Project : **CUIDADO**

Subject : **WP2.1.5 Psycho-acoustic Timbre Descriptors**

Object : Provide a set of numerical descriptors for timbre of musical sounds

Writer : **IRCAM**

V. Rioux, S. McAdams, P. Susini, G. Peeters

Control : **CUIDADO partners**

Date : **28/02/02**

Version : **V1**

Target : **CUIDADO partners**

Title : **Psycho-acoustic descriptors for Timbre Spaces**

Psycho-acoustic descriptors for Timbre Spaces

1° Introduction

"Until careful scientific work has been done on the subject, it can hardly be possible to say more about timbre than that it is multidimensional". This is all that Licklider, a prominent psychoacoustician, could assert back in 1951. Since then, a number of scientists have studied this multidimensional feature of timbre, but a consensual agreement on the number and nature of these dimensions is still to be found. In other words, this research field is still wide open.

Very recently, McAdams collected for the first time several major studies that provided perceptual timbre spaces (or "timbre spaces" for short)[McAdams & Winsberg 02]. These timbre spaces were all built according to a listening test paradigm consisting in asking subjects to rate the global timbral difference (or similarity) between all pairs of sounds in a chosen set. Hence this similarity data (highly dimensional) is analysed by means of Multidimensional Scaling Techniques which reduce the large number of similarities to a few number of relevant perceptual dimensions. Perceptual dimensions are generally supposed to correlate with models of human auditory perception, including purely signal-based models.

Following this idea, Peeters collected a number of these signal-based descriptors developed in the Music Perception and Cognition and Analysis-Synthesis teams at Ircam, and extended the available set to produce an operational toolbox [Peeters 00a]. Other descriptors available in the literature also include models of the functions of the neuro-mechanical auditory system. We hence also considered the descriptors sharpness (in acum) (Bismarck 74), roughness (in aspers) (Aures 85) and fluctuation strength(in vacil) (see Susini 00).

This extensive set of descriptors will be first reviewed in paragraph 2.

Then, we will examine possibilities and problems that arise when trying to exhibit an optimal set of generic descriptors suitable for the construction of a perceptual distance between pairs of sound samples. A proposal for such an "ideal" set will be made, together with a number of practical considerations and limitations.

2° Definition of the global descriptors pool

Several classification schemes can be applied to signal-based descriptors depending on the fact that their computation is based on time series, spectral series or under some hypotheses made on the sound itself: is it harmonic/periodic or percussive?

In the case of harmonic signal analysis, F0 estimation needs to be performed beforehand (De Cheveigné & Kawahara 02).

Table 1 Harmonic descriptors I (from Peeters [00])

std = standard deviation, spec = spectral/spectrum, amp = amplitude, lin = linear, norm = normalized
f0 = fundamental frequency, T= fundamental period

HARMONIC DESCRIPTORS

spectrum	
nrgb	energy
cgsb	spec centroid (global mean spec)
vspc	spec variation
stdb	spec spread
harmonic	
nrgh	spec energy
cgsh	spec centroid
stdh	spec spread
devs	spec deviation (of the harmonic computed from the global mean spectrum)
decs	spec slope
nrgi	mean of the instantaneous energy
cgsmax	spec centroid computed on the vector composed of the maximum amplitude [lin] of each harmonic over time
cgsmoy	spec centroid computed on the vector composed of the mean amplitude [lin] of each harmonic over time
cgsrms	spec centroid computed on the vector composed of the rms amplitude [lin] of each harmonic over time
cgsi	mean of the instantaneous spec centroid [amp lin, freq lin]
cgsidb	mean of the instantaneous spec centroid [amp dB, freq lin]
csgilog	mean of the instantaneous spec centroid [amp lin, freq log]
cgsidblog	mean of the instantaneous spec centroid [amp dB, freq log]
stdmax	spectral std computed on the vector composed of the maximum amplitude [lin] of each harmonic over time
stdmoy	spectral std computed on the vector composed of the mean amplitude [lin] of each harmonic over time
stdrms	spectral std computed on the vector composed of the rms amplitude [lin] of each harmonic over time
stdi	mean of the instantaneous spec std [amp lin, freq lin]
stdidb	mean of the instantaneous spec std [amp dB, freq lin]
stdilog	mean of the instantaneous spec std [amp lin, freq log]
stdidblog	mean of the instantaneous spec std [amp dB, freq log]

