valors que pren per defecte.

El nombre de classes queda determinat pel nombre d'elements diferents que hi ha en el model segons els tipus d'atribut pel que es vol realitzar la classificació (família, subfamília, instruments).

Per poder quantificar el percentatge d'incert tant del model en sí (GMM és un mètode estadístic de classificació que no té en compte etiquetatges previs) com de les mostres de dades sobre el model, es fa servir el següent procediment:


```

Total_classes = calcular_nombre_de_clases_model()
% generar matriu nxm amb classes x valors del model i de la classes
valors_clases = valors_x_clases()
PER i=1 FINS Total_classes FER
    % Crear model GMM per a la classe 'i' segons els seus valors
    model[i] = crear_model_gmm (valors_de_classe(i))
FPER
% Calcular el valor de la funcio de densitat de probabilita per a cada
% valor de la matriu original segons cada un dels model GMM creats per
% cada classe
PER i=1 FINS Total_mostres_test FER
    PER j=1 FINS total_clases FER
        fdp[i][j] = calcular_fdp(model[j], i)
    FPER
FPER
% Per a cada mostra, el valor de la fdp més gran correpondrà a la
% classe escollida per el mètode GMM
PER i=1 FINS total_mostres_test FER
    classe_resultat[i]= j tq fdp[i][j] és màxim 1<=j<=total_clases
FPER

```

La següent taula llista les funcions desenvolupades per aquesta fase:

Funció	Descripció
cResults = TFCGMMCheck (cFeaturesTraining, vWhat, scale, classType)	Realitza la classificació per mètode GMM del propi conjunt segons les característiques escollides per <i>vWhat</i> i deixa el resultat en la matriu <i>cResults</i> . Aquesta matriu conté la classe resultant i el valor de la fdp (funció de densitat de probabilitat). El paràmetre <i>scale</i> indica si es vol escalar les dades (1) o no (0). El paràmetre <i>classType</i> indica si es vol classificar per família, subfamília o instrument (0, 1 i 2 respectivament).
cResults =TFCGMMClassify (cFeaturesTraining, cFeaturesTest, vWhat, scale, classType)	Classifica els elements de <i>cFeaturesTest</i> segons el mètode GMM prenent com model les dades de <i>cFeaturesTraining</i> . <i>vWhat</i> especifica el subconjunt de característiques que es vol considerar. <i>Scale</i> si es vol escalar les dades en el rang [-1, +1]. <i>classType</i> indica si es vol classificar per família, subfamília o instrument (0, 1 i 2 respectivament).

Taula 8. Funcions TFCGMMCheck i TFCGMMClassify

3.3.4.2. Classificació per k-veí més proper (k-NN)

Aquest altre mètode de classificació fa servir també les funcions proporcionades per MATLAB (*knnsearch*). L'algorisme de distància aplicat serà el de distància Euclidiana. En aquest mètode es fan servir dos paràmetres addicionals:

- **k:** Indica el nombre de veïns més pròxims que es vol considerar en la classificació.
- **maxDist:** indica la distància màxima en què un k-veí és acceptat.

En el cas de què $k = 1$, l'acceptació de la classe retornada dependrà del valor del límit admès de distància. En el cas de $k > 1$, l'escenari és més complex ja que les classes resultants poden ser diferents entre elles i l'algorisme ha de prendre una decisió. S'ha optat per oferir les següents opcions:

- Si només hi ha una classe predominant entre les classes dels k-veïns, aquesta classe és l'escollida com a solució. En cas contrari el resultat és de incertesa.
- Si ha més d'una classe predominant, escollir qualsevol d'elles.
- Altre alternativa que es defineixi segons els resultats dels experiments.

Les funcions desenvolupades són les següents:

Funció	Descripció
<code>cResults = TFCKNNClassify</code> (<code>cFeaturesTraining</code> , <code>cFeaturesTest</code> , <code>vWhat</code> , <code>scale</code> , <code>classType</code> , <code>k</code> , <code>maxDist</code>)	Classifica els elements de <i>cFeaturesTest</i> segons el mètode k-NN prenent com model les dades de <i>FeaturesTraining</i> . <i>vWhat</i> especifica el subconjunt de característiques que es vol considerar. <i>Scale</i> especifica si es vol escalar les dades en el rang [-1, +1]. <i>classType</i> indica si es vol classificar per família, subfamília o instrument (0, 1 i 2 respectivament). <i>K</i> indica el nombre de veïns més pròxims que cerca i <i>maxDist</i> és la distància màxima acceptada. Aquesta versió retorna la classe més predominant quan $k > 1$. Si no existeix una única classe predominant llavors retorna la classe 'Incert'.

<code>cResults = TFCKNNClassify2 (cFeaturesTraining, cFeaturesTest, vWhat, scale, classType, k , maxDist)</code>	Classifica els elements de <i>cFeaturesTest</i> segons el mètode k-NN prenent com model les dades de <i>FeaturesTraining</i> . Els paràmetres <i>vWhat</i> , <i>scale</i> , <i>classType</i> , <i>k</i> i <i>maxDist</i> tenen el mateix significat que en <i>TFCKNNClassify</i> . Aquesta versió no requereix que la classe més predominant en el resultat quan $k > 1$ sigui única: tria qualsevol d'elles.
--	---

Taula 9. Funcions *TFCKNNClassify* i *TFCKNNClassify2*

3.3.4.3. Classificació per SVM (*Support Vector Machine*)

Per la implementació d'aquest mètode de classificació es fa servir la llibreria externa LIBSVM. Es una llibreria que permet la classificació multi-classe. El procediment amb aquesta llibreria externa és el següent:

```
Gravar_matriu_test_en_format_LIBSVM();
Gravar_matriu_Model_en_format_LIBSVM();
Escalar_dades_model();
Escalar_dades_test_segons_rangs_model();
Procedir_Aprenetatge();
Procedir_test();
Convertir_resultats_a_format_sistema();
```

Encara que es poden parametritzar molts atributs, aquests s'han de parametritzar directament en el codi de Matlab. Per defecte, es fa servir un *Kernel* de tipus *Gaussian Radial Function* (RBF) amb un valor de γ (veure capítol 2.4.3.) igual a $1/\text{característiques}$.

La funció desenvolupada és la següent:

Funció	Descripció
<code>cResults =TFCSVMClassify (cFeaturesTraining, cFeaturesTest, vWhat, classType)</code>	Classifica els elements de <i>cFeaturesTest</i> segons el mètode SVM prenent com model les dades de <i>FeaturesTraining</i> . <i>vWhat</i> especifica el subconjunt de característiques que es vol considerar. <i>classType</i> indica si es vol classificar per família, subfamília o instrument (0, 1 i 2 respectivament).

Taula 10. Funció *TFCSVMClassify*

3.3.5. Presa de decisió i mostra de resultats

Aquest mòdul s'encarrega, després de l'aplicació dels mètodes de classificació executats, de la preparació de les dades resultants i del posterior anàlisi per a la presa de decisió i mostra de resultats. L'esquema d'aquest mòdul és el següent:

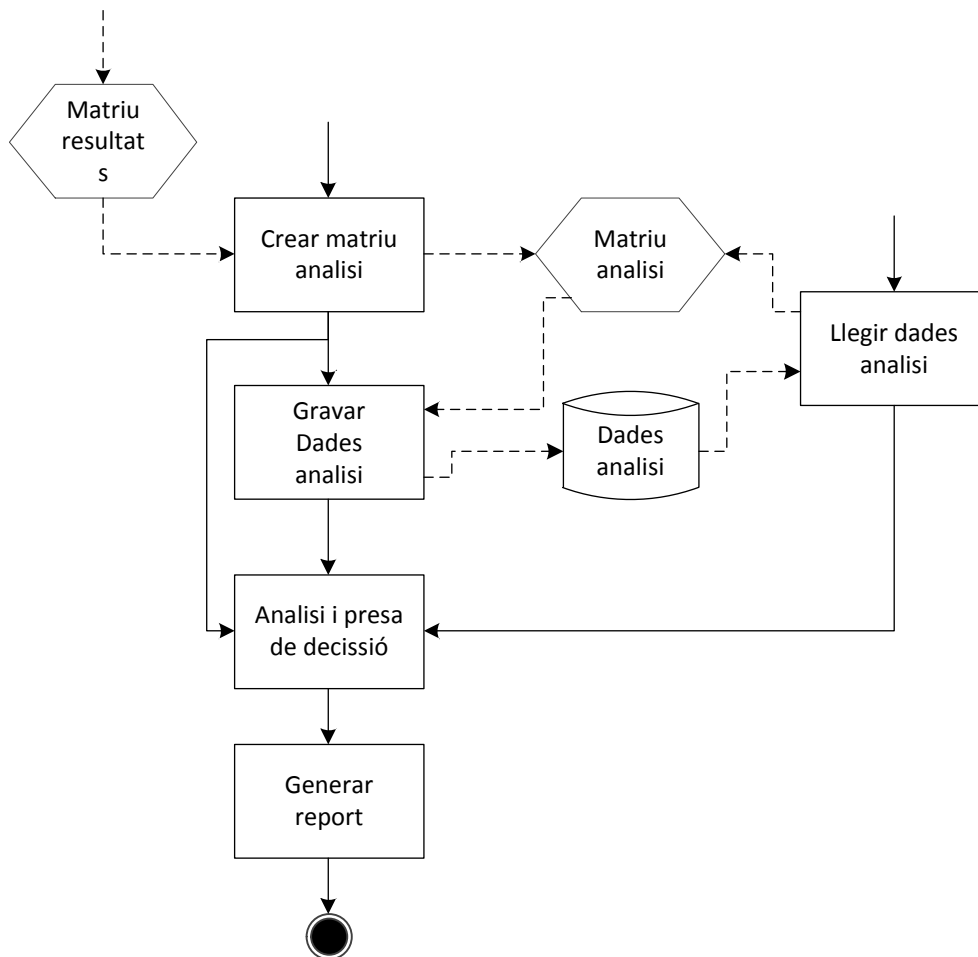


Figura 15. Diagrama del mòdul de presa de decisió i mostra de resultats

En una primera fase, es crea la matriu d'anàlisi a partir de la matriu de resultats. Aquesta matriu té tantes files com nombre de *frames* processats i conté la següent informació:

Columna	Descripció
1	Referència numèrica a la mostra de test. Aquest valor coincideix a l'índex d'arxiu dins de l'arxiu que conté la llista d'arxius a processar.

2	Nom de l'instrument que figura en la mostra de test.
3	Nom de la classe tal com figura en la mostra de test. Aquest valor pot ser el mateix nom d'instrument o la seva família o subfamília.
4	Nom de la classe resultat del mètode de classificació.
5	Número que correspon a la distància del k-veí en el cas d'haver utilitzat la classificació amb el mètode k-NN o el valor de la fdp en cas d'haver fet servir el mètode GMM. Si s'ha fet servir el mètode SVM aquest valor queda indeterminat i assignat a 0.

Taula 11. Descripció dels camps de la matriu d'anàlisi

La matriu es pot guardar en disc per a una posterior manipulació amb programari extern o, simplement, pel seu tractament en una sessió posterior. Un exemple del contingut de l'arxiu seria el següent:

```

Index Instrument      EtiquetaTest      EtiquetaClassificador      Value
1      fagot      doble_llengueta      puntejat      6.253027e+000
1      fagot      doble_llengueta      doblellequeta      6.901062e+000
1      fagot      doble_llengueta      doble_llengueta      6.901179e+000
1      fagot      doble_llengueta      doble_llengueta      6.076298e+000
1      fagot      doble_llengueta      doble_llengueta      6.584116e+000
1      fagot      doble_llengueta      doble_llengueta      6.084866e+000

```

L'exemple mostraria el resultat de la classificació segons subfamília dels primers sis *frames* de l'arxiu d'àudio referenciat per l'índex 1.

La següent taula llista les funcions desenvolupades per aquesta fase:

Funció	Descripció
cAnalytics = TFCCreateAnalytics (cFeaturesTest, cResults, classType)	Crea la matriu per anàlisi de dades a partir de la matriu de resultats del procés de classificació.
TFCSaveAnalytics (fname, cAnalytics)	Guarda en disc la matriu de dades d'anàlisi.
cAnalytics = TFCLoadAnalytics (fname)	Carrega a memòria les dades d'anàlisi gravades en disc

Taula 12. Funcions TFCCreateAnalytics, TFCSaveAnalytics i TFCLoadAnalytics

En la següent fase, el mòdul ofereix dos alternatives de resum i mostra de resultats: resum sobre el total de les mostres i sobre la llista d'arxius processats.

3.3.5.1. Resum sobre el total de mostres

Aquesta opció considera que els *frames* són mostres independents entre elles sobre les que no es té informació relacionada amb els arxius d'àudio dels quals provenen. Sota aquesta consideració, s'agrupen els resultats donant en nombre d'encerts i errades assolides.

Un exemple és el següent:

Instrument	Encerts	Errades	Total
fagot	45	10	55
acordio	34	6	40
tuba	74	12	86
flauta	33	41	74
Totals	186	69	255

Les funcions implementades per a assolir aquesta funcionalitat són les següents:

Funció	Descripció
cReport = TFCReportAnalytics (cAnalytics)	Genera informe resum a partir de la matriu d-anàlisi <i>cAnalytics</i>
TFCSaveReportAnalytics (fname, cReport)	Guarda en disc el resultat <i>cReport</i> .

Taula 13. Funcions TFCReportAnalytics i TFCSaveReportAnalytics

3.3.5.2. Resum sobre la llista d'arxius d'àudio

Aquesta opció no considera els *frames* com unitats independents sinó com el resultat de la descomposició en fragments dels arxius d'àudio d'on provenen.

El resum té en compte els encerts de cadascú dels *frames* que formen cadascú dels arxius, té en compte les dos classes amb més ponderació i, a partir d'aquests valors, decideix el resultat final de la identificació.