Table 2. Harmonic descriptors II (from Peeters [00])

devmax	spectral std computed on the vector composed of the maximum of amplitude [dB] of each harmonic over time
devmoy	spectral std computed on the vector composed of the mean of amplitude [dB] of each harmonic over time
devrms	spectral std computed on the vector composed of the rms of amplitude [dB] of each harmonic over time
devi	mean of the instantaneous spec deviation [amp lin]
devidb	mean of the instantaneous spec deviation [amp dB]
deci	mean of the instantaneous spec slope [amp lin]
decidb	mean of the instantaneous spec slope [amp dB]
flmax	spec flux using instantaneous spec centroid and cgsmax
flmoy	spec flux using instantaneous spec centroid and cgsmoy
flrms	spec flux using instantaneous spec centroid and cgsrms
fli	spec flux using instantaneous spec centroid and cgsi
vsph	harmonic spectral deviation
vsrate	speed of variation of the spectrum
magco (coherence)	sum of the variations of the instantaneous harmonic from global mean harmonics
hac	harmonic attack coherence

envelope

ltmr	log-attack time from [rms]
ltmm	log-attack time from [max]
ltmlr	log-attack time from [smoothed rms]
ltmlm	log-attack time from [smoothed max]
itmpn1	effective duration
itmpn2	effective duration [norm by file length]
itmpn3	effective duration [norm by file length and f0]
itmpn4	effective duration [norm by file length and T]

Table 3. Percussive descriptors (from Peeters [00])

lat	log-attack time
cgt	temporal centroid
stdt	temporal std
ed	effective duration
maximum	maximum value
mix	ed*cgt
LdB	rms value of the power spectrum
LdBA	rms value of the power spectrum [amp weighting dbA]
LdBB	rms value of the power spectrum [amp weighting dbB]
LdBC	rms value of the power spectrum [amp weighting dbC]
CGS	spec centroid of the power spec
CGSA	spec centroid of the power spec [amp weighting dbA]
CGSB	spec centroid of the power spec [amp weighting dbB]
CGSC	spec centroid of the power spec [amp weighting dbC]
STD	spec std of the power spec
STDA	spec std of the power spec [amp weighting dbA]
STDB	spec std of the power spec [amp weighting dbB]
STDC	spec std of the power spec [amp weighting dbC]
skew	skewness of the power spec
kurt	kurtosis of the power spec
slope	slope of the power spec

3° Timbre spaces and their Meta-analysis

Ten timbre spaces have been investigated . Timbre spaces studied by Grey (1977), Grey&Gordon (1978), used sounds resynthesized in a simplified form from analysed musical instrument sounds. The spaces derived by Krumhansl (1989) and McAdams et al. (1995) contained synthesized sounds (FM synthesis) imitating acoustic musical instruments or creating hybrids between them. Spaces studied by Iverson and Krumhansl (1993) and by Lakatos (2000) were more specifically concerned with recorded sounds of acoustic musical instruments.

A hundred sounds were hence gathered and analysed with the 48 harmonic descriptors, 21 percussive descriptors and 3 psychophysical descriptors (sharpness, roughness, fluctuation strength).

We would ideally like to expand the former attempt made by Misdariis et al. (1998) to develop a general distance model for perceptual dissimilarities among musical timbres.

3.1 Cluster analysis of the descriptors set.

Gathering all the data in a single data file (sounds vs computed descriptors), we performed a cluster analysis with the Ward method [Legendre] on the distance (euclidean) matrix composed of correlation coefficients between all pairs of descriptors. A representation of this analysis is available below.

Nine groups were formed and correspond globally to:

g1- spectral slope

g2- spectral centroïd

g3- spectral flux

g4- spectral spread (standard deviation)

g5- spectral deviation

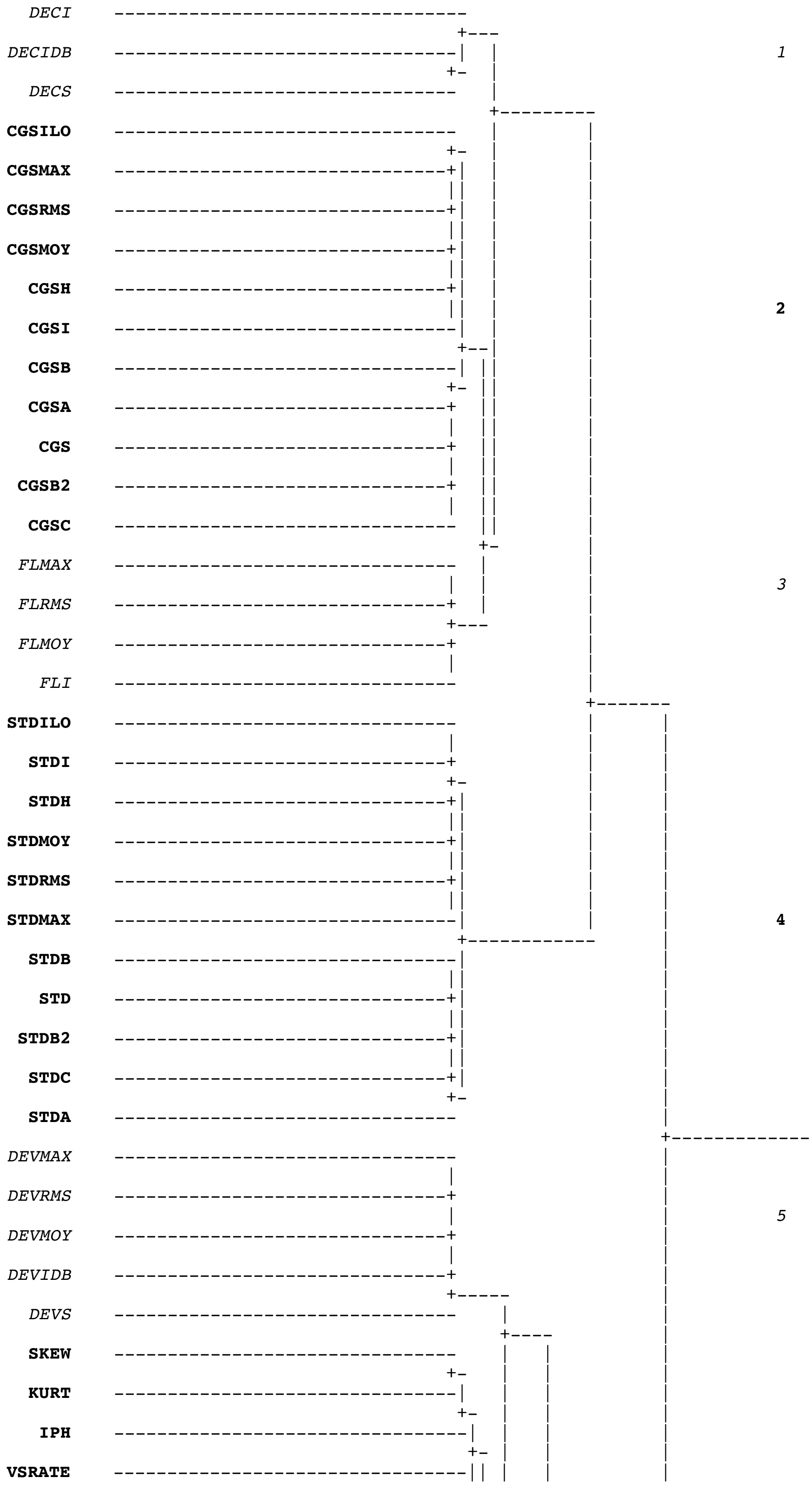
g6- spectral shape (kurtosis, skewness, slope)

g7- fluctuation/roughness

g8- rms power and energy

g9- attack time

This result confirms of course the classification presented in Figure 1.