Els camps de la matriu resultant són els següents:

Columna	Descripció
1	Nom d'arxiu
2	Nom d'instrument tal com figura en l'arxiu de test
3	Non de la família de l'instrument tal com figura en l'arxiu de test
4	Nom de la subfamília de l'instrument tal com figura en l'arxiu de test
5	Nombre de <i>frames</i> en què s'ha desglossat l'arxiu d'àudio.
6	Valor de la classe (instrument, família o subfamília) amb més ponderació segons la classificació de tots els seus <i>frames</i> .
7	Valor del percentatge de la classe sobre el total de <i>frames</i> .
8	Valor la segona classe (instrument, família o subfamília) amb més ponderació segons la classificació de tots els seus <i>frames</i> .
9	Valor del percentatge de la classe sobre el total de <i>frames</i> .
10	Nom de la classe decidit pel sistema en funció dels valors de les columnes 7 i 9.

Taula 14. Descripció de la taula resum de decisió

Les funcions implementades per assolir aquesta funcionalitat són les següents:

Funció	Descripció
cReport = TFCTReportDecision (cAnalytics)	Genera informe resum a partir de la matriu d'anàlisi <i>cAnalytics</i>
TFCTSaveReportDecision (fname, cReport)	Guarda en disc el resultat <i>cReport</i> .

Taula 15. Funcions TFCTReportDecision i TFCTSaveReportDecision

3.4. DESCRIPCIÓ DELS CORPUS

3.4.1. Corpus importat

Aquest corpus prové d'un concurs sobre mineria de dades de TunedIt Solutions en col·laboració amb el 19è Symposium de Metodologies per Systemes Intel·ligents (*ISMIS 2011 Contest: Music Information Retrieval*) (<http://tunedit.org/challenge/music-retrieval>).

El contingut del corpus del ISMIS 2011 està format per prop de 115.000 mostres de característiques musicals corresponents a 19 instruments musicals. Aquest corpus està sense arxius d'àudio i per tant no es pot aplicar parametritzacions en el moment d'obtenir les característiques acústiques.

Aquest conjunt està dividit en dos parts en la proporció 70%-30% on la part més gran es fa servir per a entrenament i l'altre part per a test.

La descripció del corpus juntament amb el tractament previ que es fa per compatibilitzar-lo amb el format descrit en l'apartat 3.3.3. és el següent:

Camp(s)	Acció
TemporalCentroid, LogAttack	Es descarten per motius de compatibilitat ja que es tracta de característiques de tipus temporal que no s'han abordat.
LogSpecCentroid, LogSpecSpread	Són dos característiques basades en l'estàndard MPEG7 (extretes després d'aplicar un filtre rectangular logarítmic en comptes d'un filtre triangular mel). Donen informació similar als camps Centroide i extensió i que també es contemplen en aquest corpus. Per tant, es descarten.
MFCC1-MFCC13	No es fa cap modificació
ZeroCrossings	No es fa cap modificació
SpecCentroid	No es fa cap modificació. És l'equivalent a 'Centroide'
SpecSpread	No es fa cap modificació. És l'equivalent a 'Extensió'
Rolloff	No es fa cap modificació. És l'equivalent a 'rolloff'

Flux	Es descarta per motius de compatibilitat ja que es tracta d'una característica temporal.
bandsCoef1-bandsCoefs33	Es descarten. Són l'equivalent als MFCC però en una escala logarítmica freqüencial dividida en 8 octaves.
bandsCoefSum	Es descarta. Es la suma dels coeficients anteriors.
prj1-prj33	Es descarten. Són els denominats <i>Audio Spectrum Projection</i> de l'estàndard MPEG-7. També donen informació del mateix tipus que els MFCC amb més temps de procés.
prjmin, prjmax, prjsum, prjdis, prjstd	Es descarten. Són valors estadístics corresponents als camps anteriors.
HamoPk1-HamoPk28	Són els valors en dB dels 28 primers pics harmònics. Es descarten perquè en el corpus moltes d'aquestes dades estan buides.
Energy	Equivalent a energia
Frameid	Identificador del fragment d'on s'ha tret l'àudio. No té sentit en aquest sistema i es descarta.
Note	Nota musical associada a la mostra. Es descarta.
Playmethod	Té els valors de <i>blown, string o struck</i> . Equival al camp família categoritzat com corda i bufat.
Class1	Pren els valors <i>aerophone o cordophone</i> . Es descarta perquè no aporta informació addicional ja que correspon al camp família.
Class2	Pren els valors de <i>free-reed, double-reed, rubbed_composite, single-reed, bevel, lip-vibrated, plucked, keyboard</i> . Equival al camp subfamília amb els següents corresponents valors: lliure_llengueta, doble_llengueta, fregat, llengueta_simple, bisell, vibracio_llavis, puntejat i colpejat.
Instrument:	Nom de l'instrument. Pren els valors <i>Accordion, Bassoon, Cello, Clarinet, DoubleBass, EnglishHorn, Flute, Frenchhorn, Guitar, Oboe, Piano, Piccolo, Saxophone, SynthBass, Trombone, Trumpet, Tuba, Viola, Violin</i> . El <i>SynthBass</i> no es té en compte i per tant les mostres correspondran als instruments acordio, fagot, violoncel, clarinet, contrabaix, corn_angles, flauta, trompa, guitarra, oboe, piano, piccolo, saxofon, trombo, trompeta, tuba, viola i violi.

Taula 16. Descripció dels camps del corpus importat de TunedIt

3.4.2. Corpus de generació pròpia

Aquest corpus ha estat generat a partir d'un teclat electrònic CASIO de gama baixa amb la capacitat d'emular sons de diferents instruments juntament amb el programari Adobe Audition pel procés d'enregistrament.

S'ha connectat la sortida d'àudio del teclat electrònic amb la entrada *line-in* de l'ordinador. Per obtenir les mostres d'aprenentatge, amb l'Adobe Audition s'han gravat 18 arxius (un per cada instrument) en el que s'ha enregistrat cada una de les notes que formen part de l'extensió de l'instrument amb una duració aproximada de dos segons i amb una separació entre nota i nota d'un segon aproximadament.

En el procés d'extracció de característiques, s'ha aplicat una finestra de 50 ms amb un solapament del 50%. A més, l'algorisme implementat descarta els silencis deixats entre nota i nota.

Per a la generació de les mostres de test, utilitzant el mateix teclat electrònic, s'han gravat arxius amb una seqüència aleatòria de notes corresponents a cada un dels instruments tractats. Tots dos enregistraments estan gravats en mono, amb una freqüència de mostreig de 44.000 Hz i 16 bits de resolució.

En comparació al corpus importat de TunedIt, el corpus creat és molt més reduït en nombre de mostres (24.700 d'entrenament i 4.700 de test), de menys qualitat i els resultats reflexaran aquesta particularitat.

Amb l'objectiu de comprovar el funcionament del sistema en un altre escenari, també s'han creat un altre corpus de test amb gravacions d'instruments aïllats però enregistrant més d'una nota a l'uníson.

3.4.3. Microsoft Musical Instruments

De manera addicional als corpus generats de forma electrònica, s'ha afegit un corpus amb gravacions d'extractes d'instruments musicals sonant aïlladament procedents d'una enciclopèdia multimèdia titulada Microsoft Musical Instruments (any 1992).

3.5. MODE D'OPERACIO

Encara que els dos corpus mantenen el mateix tipus de dades, no poden barrejar-se i s'ha de treballar independentment amb cadascú d'ells.

Així doncs, la fase de parametrització amb el corpus importat no existeix com tal ja que correspon directament al mateix corpus una vegada modificat segons les pautes descrites en l'apartat 3.4.1.

Les mostres model i test del corpus importat ja estan processats i només s'han de carregar en el sistema per parametritzar les classificacions i analitzar els resultats (apartats 3.3.4. i 3.3.5.).

Pel que fa al corpus propi, aquest sí que fa ús de tot el sistema en el que en la primera fase hi ha un processament de l'àudio per tal de extreure-hi les seves característiques musicals (apartat 3.3.2) i, a partir d'aquest punt, seguir el mateix procés descrit pel corpus importat.

CAPITOL 4

PROVES I EXPERIMENTS

4.1. INTRODUCCIO

L'objectiu dels experiments són analitzar com afecten les parametritzacions en el procés de classificació en el resultat final. Les parametritzacions venen donades per múltiples combinacions de variables tals com:

- Nombre i selecció de característiques acústiques escollides.
- Mètode de classificació utilitzat.
- Nombre de mostres.
- Tipus de classes presents.

Per altra banda, també és interessant investigar si hi ha algun grup d'instruments en particular que siguin més fàcils d'identificar en un mètode de classificació que en un altre. Per finalitzar, convé comprovar si els experiments amb els dos corpus tenen resultats en comú.

4.1. CORPUS IMPORTAT DE TUNEDIT

4.1.1. Rellevància de cada característica acústic

La primera sèrie d'experiments ha consistit en estudiar la rellevància, per separat, de les 18 característiques acústiques (centroide, extensió, 13 coeficients MFCC, *rolloff*, *zerocross* i energia) segons els tres mètodes de classificació estudiats. El coeficients MFCC s'han

considerat com un vector amb longitud creixent de manera que quan es considera, per exemple el cinquè coeficient, implícitament també es consideren els coeficients 1-4.

De cara a valorar els resultats, les proporcions d'encert aleatori són del 50% (família), 33% (subfamília) i 6% (instrument). Els resultats són els següents:

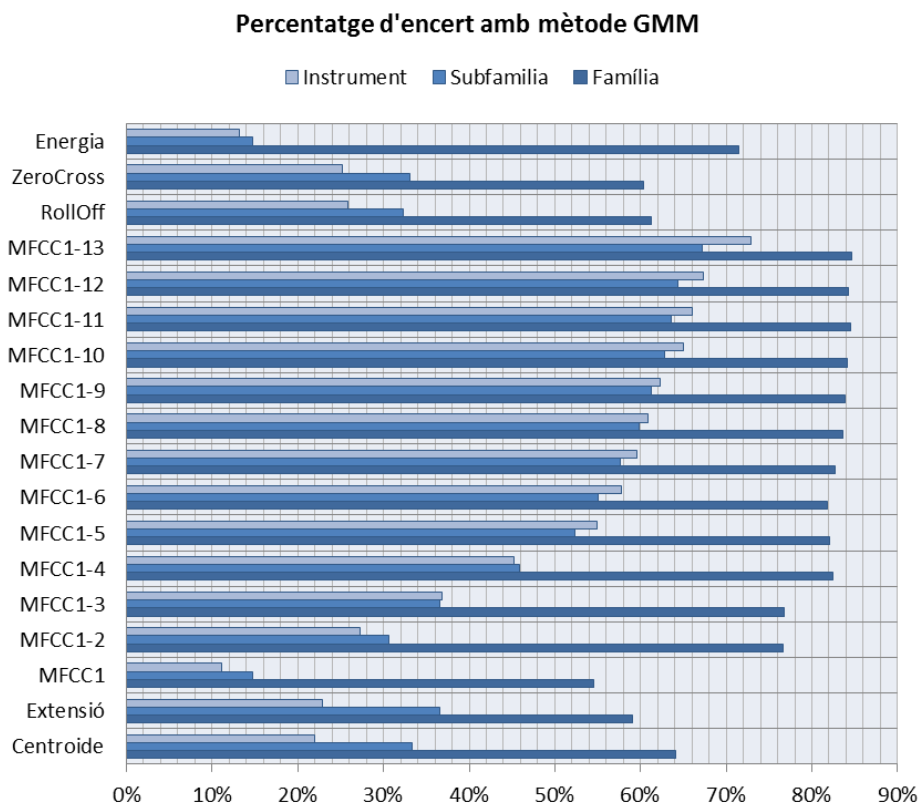


Figura 16. Percent. d'encerts de les característiques aïllades aplicant GMM

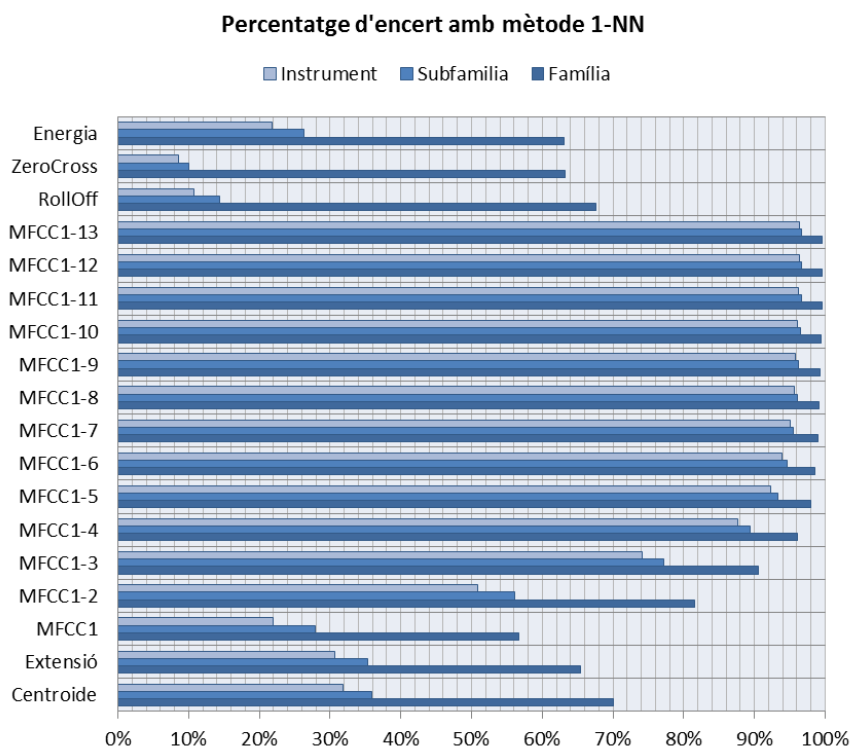


Figura 17. Percent. d'encerts de les característiques aïllades aplicant 1-NN

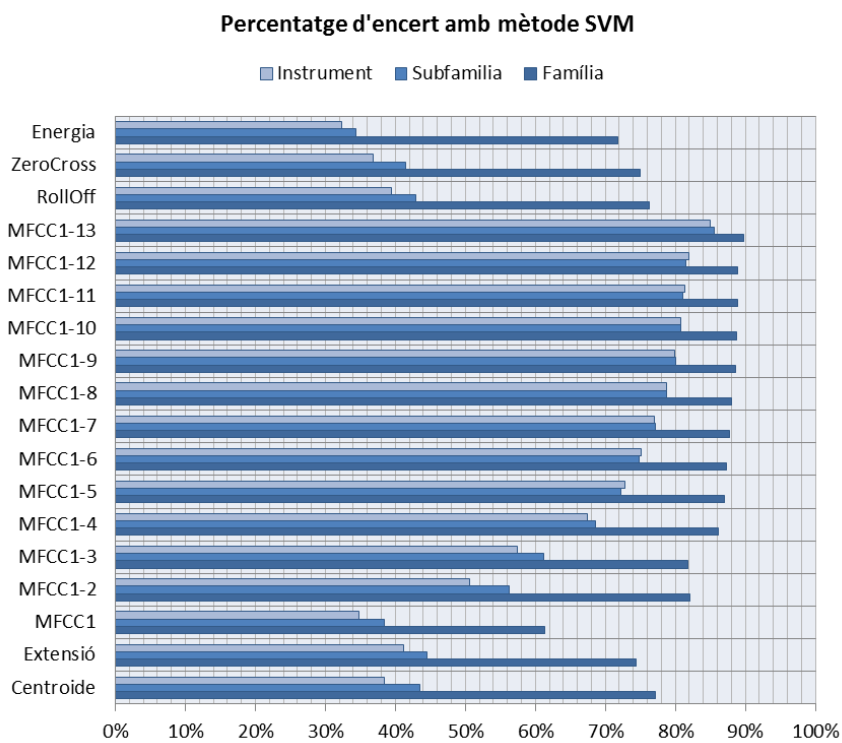


Figura 18. Percent. d'encerts de les característiques aïllades aplicant SVM

Els resultats indiquen que aplicant 1-NN s'aconsegueixen els millors resultats on arriba a un 96.43 % d'encerts en l'instrument utilitzant només els primers 13 coeficients MFCCs.

Per altra banda, característiques de menys pes com, per exemple, *zerocross* i *rolloff* donen un percentatge d'encert molt baix en 1-NN (8.55 % i 10.80 %) mentre que aplicant GMM o SVM el percentatge és molt superior (36.82 % i 39.40 % per a SVM).

4.1.2. Combinació de paràmetres acústics

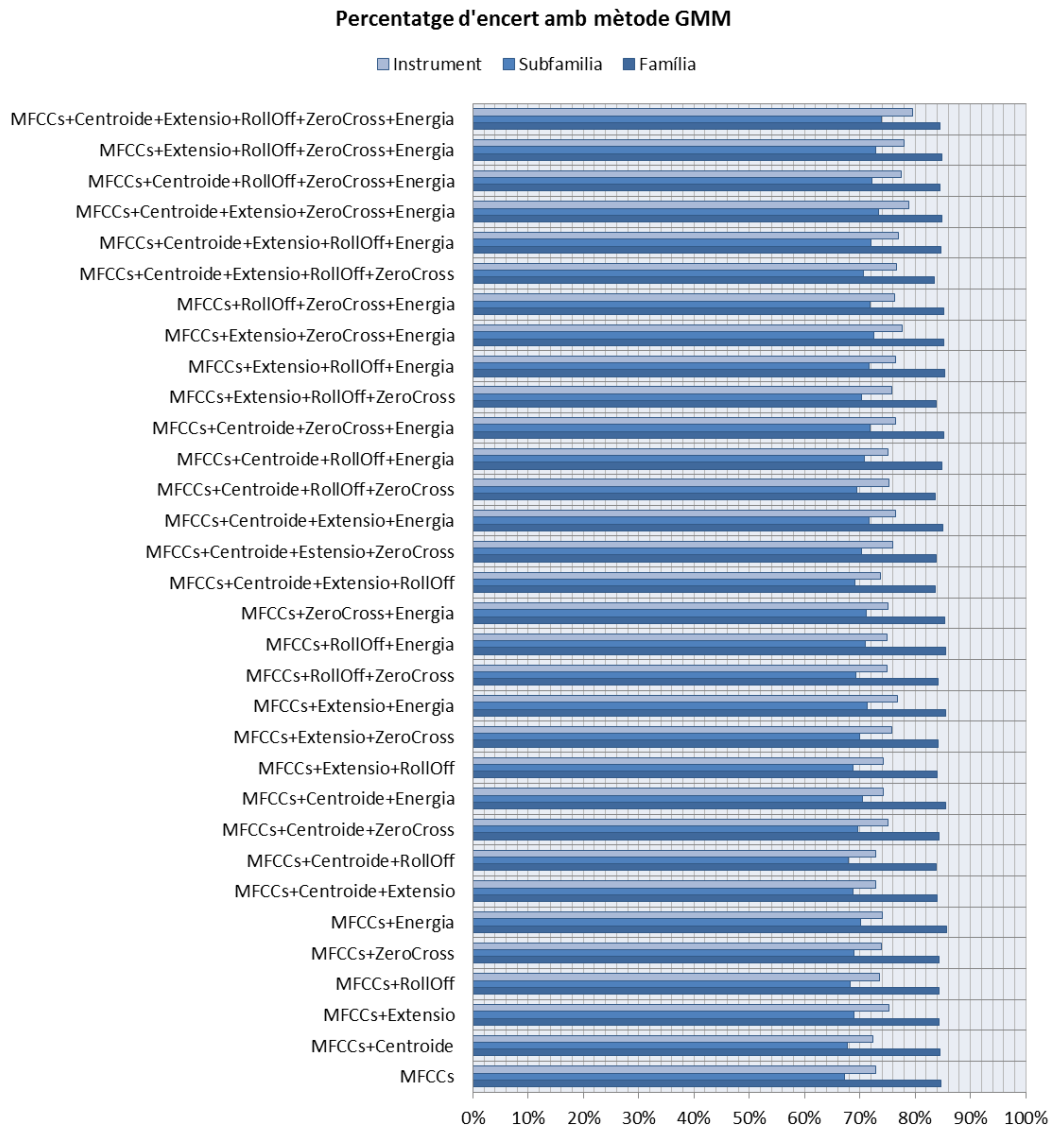


Figura 19. Percent. d'encerts amb combinació de característiques aplicant GMM

Percentatge d'encerts aplicant 1-NN

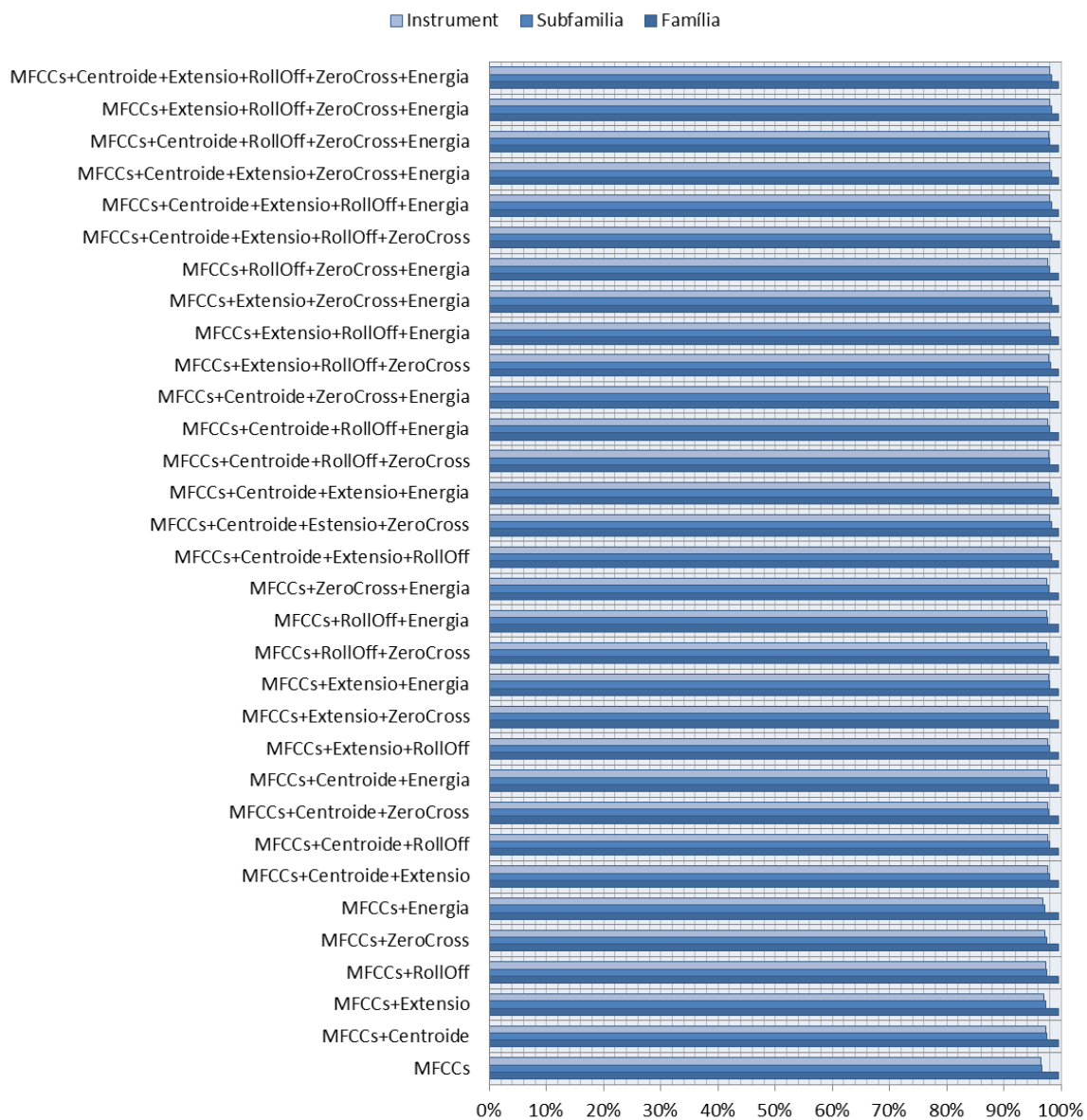


Figura 20. Percent. d'encerts amb combinació de característiques aplicant 1-NN

Percentatge d'encerts aplicant SVM

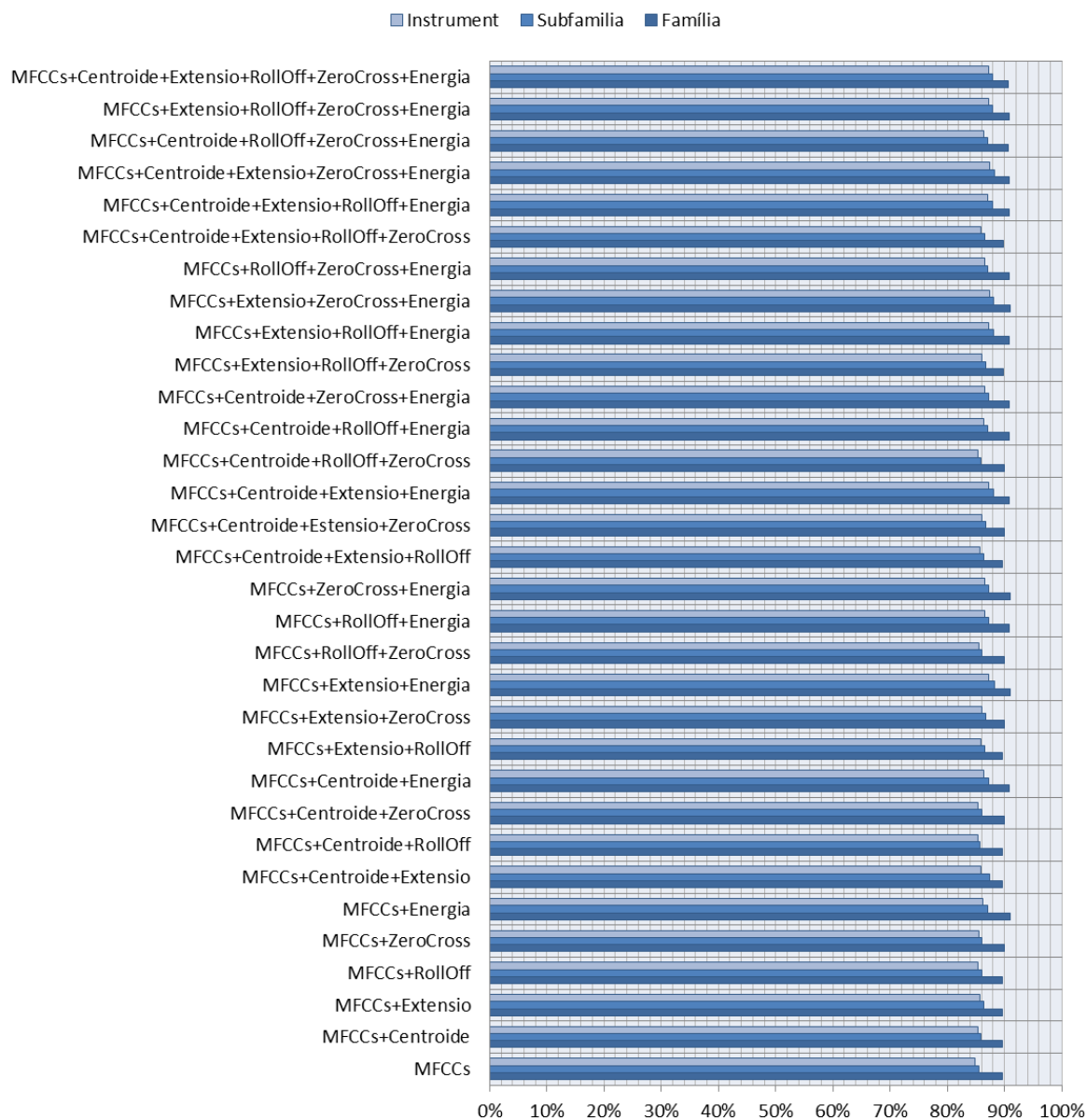


Figura 21. Percent. d'encerts amb combinació de característiques aplicant SVM

Vist que els coeficients MFCCs eren el que més percentatge tenien, l'objectiu d'aquests experiments han sigut veure si hi havia alguna combinació de característiques que juntament amb els MFCCs incrementés el percentatge.