DECI
DECIDB
DECS
CGSILO
CGSMAX
CGSRMS
CGSMOY
CGSH
CGSI
CGSB
CGSA
CGS
CGSB2
CGSC
FLMAX
FLRMS
FLMOY
FLI
STDILO
STDI
STDH
STDMOY
STDRMS
STDMAX
STDB
STD
STDB2
STDC
STDA
DEVMAX
DEVRMS
DEVMOY
DEVIDB
DEVS
SKEW
KURT
IPH
VSRATE
1
2
3
4
5

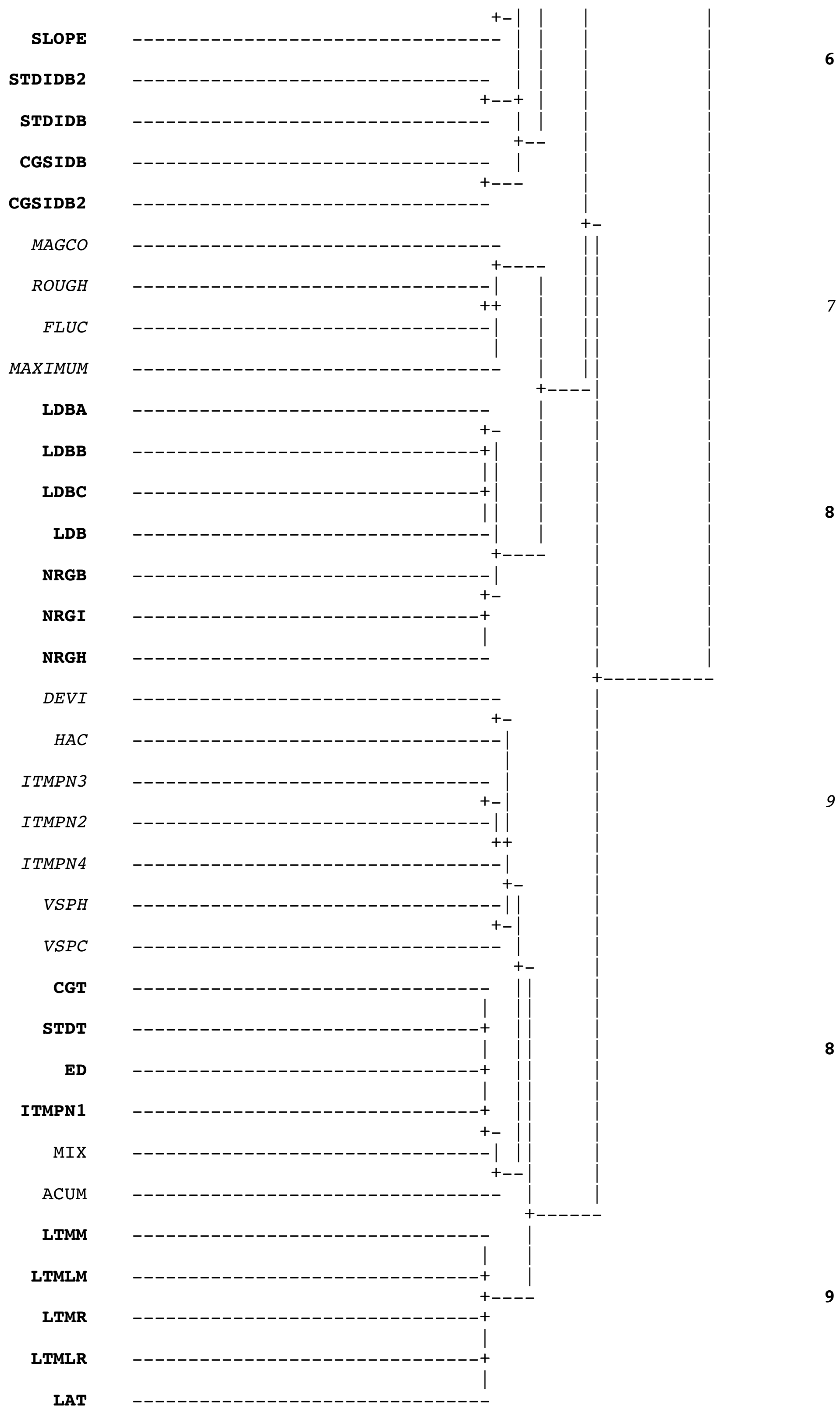


Figure 1. Cluster analysis of descriptors "similarities". The height at which two descriptors join n the tree (following the branches) reflects their degree of similarity.

3.2 Incremental multiple regression analysis

We performed a stepwise multiple regression analysis on each dimension of each perceptual timbre space. All these results are displayed in table 4.

Table 4. Stepwise multiple regression analysis performed on various timbre space dimensions with all descriptors. For each dimension, up to three descriptors can be used to explain most of the variance when combined linearly. The value s in parentheses are the cumulated regression coefficient. Explained variance is the square of this value.

Grey

G_3Ds

Dim 1	CGSA(0.882)	flrms(0.929)		spec. Centroid
Dim 2	acum(0.658)	decs(0.748)	vspc(0.877)	spec. flux
Dim 3	ltmlm(0.557)	devmax(0.825)	skew(0.891)	attack / spec. Deviation

Grey & Gordon

GG_3Ds

Dim 1	acum(0.309)	itmpn3(0.499)	stdt(0.643)	low correlation
Dim 2	cgsmax(0.884)	ltmm(0.911)		spec. Centroid
Dim 3	STDB(0.518)	devmax(0.758)	vsrate(0.908)	spec. Spread

Krumhansl

K_4Ds

Dim 1	ltmlm(0.683)	stdidb(0.815)	rough(0.883)	attack
Dim 2	LdBB(0.567)	*stdilo(0.674)*	iph(0.822)	energy
Dim 3	cgsi(0.822)	*stdilo(0.879)*	stdh(0.911)	spec. Centroid
Dim 4	STDA(0.396)	devs(0.567)	ltmlm(0.719)	spec. Spread

McAdams

M_3Dn

Dim 1	devs(0.430)	maximum(0.759)	fluc(0.819)	spec. Deviation
Dim 2	fluc(0.694)	*cgsidb(0.784)*	vspc(0.851)	fluctuation strength
Dim 3	hac(0.359)	*cgsidb(0.520)*	decs(0.735)	attack

Iverson

IKW_3Ds

Dim 1	ltmlm(0.594)	devs(0.738)	stdb(0.895)	attack
Dim 2	cgsilo(0.672)	itmpn3(0.792)	nrgi(0.837)	spec. Centroid
Dim 3	itmpn3(0.761)	fluc(0.910)	magco(0.955)	effective duration

Lakatos

LH_2Ds wind/string

Dim 1	itmpn2(0.896)	rough(0.937)		effective duration
Dim 2	LdB(0.675)	decs(0.828)	iph(0.876)	energy

LP_2Ds Percussions

Dim 1	cgt(0.804)	mix(0.908)	LdBA(0.952)	temporal centroid
Dim 2	rough(0.868)	CGSC(0.975)	maximum(0.988)	roughness

LC_3Ds Combined

Dim 1	cgt(0.790)	ltmr(0.916)	hac(0.949)	temporal centroid
Dim 2	rough(0.645)	stdt(0.770)	CGSC(0.858)	roughness
Dim 3	stdi(0.638)	flmoy(0.782)	cgsilo(0.872)	spec. Spread

From inspection of table4 we can conclude that from the original list of 71 descriptors, 18 descriptors can be extracted and classified according to tables 1 to 3 and figure 1 :

group2: spectral centroid : CGSA, cgsi, cgsilo, cgsmax

group4: spectral spread (standard deviation) : STDA, STDB, stdi

group5 : spectral deviation : devs

group8 : energy : LdBB, LdB

group9 : effective duration / attack time : itmpn2, itmpn3, ltmlm(2), hac, cgt

plus the psychopyhisical descriptors sharpness(2), roughness(2) and fluctuation strength.

Considering that the computation of sharpness (acum) is based on a center of gravity of the specific loudness (loudness in each auditory frequency sub-bands), we can reasonably consider that this descriptor belongs to group2. Moreover the last two descriptors are rather close making up a new group.