Els resultats diuen que, a mesura que s'incorporen característiques, el percentatge d'encert s'incrementa. No obstant, aquest percentatge és baix i reflexa poca variació en l'elecció i nombre de característiques escollides. Es veu com, per exemple, aplicant GMM tenint en compte totes les característiques s'arriba a un nivell d'encert del 79.49%. Aquest nivell és de 72.86% quan només es considera els coeficients MFCC.

4.1.3. Variacions en el nombre de mostres

Per aquest experiment es tenen en compte totes les característiques.

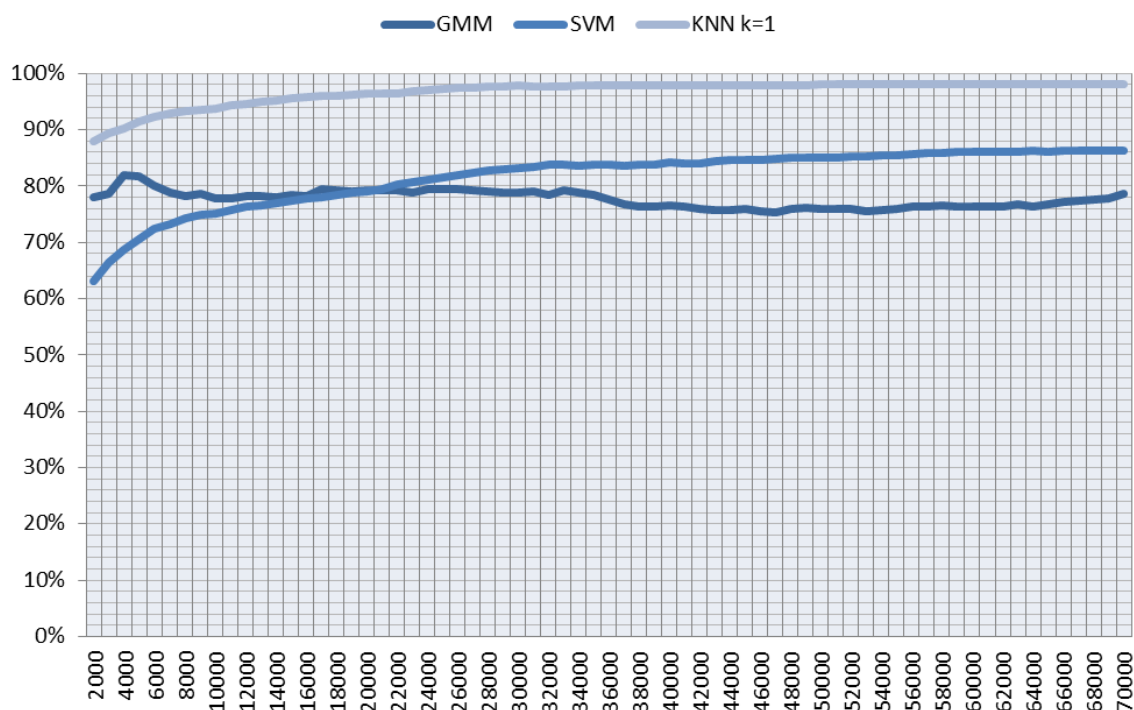


Figura 22. Percent. d'encert segons el nombre de mostres d'aprenentatge

La gràfica mostra molt clarament com el nombre de mostres no sembla ser una consideració a tindre en compte si s'aplica el mètode GMM on, en aquests experiments, assoleix millors resultats que aplicant SVM quan el model no sobrepassa les 20.000 mostres. En canvi, el percentatge d'èxit dels altres dos mètodes sí que depèn del nombre de mostres, arribant de manera asíptica a un nivell màxim.

4.1.4. Variacions en el metode k-NN

L'objectiu d'aquest joc de proves és veure com afecta l'increment del valor k en els resultats.

En un primer experiment, l'algorisme k-NN està condicionat a retornar la classe dominant quan $k > 1$. Si no existeix una classe dominant, el resultat és considerat incert.

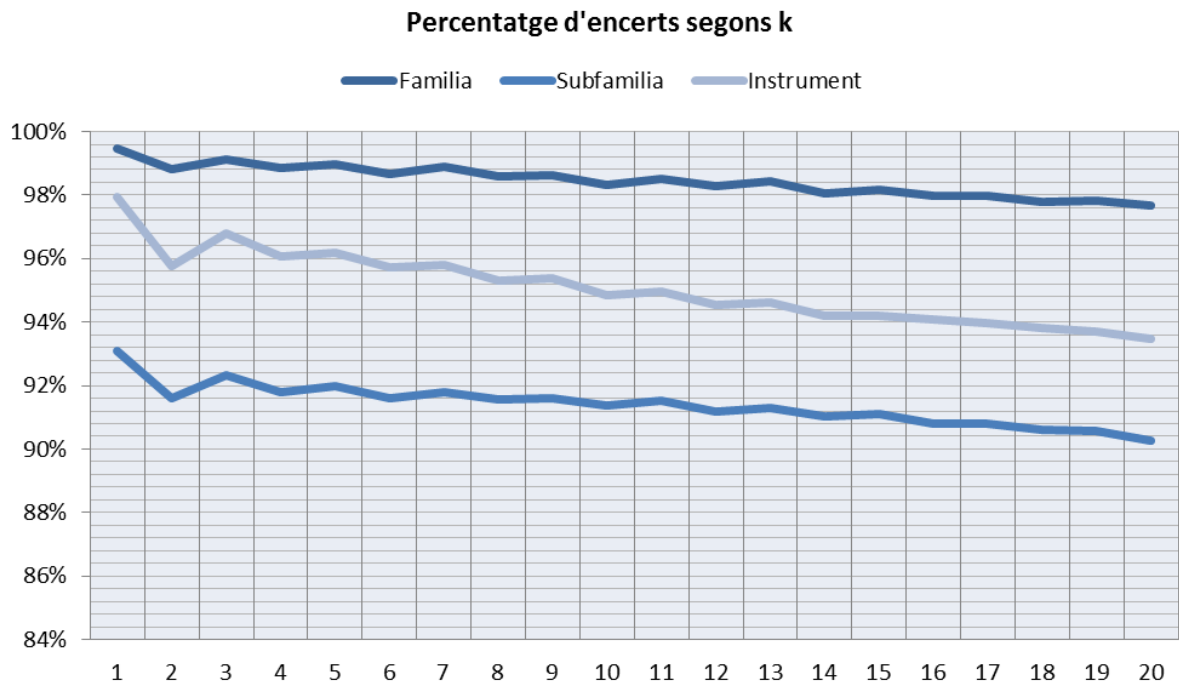


Figura 23. Percentatge d'incerts en funció de k aplicant k-NN sobre totes les característiques

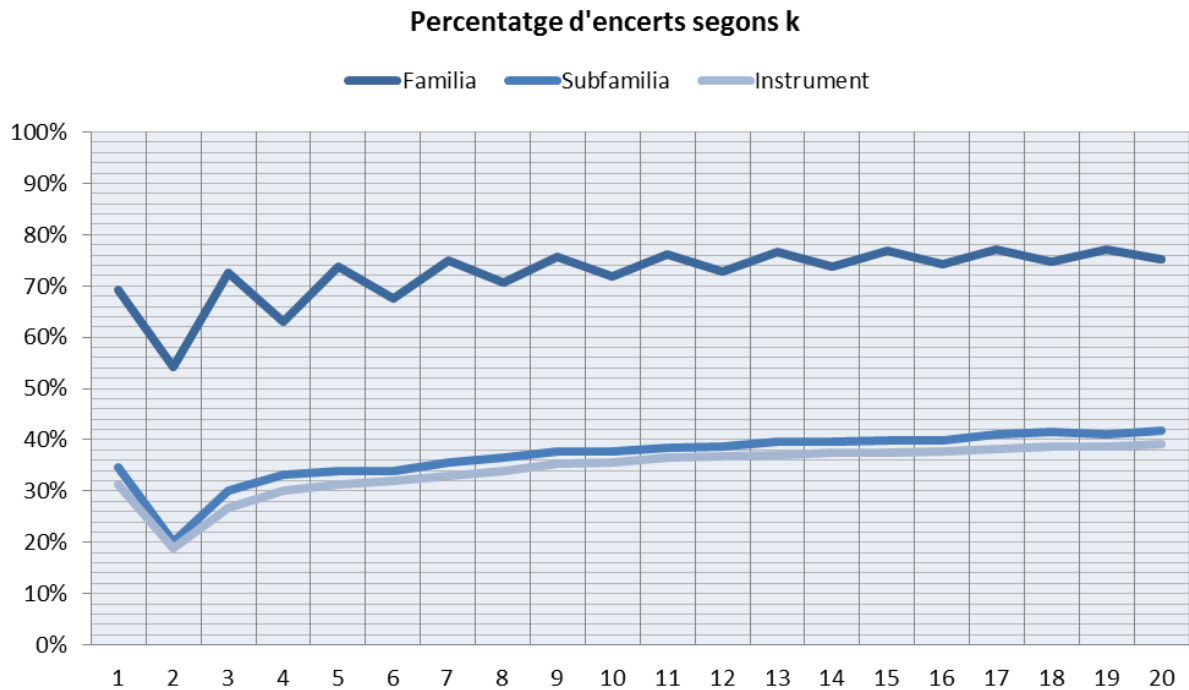


Figura 24. Percentatge d'encerts en funció de k aplicant k-NN sobre centroide

Els resultats són ben diferents si es tenen en compte totes les característiques o només una d'elles amb poc pes relatiu. En el primer cas, a mesura que s'incrementa k, el percentatge d'encerts disminueix linealment i en el segons cas s'incrementa.

Amb la finalitat de proporcionar algun mecanisme per a no debilitar l'algorisme k-NN en aquest escenari a mesura que s'incrementa k, es decideix desenvolupar una varietat que combini k-NN amb altre mètode.

La varietat afegeix l'aplicació del mètode SVM quan no hi ha coincidència de classe de tots els k-veïns més propers trobats amb k-NN. En aquest cas, els k-veïns són les mostres que es fan servir com model en un classificador SVM.

La varietat es descriu de la següent manera:

```

k_veïns = Obtenir_k_NN (MODEL, TEST);
SI totes_les_Classes_dels k_veïns SON iguals LLAVORS
    Classe_Classificador = Classe k_veïns
SINO
    Classe_Classificador = Obtenir_ClasseSVM (k_veïns, mostra_Test)
FSI

```

La descripció de la funció afegida és la següent:

Funció	Descripció
<code>cResults = TFCKNNClassify3</code> (cFeaturesTraining, cFeaturesTest, vWhat, scale, classType, k, maxDist)	Classifica els elements de <i>cFeaturesTest</i> segons el mètode k-NN prenent com model les dades de <i>FeaturesTraining</i> . <i>vWhat</i> especifica el subconjunt de característiques que es vol considerar. <i>Scale</i> si es vol escalar les dades en el rang [-1, +1]. <i>classType</i> indica si es vol classificar per família, subfamília o instrument (0, 1 i 2 respectivament). <i>K</i> indica el nombre de veïns més pròxims que cerca i <i>maxDist</i> és la distància màxima acceptada. Aquesta versió retorna la classe més predominant quan $k > 1$. Si no existeix una única classe predominant retorna l'elecció segons el mètode SVM creant un model amb els k-veïns retornats per l'algorisme k-NN.

Taula 17. Funció TFCKNNClassify3

Els resultat d'aplicar aquesta variant sobre el mateix tipus de conjunts és:

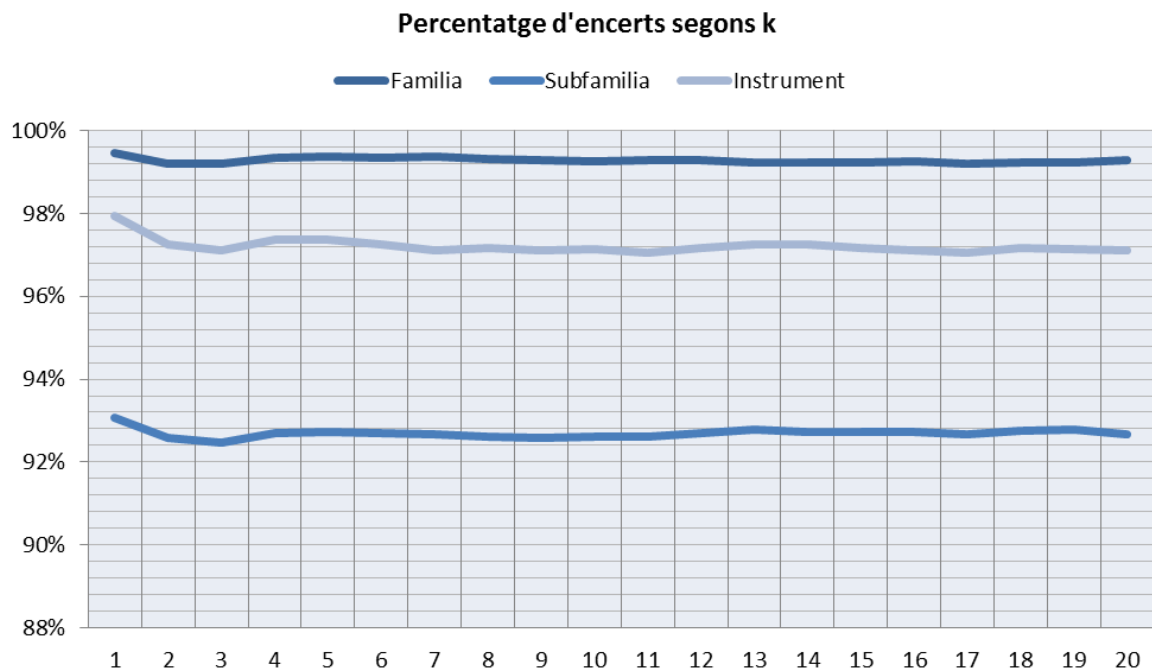


Figura 25. Percent. d'encerts en funció de k aplicant k-NN (totes les característiques) amb suport de SVM

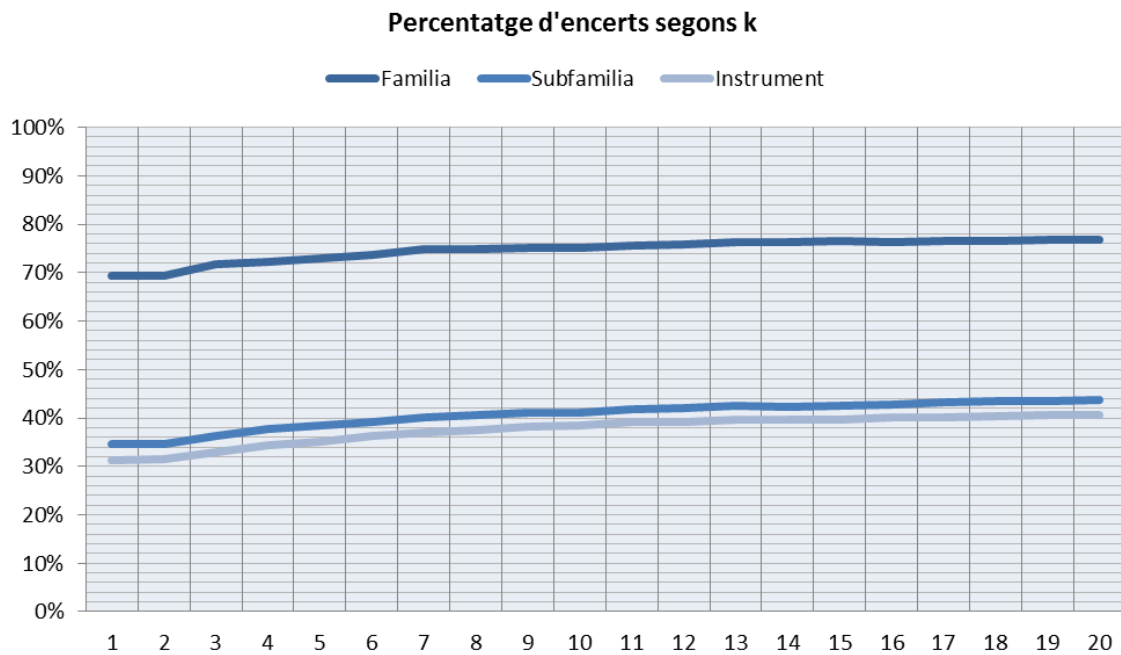


Figura 26. Percent. d'encerts en funció de k aplicant k-NN (centroide) amb suport de SVM

4.1.5. Matrius de confusió

Les matrius de confusió donen informació dels resultats a nivell de valor de la classe (família, subfamília o instrument). Amb aquestes matrius es pot estudiar aspectes com:

- Quin són els instruments més fàcils o difícils d'encertar
- Amb quins instruments es sol confondre un instrument en particular

Les taules següent mostren les matrius de confusió amb l'aplicació dels mètodes de classificació GMM, 1-NN i SVM. Les files marquen els resultats de la classificació.

	Acordio	Clarinet	Contrabaix	Corn angles	Fagot	Flauta	Guitarra	Oboe	Piano	Piccolo	Saxofon	Trombo	Trompa	Trompeta	Tuba	Viola	Violi	Violoncel	Total
Acordio	71	1															1		73
Clarinet	1	363	1		4		2	8	8		6			2		2	7	3	407
contrabaix		3	140		4	2	7		14							2		4	176
corn angles				67					1		1						1		70
fagot			1		265				3										269
flauta		1	3	1		272	6		2	4	1	10					1		301
guitarra	4	4	279		1	1	737	1	404		1			22		3	5	7	1469
oboe		1		1				62	2		2						1		69
piano			4		1		13		249			1	1			1	1		271
piccolo										112							1	1	114
saxofon		2		1			2	5	10		194	1		1		5			221
trombo						1		2	1		4	172	27						207
trompa								1	2			3	97	2	6				111
trompeta		12		1			4	38	1		9	1	3	411		4	5		489
tuba															168				168
viola	1		2		1		11							1		110	7	3	136
violi	5	5	1				5	1			2			2		24	175	1	221
violoncel		1	44	1			5		9		3	1		1		4		159	228
Total	82	393	475	72	276	276	792	118	706	116	224	188	128	442	174	156	205	177	5000

Figura 27. Matriu de confusió (instruments) aplicant GMM

	Bufst	Corda	Total		Fregat	Colpejat	Puntejat	Lliure llengueta	Llengueta simple	Doble llengueta	Bisell	Vibracio llavis	Total
Acordio	66	7	73	Acordio	4			68	1				73
Clarinet	321	86	407	Clarinet	7	9	2	1	375	11		2	407
contrabaix	6	170	176	contrabaix	82	15	56		5	15	1	2	176
corn angles	69	1	70	corn angles	4	1			6	58		1	70
fagot	169	100	269	fagot		18	2			249			269
flauta	273	28	301	flauta		3	10		2	6	278	2	301
guitarra	62	1407	1469	guitarra	23	420	987	2	5	1		31	1469
oboe	66	3	69	oboe	1	2			3	63			69
piano	6	265	271	piano	3	254	9		1	3		1	271
piccolo	114		114	piccolo	4						110		114
saxofon	194	27	221	saxofon	5	9	8		171	17		11	221
trombo	165	42	207	trombo		43	5		17	14	10	118	207
trompa	63	48	111	trompa		58	1			3		49	111
trompeta	488	1	489	trompeta	10		1	1	19	36		422	489
tuba		168	168	tuba		108	60						168
viola	61	75	136	viola	104		15	5	4			8	136
violi	101	120	221	violi	189		7	7	8	2	4	4	221
violoncel	41	187	228	violoncel	127	11	68		16	3	3		228
Total	2265	2735	5000	Total	563	951	1231	84	633	481	406	651	5000

Figura 28. Matriu de confusió (familia i subfamilia) aplicant GMM

	acordio	clarinet	contrabaix	corn angles	fagot	flauta	guitarra	oboe	piano	piccolo	saxofon	trombo	trompa	trompeta	tuba	viola	violi	violoncel	Total
acordio	71		1			1													73
clarinet		395							7		2	1		1			1		407
contrabaix		1	114		4		42		5		4				2	1		3	176
corn angles				69			1												70
fagot					268											1			269
flauta						297	2		2										301
guitarra	2	2	2		3	2	1428		8					7	11	1	1	2	1469
oboe					3		2	60	2								1		69
piano					4	1	50		207				3		5			1	271
piccolo										114									114
saxofon			1						3		213	1		1			2		221
trombo						1			3			197	5	1					207
trompa							2		6			3	99		1				111
trompeta					1		1		1			1	2	481				2	489
tuba															168				168
viola		4	1				4		1		1			4		110	7	4	136
violi	1	3	1				2		1		1	1	1	2		10	197	1	221
violoncel		5	1				30					1	1			3		187	228
Total	74	410	121	73	279	302	1564	60	246	114	222	205	110	498	188	127	209	198	5000

Figura 29. Matriu de confusió (instruments) aplicant 1-NN

	bufat	Corda	total		fregat	colpejat	puntejat	Lliure llengueta	Llengueta simple	doble llengueta	bisell	vibracio llavis	total
acordio	71	2	73	acordio	2			70			1		73
clarinet	398	9	407	clarinet	2	7			396			2	407
contrabaix	10	166	176	contrabaix	121	4	40		5	4		2	176
corn angles	69	1	70	corn angles			1			68		1	70
fagot	269		269	fagot						268		1	269
flauta	296	5	301	flauta		2	2				297		301
guitarra	26	1443	1469	guitarra	9	8	1425	2	2	3	2	18	1469
oboe	66	3	69	oboe	1	2	2			63		1	69
piano	15	256	271	piano	1	205	50			4	1	10	271
piccolo	113	1	114	piccolo	1						113		114
saxofon	216	5	221	saxofon	3	3			213			2	221
trombo	204	3	207	trombo		3					1	203	207
trompa	103	8	111	trompa		6	2					103	111
trompeta	485	4	489	trompeta	3		1			1		484	489
tuba	168		168	tuba								168	168
viola	9	127	136	viola	122	1	4		5			4	136
violi	8	213	221	violi	210	1	1	1	4			4	221
violoncel	8	220	228	violoncel	193		27		5		1	2	228
Total	2534	2466	5000	Total	668	242	1555	73	630	411	416	1005	5000

Figura 30. Matriu de confusió (família i subfamília) aplicant 1-NN

	acordio	clarinet	contrabaix	corn angles	fagot	flauta	guitarra	oboe	piano	piccolo	saxofon	trombo	trompa	trompeta	tuba	viola	violi	violoncel	total
acordio	67	1												5					73
clarinet		361		1		5	2	5		7	3			18		2	2	1	407
contrabaix		9	85		1	3	52		2	3					2	3	2	14	176
corn angles		5		51			2	2	2	7				1					70
fagot			2		249		5		11						2				269
flauta						277	10		2			12							301
guitarra	3	3	4		3	4	1430		1	1	1	1		16	2				1469
oboe		3		6				44	3		2			11					69
piano					1	1	166		94		1		4		4				271
piccolo										113								1	114
saxofon		15				6	1		12		172	1		10		4			221
trombo		1				5			6		2	167	24	2					207
trompa							2		16			6	82	3	2				111
trompeta	1	1	1	2			2				3	2	2	474					489
tuba						19		3							146				168
viola		10			1		13				3			5		81	20	3	136
violi	8	16	1				7			2	3			8		7	168	1	221
violoncel		6	2			4	54		1	1	4	1		6		1	1		147
Total	79	431	95	60	255	300	1768	48	158	117	208	193	112	559	158	98	194	167	5000

Figura 31. Matriu de confusió (instruments) aplicant SVM

	bufat	Corda	total		fregat	colpejat	puntejat	Lliure llengueta	Llengueta simple	doble llengueta	bisell	vibració llavis	total
acordio	58	15	73	acordio	7			63	1	1		1	73
clarinet	383	24	407	clarinet	16	3	3		359	1	1	24	407
contrabaix	37	139	176	contrabaix	105		50		9	6	1	5	176
corn angles	68	2	70	corn angles			2		12	47	2	7	70
fagot	239	30	269	fagot	10	9	6			235	1	8	269
flauta	285	16	301	flauta			11		3	2	272	13	301
guitarra	57	1412	1469	guitarra	4		1427	3	4	4	3	24	1469
oboe	68	1	69	oboe		2			5	41		21	69
piano	36	235	271	piano		69	167		1	1		33	271
piccolo	113	1	114	piccolo	2						112		114
saxofon	214	7	221	saxofon	5	4			185		6	21	221
trombo	205	2	207	trombo		9			2		8	188	207
trompa	106	5	111	trompa		7	2					102	111
trompeta	486	3	489	trompeta	3		3	1	3			479	489
tuba	22	146	168	tuba			66					102	168
viola	44	92	136	viola	106		13		10	1	1	5	136
violi	34	187	221	violi	189		5	5	12		5	5	221
violoncel	69	159	228	violoncel	149		51		18		7	3	228
Total	2524	2476	5000	Total	596	103	1806	72	624	339	419	1041	5000

Figura 32. Matriu de confusió (família i subfamília) aplicant SVM

Els resultats són en certa manera sorprenents. A nivell global, a arrel dels experiments anteriors, es sabia que el percentatge d'encerts seria molt alt. Però ara es pot veure la part d'implicació que hi té cada instrument en el global. I aquesta implicació varia entre els classificadors utilitzats. Per exemple, la identificació de la guitarra (1469 mostres) és un problema aplicant GMM (279 mostres confoses amb un contrabaix i 404 amb un piano). Però és perfectament identificable aplicant SVM o KNN (1430 i 1428 sobre el total).

4.2. CORPUS DE GENERACIO PROPIA

4.2.1. Rellevància de cada característica acústica

A l'igual que amb el corpus importat, la primera sèrie d'experiments ha consistit en estudiar la rellevància, per separat, de les 18 característiques acústiques (centroide, extensió, 13 coeficients MFCC, *rolloff*, *zerocross* i energia) segons els tres mètodes de classificació estudiats. El coeficients MFCC s'han considerat com un vector amb longitud creixent de manera que quan es considera, per exemple, el cinquè coeficient implícitament s'estan considerant també els coeficients 1-4.

No obstant, en aquest experiments s'estan tenint en compte els *frames* dels arxius d'àudio com unitats independents i, per tant, el resultats són sobre els *frames* i no sobre els arxius. D'aquesta manera s'equipara l'estudi amb els experiments fets amb el corpus importat.

De cara a valorar els resultats, les proporcions d'encert aleatori són del 50% (família), 33% (subfamília) i 6% (instrument).

Els resultats són els següents:

Percentatge d'encerts aplicant GMM

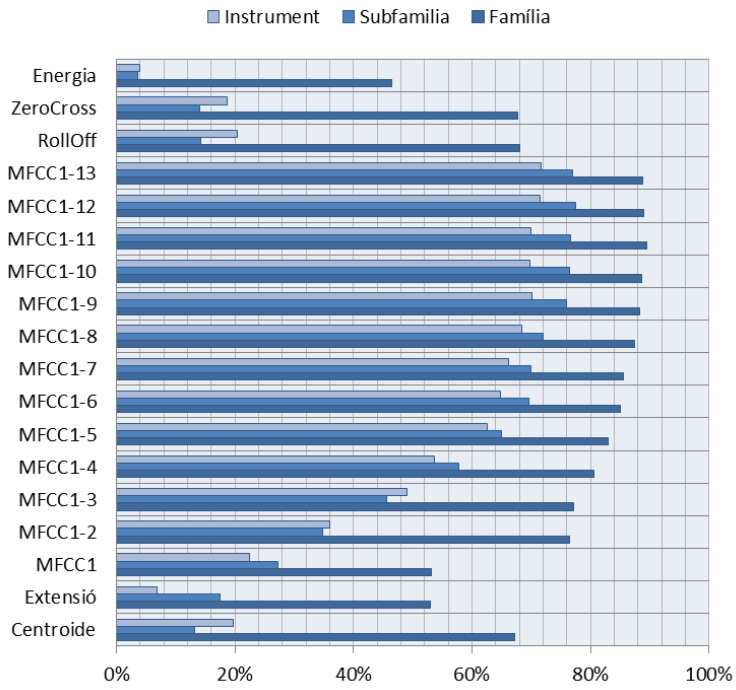


Figura 33. Percent. d'encerts aplicant GMM

Percentatge d'encerts aplicant 1-NN

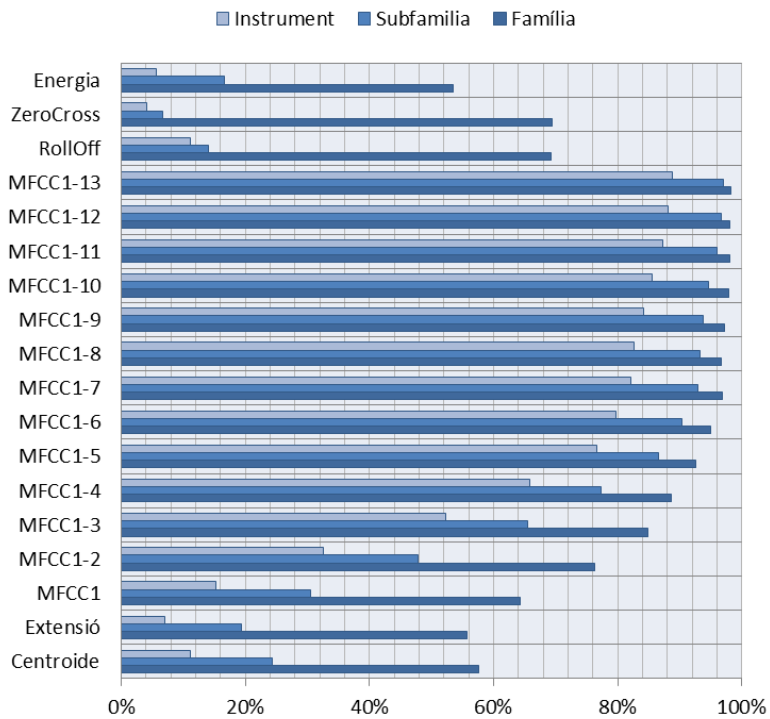


Figura 34. Percent. d'encerts aplicant 1-NN

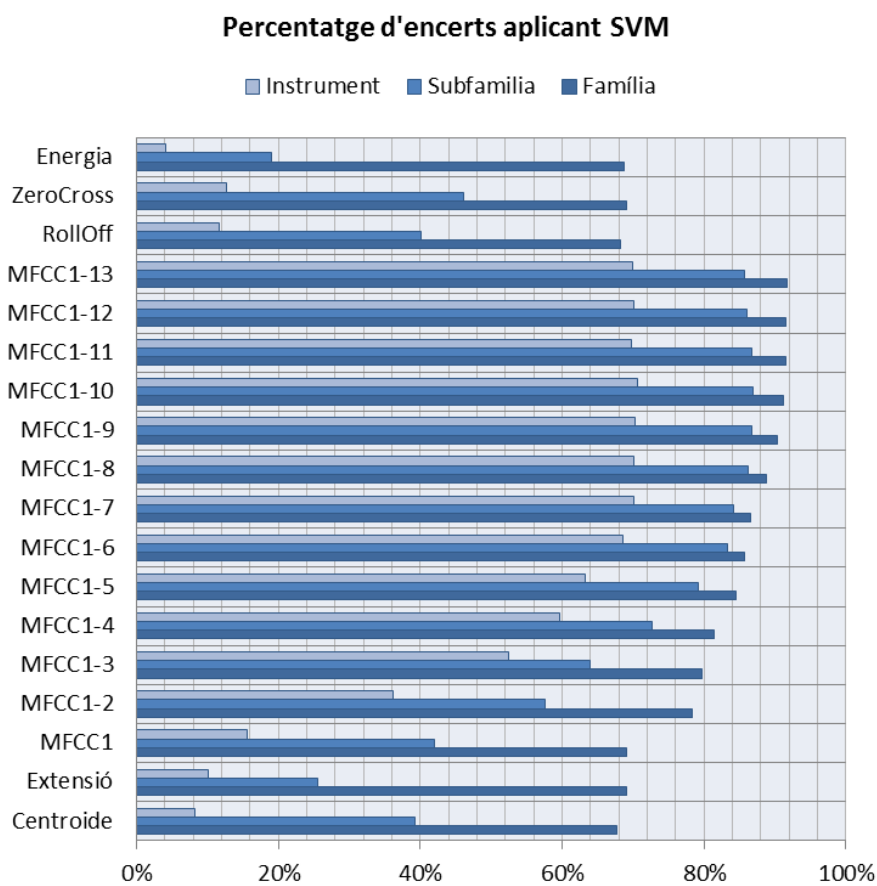


Figura 35. Percent. d'encerts aplicant SVM

Un primer anàlisi indica que el nivell d'encert de les característiques de menys pes relatiu (centroide, zerocross, rolloff, energia i extensió) està bastant per sota de l'esperat segons els resultats amb l'altre corpus. Una possible causa és el limitat nombre de mostres de les que es disposa.

A l'igual que amb el corpus anterior, k-NN dona millors resultats i SVM potencia més la rellevància de les característiques de menys pes.

4.2.2. Combinació de totes les característiques

En la següent figura es mostra el resultat de l'aplicació dels mètodes de classificació amb models creats amb totes les característiques acústiques:

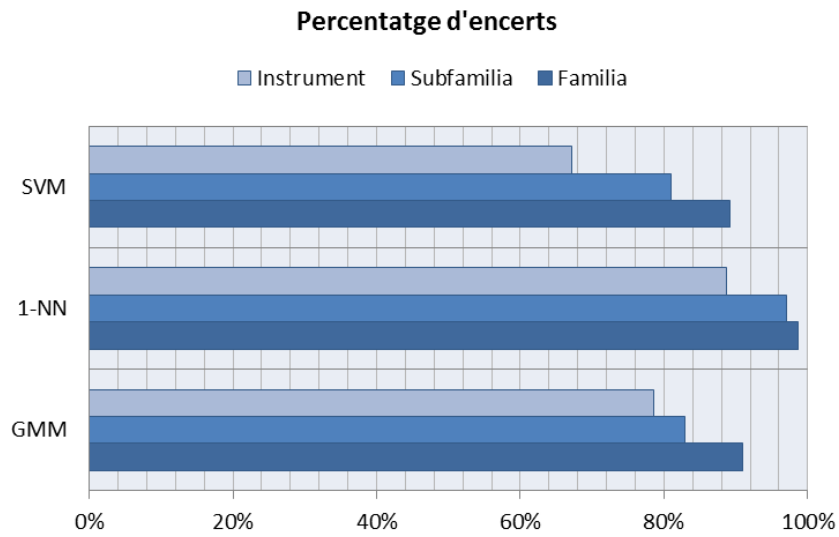


Figura 36. Percent. d'encert segons mètode amb totes les característiques

A l'igual que en els experiments anteriors k-NN és el mètode guanyador.

4.2.3. Variacions en el mètode k-NN

S'ha fet servir la variant que incorpora SVM.

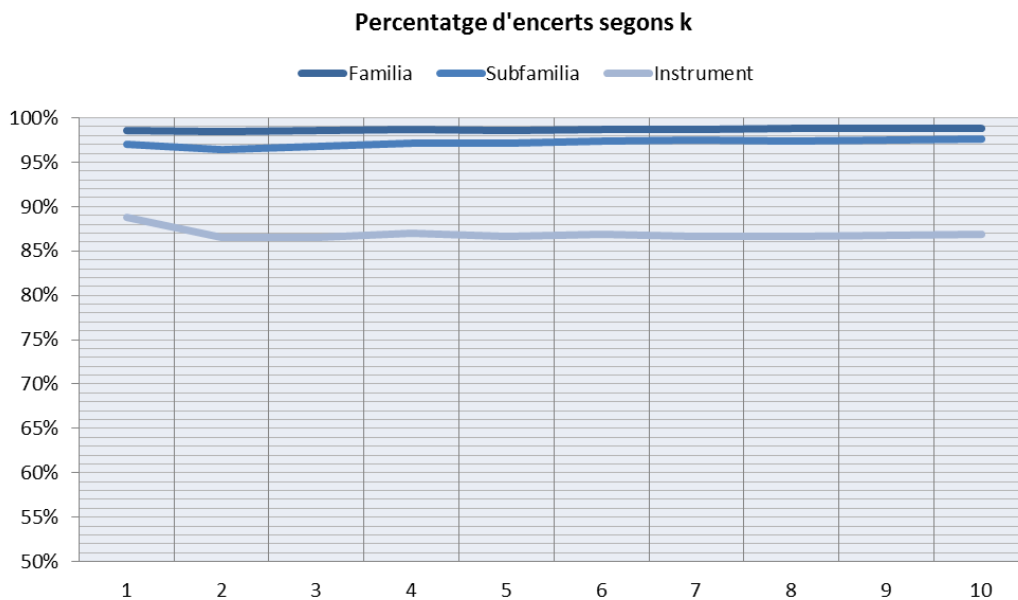


Figura 37. Percentatge d'encerts en funció de k aplicant k-NN sobre totes les característiques

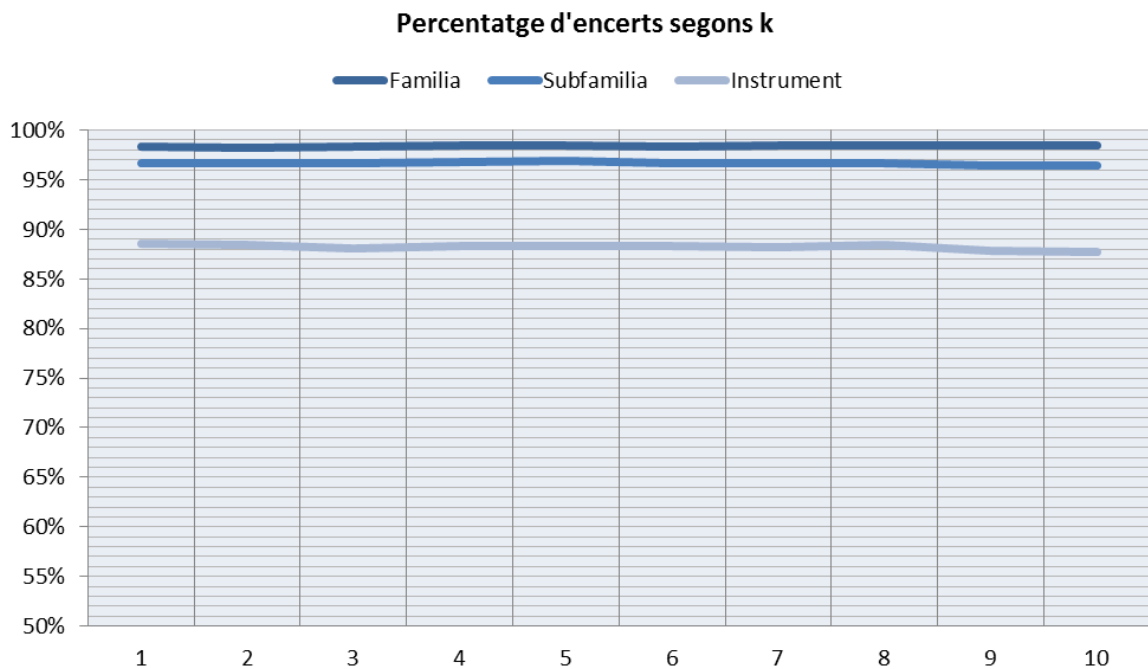


Figura 38. Percentatge d'encerts en funció de k aplicant k-NN sobre els 13 MFCCs

Els resultats són també semblants que amb el corpus anterior. No obstant, degut al poc pes que es veien de les característiques de poc pes, un experiment amb utilitzant només els MFCC no presenta el desnivell que es presenta quan es consideren totes les característiques.

4.2.4. Decisió del sistema a nivel d'arxius

Aquest conjunt de resultats no es poden comparar amb el del corpus importat. Ara els resultats mostrats són en referència als arxius que s'han fet servir de test.

4.2.4.1. Arxius de test de generació pròpia

Aquests arxius s'han generat amb les mateixes condicions de gravació que les utilitzades per a obtenir les mostres del models d'entrenament. La presentació és en forma de taules en comptes de gràfics per a poder visualitzar més informació. L'elecció de la classe es basa en la més predominant entre tots els *frames*. Inclou de forma addicional la segona classe més predominant.

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
acordio.wav	acordio	bufat	lliure_llengueta	66	acordio	62.12%	violi	21.21%	SI
flauta.wav	flauta	bufat	bisell	271	flauta	92.25%	piccolo	5.90%	SI
piccolo.wav	piccolo	bufat	bisell	249	piccolo	93.17%	flauta	5.62%	SI
fagot.wav	fagot	bufat	doble_llengueta	377	fagot	85.94%	corn_angles	5.84%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	113	corn_angles	83.19%	fagot	9.73%	SI
oboe.wav	oboe	bufat	doble_llengueta	275	oboe	54.55%	corn_angles	30.91%	SI
trompa.wav	trompa	bufat	vibracio_llavis	449	trompa	38.31%	fagot	34.30%	SI
trombo.wav	trombo	bufat	vibracio_llavis	308	trombo	92.53%	tuba	4.55%	SI
trompeta.wav	trompeta	bufat	vibracio_llavis	301	trompeta	81.06%	trombo	10.30%	SI
tuba.wav	tuba	bufat	vibracio_llavis	302	tuba	80.13%	trombo	13.58%	SI
clarinet.wav	clarinet	bufat	llengueta_simple	252	clarinet	99.60%	flauta	0.40%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	298	saxofon	94.63%	piano	3.02%	SI
guitarra.wav	guitarra	corda	puntejat	241	guitarra	97.51%	piano	1.24%	SI
violoncel.wav	violoncel	corda	fregat	344	violoncel	64.83%	contrabaix	23.26%	SI
contrabaix.wav	contrabaix	corda	fregat	214	contrabaix	98.60%	guitarra	1.40%	SI
viola.wav	viola	corda	fregat	353	viola	82.44%	violi	10.76%	SI
violi.wav	violi	corda	fregat	138	viola	81.16%	violi	17.39%	NO
piano.wav	piano	corda	colpejat	165	piano	93.33%	saxofon	1.82%	SI

Taula 18. Resultats de classificació de l'instrument aplicant GMM

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
acordio.wav	acordio	bufat	lliure_llengueta	66	acordio	81.82%	violoncel	18.18%	SI
flauta.wav	flauta	bufat	bisell	271	flauta	88.56%	piccolo	6.27%	SI
piccolo.wav	piccolo	bufat	bisell	249	piccolo	99.20%	violi	0.40%	SI
fagot.wav	fagot	bufat	doble_llengueta	377	fagot	84.62%	corn_angles	9.81%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	113	corn_angles	74.34%	oboe	24.78%	SI
oboe.wav	oboe	bufat	doble_llengueta	275	oboe	80.73%	corn_angles	15.27%	SI
trompa.wav	trompa	bufat	vibracio_llavis	449	trompa	84.41%	trompeta	4.23%	SI
trombo.wav	trombo	bufat	vibracio_llavis	308	trombo	86.69%	tuba	8.12%	SI
trompeta.wav	trompeta	bufat	vibracio_llavis	301	trompeta	97.67%	trombo	1.00%	SI
tuba.wav	tuba	bufat	vibracio_llavis	302	tuba	96.69%	trompa	1.99%	SI
clarinet.wav	clarinet	bufat	llengueta_simple	252	clarinet	99.21%	trompa	0.79%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	298	saxofon	98.32%	piano	1.68%	SI
guitarra.wav	guitarra	corda	puntejat	241	guitarra	96.27%	trompa	1.66%	SI
violoncel.wav	violoncel	corda	fregat	344	violoncel	90.70%	acordio	2.62%	SI
contrabaix.wav	contrabaix	corda	fregat	214	contrabaix	90.65%	violoncel	7.48%	SI
viola.wav	viola	corda	fregat	353	viola	77.34%	violi	21.25%	SI
violi.wav	violi	corda	fregat	138	violi	52.17%	viola	47.10%	SI
piano.wav	piano	corda	colpejat	165	piano	97.58%	saxofon	1.21%	SI

Taula 19. Resultats de classificació de l'instrument aplicant 1-NN

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Família	Subfamília		Classe	Percentatge	Classe	Percentatge	
acordio.wav	acordio	bufat	lliure_llengueta	66	acordio	43.94%	violi	42.42%	SI
flauta.wav	flauta	bufat	bisell	271	flauta	72.32%	piccolo	22.88%	SI
piccolo.wav	piccolo	bufat	bisell	249	piccolo	98.80%	violi	0.80%	SI
fagot.wav	fagot	bufat	doble_llengueta	377	fagot	43.24%	corn_angles	32.63%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	113	corn_angles	81.42%	trompeta	17.70%	SI
oboe.wav	oboe	bufat	doble_llengueta	275	oboe	66.91%	corn_angles	22.18%	SI
trompa.wav	trompa	bufat	vibracio_llavis	449	trompeta	23.83%	tuba	18.49%	SI
trombo.wav	trombo	bufat	vibracio_llavis	308	trombo	78.90%	tuba	12.34%	SI
trompeta.wav	trompeta	bufat	vibracio_llavis	301	trompeta	91.36%	trombo	3.65%	SI
tuba.wav	tuba	bufat	vibracio_llavis	302	tuba	98.68%	guitarra	1.32%	SI
clarinet.wav	clarinet	bufat	llengueta_simple	252	clarinet	98.41%	trompa	0.79%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	298	saxofon	90.60%	piano	8.72%	SI
guitarra.wav	guitarra	corda	puntejat	241	guitarra	97.93%	piano	1.66%	SI
violoncel.wav	violoncel	corda	fregat	344	violoncel	55.81%	trompa	11.05%	SI
contrabaix.wav	contrabaix	corda	fregat	214	contrabaix	76.17%	guitarra	21.96%	SI
viola.wav	viola	corda	fregat	353	violi	46.46%	trompeta	26.35%	NO
violi.wav	violi	corda	fregat	138	trompeta	48.55%	violi	44.93%	NO
piano.wav	piano	corda	colpejat	165	piano	69.70%	saxofon	17.58%	SI

Taula 20. Resultats de classificació de l'instrument aplicant SVM

Com es pot veure de les taules, els resultats són molt bons amb una diferència entre el primer (candidat escollit) i segon candidat prou ample amb petites excepcions (p.e. acordió).

El fet de què la decisió es faci tenint en compte tots els *frames* que formen l'arxiu d'àudio original, incrementa de manera significativa el percentatge d'encerts, on s'arriba al 100 % aplicant 1-NN. En el cas d'aplicar GMM només s'ha produït un error que podria considerar-se comú pel tipus d'instruments (es confon un violí amb una viola). No obstant, encara que els resultats aplicant SVM són igualment bons (només dos errades) el que sorprèn és l'instrument escollit en una de les decisions errònies (trompeta per violí).

4.2.4.2. Gravacions de Musical Instruments

En aquest joc de proves les condicions de gravació no són les mateixes que les que van ser aplicades per la fase d'aprenentatge. Per tant, el percentatge d'encerts es ressentirà d'aquesta situació.

Es resultats són els següents:

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	bufat	98.90%	corda	1.10%	SI
oboesolo.wav	oboe	bufat	doble_llengüeta	793	bufat	99.87%	corda	0.13%	SI
hornsolo.wav	trompa	bufat	vibracio_llavis	573	bufat	91.80%	corda	8.20%	SI
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	bufat	96.64%	corda	3.36%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	bufat	92.02%	corda	7.98%	SI
tubasolo.wav	tuba	bufat	vibracio_llavis	390	bufat	98.46%	corda	1.54%	SI
celosolo.wav	violoncel	corda	fregat	683	corda	94.58%	bufat	5.42%	SI
clarsolo.wav	clarinet	bufat	llengüeta_simple	320	bufat	50.63%	corda	49.38%	SI
violsolo.wav	violí	corda	fregat	633	bufat	87.84%	corda	12.16%	NO
vilasolo.wav	viola	corda	fregat	738	corda	83.33%	bufat	16.67%	SI

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	bisell	43.09%	doble_llengüeta	24.86%	SI
oboesolo.wav	oboe	bufat	doble_llengüeta	793	bisell	50.69%	doble_llengüeta	45.27%	NO
hornsolo.wav	trompa	bufat	vibracio_llavis	573	vibracio_llavis	58.46%	doble_llengüeta	15.01%	SI
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	vibracio_llavis	77.61%	llengüeta_simple	9.70%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	doble_llengüeta	51.68%	vibracio_llavis	36.34%	NO
tubasolo.wav	tuba	bufat	vibracio_llavis	390	vibracio_llavis	98.72%	puntejat	1.03%	SI
celosolo.wav	violoncel	corda	fregat	683	fregat	73.21%	llengüeta_simple	13.03%	SI
clarsolo.wav	clarinet	bufat	llengüeta_simple	320	fregat	77.81%	llengüeta_simple	15.31%	NO
violsolo.wav	violí	corda	fregat	633	fregat	33.02%	doble_llengüeta	27.80%	SI
vilasolo.wav	viola	corda	fregat	738	fregat	57.99%	colpejat	32.52%	SI

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	flauta	38.12%	violí	24.31%	SI
oboesolo.wav	oboe	bufat	doble_llengüeta	793	flauta	45.27%	oboe	44.51%	NO
hornsolo.wav	trompa	bufat	vibracio_llavis	573	clarinet	49.74%	trompa	33.86%	NO
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	trompeta	75.00%	flauta	13.06%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	trompa	70.17%	flauta	12.82%	NO
tubasolo.wav	tuba	bufat	vibracio_llavis	390	tuba	49.49%	trompa	27.44%	SI
celosolo.wav	violoncel	corda	fregat	683	violoncel	50.51%	piano	22.40%	SI
clarsolo.wav	clarinet	bufat	llengüeta_simple	320	piano	40.63%	violoncel	33.13%	NO
violsolo.wav	violí	corda	fregat	633	piano	47.55%	violoncel	20.54%	NO
vilasolo.wav	viola	corda	fregat	738	piano	75.61%	flauta	10.98%	NO

Taula 21. Resultats de classificació aplicant GMM

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	bufat	83.70%	corda	16.30%	SI
oboesolo.wav	oboe	bufat	doble_llengueta	793	bufat	99.37%	corda	0.63%	SI
hornsolo.wav	trompa	bufat	vibracio_llavis	573	bufat	95.81%	corda	4.19%	SI
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	bufat	92.16%	corda	7.84%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	bufat	99.79%	corda	0.21%	SI
tubasolo.wav	tuba	bufat	vibracio_llavis	390	bufat	99.74%	corda	0.26%	SI
celosolo.wav	violoncel	corda	fregat	683	corda	78.77%	bufat	21.23%	SI
clarsolo.wav	clarinet	bufat	llengueta_simple	320	corda	56.88%	bufat	43.13%	NO
violsolo.wav	violí	corda	fregat	633	corda	61.93%	bufat	38.07%	SI
vilasolo.wav	viola	corda	fregat	738	corda	87.53%	bufat	12.47%	SI

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	bisell	37.29%	vibracio_llavis	31.77%	SI
oboesolo.wav	oboe	bufat	doble_llengueta	793	vibracio_llavis	87.89%	doble_llengueta	7.31%	NO
hornsolo.wav	trompa	bufat	vibracio_llavis	573	llengueta_simple	65.10%	vibracio_llavis	24.08%	NO
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	vibracio_llavis	91.42%	fregat	7.84%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	vibracio_llavis	99.37%	doble_llengueta	0.42%	SI
tubasolo.wav	tuba	bufat	vibracio_llavis	390	vibracio_llavis	98.21%	doble_llengueta	1.54%	SI
celosolo.wav	violoncel	corda	fregat	683	fregat	57.39%	colpejat	19.77%	SI
clarsolo.wav	clarinet	bufat	llengueta_simple	320	fregat	44.69%	llengueta_simple	29.06%	NO
violsolo.wav	violí	corda	fregat	633	fregat	53.71%	vibracio_llavis	21.48%	SI
vilasolo.wav	viola	corda	fregat	738	colpejat	65.18%	fregat	22.36%	NO

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	flauta	33.43%	trompa	23.20%	SI
oboesolo.wav	oboe	bufat	doble_llengueta	793	trompa	87.64%	oboe	4.41%	NO
hornsolo.wav	trompa	bufat	vibracio_llavis	573	clarinet	64.57%	tuba	14.14%	NO
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	trompeta	85.07%	trompa	6.34%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	trompa	76.47%	trombo	16.81%	NO
tubasolo.wav	tuba	bufat	vibracio_llavis	390	trompa	51.03%	tuba	45.90%	NO
celosolo.wav	violoncel	corda	fregat	683	contrabaix	33.53%	violoncel	23.57%	NO
clarsolo.wav	clarinet	bufat	llengueta_simple	320	violoncel	41.56%	clarinet	29.06%	NO
violsolo.wav	violí	corda	fregat	633	violoncel	24.01%	viola	17.54%	NO
vilasolo.wav	viola	corda	fregat	738	piano	65.18%	violí	14.77%	NO

Taula 22. Resultats de classificació aplicant 1-NN

Dades originals				frames	1er candidat		2on candidat		Encert
Arxiu	Instrument	Familia	Subfamília		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	bufat	89.23%	corda	10.77%	SI
oboesolo.wav	oboe	bufat	doble_llengueta	793	bufat	100.00%		0.00%	SI
hornsolo.wav	trompa	bufat	vibracio_llavis	573	bufat	99.83%	corda	0.17%	SI
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	bufat	91.42%	corda	8.58%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	bufat	96.22%	corda	3.78%	SI
tubasolo.wav	tuba	bufat	vibracio_llavis	390	bufat	99.74%	corda	0.26%	SI
celosolo.wav	violoncel	corda	fregat	683	corda	72.04%	bufat	27.96%	SI
clarsolo.wav	clarinet	bufat	llengueta_simple	320	corda	72.81%	bufat	27.19%	NO
violsolo.wav	violi	corda	fregat	633	bufat	84.04%	corda	15.96%	NO
vilasolo.wav	viola	corda	fregat	738	corda	91.73%	bufat	8.27%	SI
Dades originals				frames	1er candidat		2on candidat		Encert
Arxiu	Instrument	Familia	Subfamília		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	vibracio_llavis	27.07%	bisell	25.97%	NO
oboesolo.wav	oboe	bufat	doble_llengueta	793	bisell	47.29%	doble_llengueta	35.69%	NO
hornsolo.wav	trompa	bufat	vibracio_llavis	573	vibracio_llavis	36.65%	doble_llengueta	28.80%	SI
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	vibracio_llavis	83.21%	fregat	16.04%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	doble_llengueta	69.54%	vibracio_llavis	23.53%	NO
tubasolo.wav	tuba	bufat	vibracio_llavis	390	vibracio_llavis	96.67%	doble_llengueta	1.79%	SI
celosolo.wav	violoncel	corda	fregat	683	fregat	87.99%	vibracio_llavis	6.44%	SI
clarsolo.wav	clarinet	bufat	llengueta_simple	320	fregat	89.06%	vibracio_llavis	4.06%	NO
violsolo.wav	violi	corda	fregat	633	fregat	37.44%	vibracio_llavis	28.59%	SI
vilasolo.wav	viola	corda	fregat	738	fregat	84.28%	doble_llengueta	6.64%	SI
Dades originals				frames	1er candidat		2on candidat		Encert
Arxiu	Instrument	Familia	Subfamília		Classe	Percentatge	Classe	Percentatge	
flutsolo.wav	flauta	bufat	bisell	362	flauta	33.15%	trompa	26.52%	SI
oboesolo.wav	oboe	bufat	doble_llengueta	793	flauta	38.97%	trompa	26.36%	NO
hornsolo.wav	trompa	bufat	vibracio_llavis	573	clarinet	45.55%	fagot	27.05%	NO
trumsolo.wav	trompeta	bufat	vibracio_llavis	268	trompeta	84.70%	violi	7.09%	SI
ttbnsolo.wav	trombo	bufat	vibracio_llavis	476	fagot	36.97%	trompa	21.85%	NO
tubasolo.wav	tuba	bufat	vibracio_llavis	390	tuba	66.92%	fagot	32.05%	SI
celosolo.wav	violoncel	corda	fregat	683	violoncel	63.54%	trompa	18.89%	SI
clarsolo.wav	clarinet	bufat	llengueta_simple	320	violoncel	32.81%	clarinet	17.81%	NO
violsolo.wav	violi	corda	fregat	633	violoncel	33.81%	trompa	15.48%	NO
vilasolo.wav	viola	corda	fregat	738	violoncel	37.53%	violi	24.66%	NO

Taula 23. Resultats de classificació aplicant SVM

En aquests experiments, els millors resultats a nivell de subfamília s'aconsegueixen quan s'aplica el mètode GMM amb un 60 % d'encerts. A nivell d'instruments, el mètode SVM és el que més encerts obté però el percentatge és baix (40%). La raó d'aquests resultats és conseqüència de la qualitat de les mostres que s'han fet servir per aprenentatge (sintetitzades amb un teclat electrònic de baixa gama).

4.2.4.3. Gravacions polifòniques

Finalment, s'ha enregistrat sons polifònics corresponents a 4 instruments per a veure si el sistema pot reconèixer els instruments. Les gravacions polifòniques contenen de 2 a quatre notes simultànies. Els resultats han sigut els següents:

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	bufat	81.52%	corda	18.48%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	bufat	54.15%	corda	45.85%	SI
flauta.wav	flauta	bufat	bisell	203	bufat	99.01%	corda	0.99%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	bufat	100.00%		0.00%	SI

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	llengueta_simple	71.55%	vibracio_llavis	14.37%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	llengueta_simple	79.48%	colpejat	20.09%	SI
flauta.wav	flauta	bufat	bisell	203	bisell	69.95%	vibracio_llavis	9.36%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	doble_llengueta	76.28%	vibracio_llavis	18.60%	SI

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	clarinet	82.40%	piano	10.56%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	saxofon	59.83%	piano	37.12%	SI
flauta.wav	flauta	bufat	bisell	203	flauta	81.77%	corn_angles	9.36%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	corn_angles	56.28%	fagot	17.67%	SI

Taula 24. Resultats de classificació aplicant GMM

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	bufat	87.98%	corda	12.02%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	bufat	81.22%	corda	18.78%	SI
flauta.wav	flauta	bufat	bisell	203	bufat	98.52%	corda	1.48%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	bufat	97.67%	corda	2.33%	SI

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	llengueta_simple	79.18%	vibracio_llavis	8.50%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	llengueta_simple	81.22%	colpejat	18.78%	SI
flauta.wav	flauta	bufat	bisell	203	bisell	81.28%	vibracio_llavis	16.26%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	vibracio_llavis	50.23%	doble_llengueta	47.44%	NO

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	clarinet	79.18%	trombo	7.92%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	saxofon	81.22%	piano	18.78%	SI
flauta.wav	flauta	bufat	bisell	203	flauta	81.28%	trompa	8.87%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	corn_angles	36.28%	trompeta	27.91%	SI

Taula 25. Resultats de classificació aplicant 1-NN

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	bufat	83.87%	corda	16.13%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	bufat	74.24%	corda	25.76%	SI
flauta.wav	flauta	bufat	bisell	203	bufat	99.51%	corda	0.49%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	bufat	100.00%		0.00%	SI

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	llengueta_simple	46.04%	fregat	25.22%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	llengueta_simple	88.65%	colpejat	4.80%	SI
flauta.wav	flauta	bufat	bisell	203	bisell	65.52%	vibracio_llavis	17.24%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	doble_llengueta	82.33%	vibracio_llavis	17.21%	SI

Arxiu	Dades originals			frames	1er candidat		2on candidat		Encert
	Instrument	Familia	Subfamilia		Classe	Percentatge	Classe	Percentatge	
clarinet.wav	clarinet	bufat	llengueta_simple	341	clarinet	74.49%	trompeta	10.85%	SI
saxofon.wav	saxofon	bufat	llengueta_simple	229	saxofon	97.38%	piano	2.62%	SI
flauta.wav	flauta	bufat	bisell	203	flauta	70.94%	trompeta	10.84%	SI
corn_angles.wav	corn_angles	bufat	doble_llengueta	215	corn_angles	76.74%	trompeta	21.86%	SI

Taula 26. Resultats de classificació aplicant SVM

Els experiments han donat un 100 % d'encerts en tots els casos. Demostren que la simultaneïtat de notes no ha afectat massa en els resultats (tenint en compte que es tracta del mateix instrument) encara que, tot i això, sí que ha fet que el percentatge associat al primer de candidat s'hagi reduït. Això posa de manifest la complexitat que es crea a l'hora de reconèixer instruments en fraseig polifònic: les característiques acústiques associades als sons aïllats dels instruments no seran prou informació en fraseig polifònic d'instruments diferents.

CONCLUSIONS

En aquest treball, després d'una breu exposició dels conceptes acústics a nivell físic i fisiològic, s'han analitzat els paràmetres acústics més rellevants per a la identificació d'instruments musicals. També, s'han descrit el concepte i funcionament dels mètodes de classificació *Gaussian Mixture Model*, *k-Nearest Neighbor* i *Support Vector*. Amb aquests fonaments, l'objectiu era estudiar el comportament dels diferents classificadors i paràmetres acústics a partir de la configuració de diferents escenaris de treball.

Un dels contratemps que es solen presentar en aquest tipus de treballs de reconeixement de patrons és la manca d'un volum de mostres que sigui suficientment gran com per a treure unes conclusions amb prou fonaments estadístics. Això s'ha aconseguit aquí gràcies a que s'ha incorporat un conjunt de més de 115.000 mostres que, encara que no es disposava de la informació d'àudio original, han sigut de gran valor pels objectius del projecte.

Tot i això no s'ha deixat la banda la part de creació d'un corpus propi per a l'estudi de les fases d'extraccions de característiques que hi ha present abans d'aplicar els algorismes de generació de models i classificació.

Per altra banda, amb la finalitat d'assolir l'objectiu d'aquest treball, s'ha desenvolupat una sèrie de funcions en Matlab que permeten, d'una manera simple, desenvolupar i planificar els diferents experiments que s'han portat a terme.

Dels experiments realitzats s'ha pogut observar que no hi ha molt benefici en afegir combinacions de les característiques de menys pes als MFCC per a obtenir un major nombre d'encerts. Encara que s'incrementa lleugerament, amb els MFCC de per sí, ja s'obtenen uns resultats prou alts en comparació a la resta.

Una dada molt interessant ha sigut veure com en funció del mètode de classificació aplicat, el percentatge d'encerts a partir de les característiques de poc pes relatiu variava

apreciablement. Quan amb *zerocross* s'obté un nivell d'encert del 8.55 % si s'aplica el mètode k-NN, amb l'aplicació del mètode SVM aquest nivell arriba fins al 36.82 %.

La possibilitat de treballar amb una quantitat molt gran de mostres ha sigut útil per a estudiar el comportament dels classificadors segons el seu nombre. L'estudi revela de manera clara que el mètode del Model de Mescles de Gaussians està menys condicionat al nombre de mostres mentre que l'eficàcia dels mètodes SVM i k-NN sembla dependre del nombre de mostres.

S'ha desenvolupat una variant del mètode k-NN per a k majors de 1, amb la finalitat d'evitar un descens en el percentatge d'encerts en aquest casos. L'assumpció inicial era que l'increment de k faria pujar el nivell d'encerts però els experiments semblen refutar-lo ja que el conjunt de candidats resultants crea confusió a l'hora de prendre la decisió.

La variant desenvolupada ha fet servir el mètode SVM per a resoldre les confusions que es creen quan no tots els k-candidats pertanyen a la mateixa classe. Encara i això, els resultats milloren el nivell d'encerts quan k és més gran que 1, però només arriben a superar de molt poc els obtinguts directament amb el mètode SVM.

A l'igual que els paràmetres acústics tenen una rellevància en funció del mètode de classificació utilitzat, el mateix passa a l'hora d'identificar els instruments. Les matrius de confusió creades a tal efecte han demostrat com, per exemple, la guitarra és molt més difícil d'identificar amb el mètode GMM que amb els altres dos. Seria interessant aprofundir en aquest camp per a un possible mètode jeràrquic de classificació en el que, en funció del node, s'aplica el mètode més adient.

L'estudi amb les dades de generació pròpia confirma que el nombre de mostres és molt important per a enllestir un estudi d'aquest tipus. Encara que proporcionalment els resultats en nombre d'encerts són similars, a nivell absoluts són més baixos.

De la identificació a nivell d'arxiu d'àudio es veu com aquests depenen molt del procés d'elaboració comú que tinguin les mostres de model i les de test. Quan els dos tipus de mostres han estat enregistrats amb els mateixos equips i mètode d'enregistrament, l'èxit

de les proves arriba quasi al 100 % en la majoria dels casos. Quan hi ha variacions en aquest procés, l'índex d'encerts disminueix.

Finalment, s'ha volgut provar el sistema amb sons polifònics d'un mateix instrument amb uns resultats més alts dels esperats. No obstant, és un avís de què s'ha d'investigar en més direccions per a escenaris polifònics de més complexitat.

BIBLIOGRAFIA

- [1] Ulrichy Michels, *Atlas de música, I.*: Alianza Editorial, 1989.
- [2] wikipedia. [Online]. <http://www.wikipedia.org>
- [3] David Lapp. (2003) The physics of music and musical instruments. [Online].
http://www.tufts.edu/as/wright_center/workshop/workshop_archives/physics_2003_wkshp/book.htm
- [4] Joaquin Zamacois, *Teoria de la musica, Libro II.*: Editorial Labor, 1990.
- [5] S. S. Stevens, Volkman, and E. B. Newman, "A scale for the measurement of the psychological magnitude pitch," *Journal of Acoustical Society of America*, vol. 8, no. 3, pp. 185-190, 1937.
- [6] Schmidt-Jones. (2011, March) Timbre: The Color of Music. [Online].
<http://cnx.org/content/m11059/latest/>
- [7] Antti Eronen. (2001, April) Automatic Musical Instrument Recognition. [Online].
http://www.cs.tut.fi/sgn/arg/music/eronen/eronen_msc.pdf
- [8] Keith Dana Martin. (1999, June) Sound Source Recognition: A Theory of Computational Model. [Online]. <http://sound.media.mit.edu/Papers/kdm-pdhtesis.pdf>
- [9] Nicolas D. Chétry. (2006, April) Computers Models for Musical Instrument Identification. [Online].
<http://www.elec.qmul.ac.uk/digitalmusic/papers/2006/Chetry06-phdthesis.pdf>
- [10] Tae Hong Park. (2004, November) Towards Automatic Musical Instrument Timbre Recognition. [Online].

<http://www.tulane.edu/~park/publications/ParkPrincetonThesis2004.pdf>

- [11] Tetsuro Kitahara. (2006) Computational Musical Instrument Recognition and its Application to Content-based Music Information Retrieval. [Online]. <http://www.winnie.kuis.kyoto-u.ac.jp/~kitahara.pdf>
- [12] Essid, Richard, and David. (2004, June) Efficient musical instrument recognition on solo performance using basic features. [Online]. <http://perso.telecom-paristech.fr/~grichard/Publications/aes25.pdf>
- [13] Agostini, Longari, and Pollastri. (2006) Content-Based Classification of Musical Instrument Timbres. [Online]. http://www.music.mcgill.ca/~ich/classes/mumt611_07/MIR/agostiniTimbre.pdf
- [14] Deng, Simmermacher, and Cranefield. (2007, August) A study of Feature Analysis for Musical Instrument Classification. [Online]. <http://otago.ourarchive.ac.nz/bitstream/handle/10523/957/dp2007-04.pdf?sequence=3>