It is clear that the presence of group8 (related to the global energy of a signal and hence to its perceived loudness) is not really relevant when assessing timbre. In fact all psychoacoustic experiences concerning timbre generally follow the same paradigm of adjusting pitch, duration and loudness to a constant value across all the sounds of the set.

Conclusions

We examined various timbre spaces available in the literature and came up from a large number of descriptors (71) to a reduced and optimal set of 5 psychoacoustic descriptors which should be relevant for the spectral centroid, the spectral spread, the spectral deviation, the effective duration and attack time and finally roughness and fluctuation strength..

It is important to realize that as we quoted earlier, these studies are still experimental and should be used with care before trying to for example build a generalized timbral distance between any kind of sounds. It is obvious that the previous results were obtained with an a priori knowledge of the class of sounds that were used for the listening tests. Moreover results obtained by McAdams[2002] clearly show thath there also exist class of listeners. Some more work is definitely needed in order to assess the possibility of automatically finding such class of sounds and of listeners.

Acknowledgements

We wish to thank several contributors, namely Alain DeCheveigné, Jeremy Marozeau, and Olivier Houix.

References

[Aures 85] Aures, W. *A procedure for calculating roughness*. Acustica **58**.

[Bismarck 74] Bismarck, G. *Sharpness as an attribute of the timbre of steady sounds*. Acustica **30**.

[Chessel et al. 01] Chessel, D., Dufour A.B., Thioulouse, J. *Analyses procustéennes*. Lecture notes, available at: http://pbil.univ-lyon1.fr/R/fichestd/tdr55.pdf, In French.

[De Cheveigné & Kawahara 02] De Cheveigné, A., Kawahara, H. *YIN, a fundamental frequency estimator for speech and music*, accepted for publication in J. Acoust. Soc. Am.

[Grey 77] Grey, J.M. *Multidimensional perceptual scaling of musical timbres*. J. Acoust. Soc. Am. **61**, 1270-1277.

[Grey & Gordon 78] Grey, J.M and Gordon, J.W. *Perceptual effects of spectral modifications on musical timbres*, . J. Acoust. Soc. Am. **63**, 1493-1500.

[Herrera et al. 99] Herrera, P., Serra, X., Peeters, G. *Audio Descriptors and Descriptors Schemes in the Context of MPEG-7*. Proceedings of the ICMC1999.

[Iverson & Krumhansl 93] Iverson, P. and Krumhansl, C.L. *Isolating the dynamic attributes of musical timbre* J. Acoust. Soc. Am. **94**, 2595-2603.

[Jonhson & Wichern 98] Jonhson R. A., Wichern D. W. *Applied Multivariate Statistical Analysis*. Fourth Edition. Prentice Hall, 1998. 782-789.

[Krumhansl 89] Krumhansl, C.L. *Why is musical timbre so hard to understand?* in *Structure and Perception of Electroacoustics Sound and Music*, edited by S. Nielzen O. Olsson (Excerpta Medica, Amsterdam), 44-53.

[Lakatos] Lakatos, S. *A common perceptual space for harmonic and percussive timbres*, Percept. Psychophys.

[Legendre and Legendre 98] Legendre, P. and Legendre, L. *Numerical Ecology*, 2nd edition, Elsevier Science. Amsterdam, 1998.

[McAdams & Winsberg 02] McAdams, S., Winsberg, S. *A meta-analysis of timbre space. I: Multidimensional scaling of group data with common dimensions, specificities, and latent subject classes*. accepted for publication in J. Acoust. Soc. Am.

[Misdariis et al. 98] Misdariis, N., Smith, B., Pressnitzer, D., Susini, P., McAdams, S. *Validation of a Multidimensional Distance Model for Perceptual Dissimilarities among Musical Timbres*. Validation 16th Int. Congress on Acoustics, 135th meeting ASA, Seattle, USA, 20th -26th June 1998.

[Peeters et al. 00] Peeters, G., McAdams, S., Herrera, P. *Instrument Sound Description in the Context of MPEG-7*. Proceedings of ICMC2000 (International Computer Music Conference), Berlin, Germany, August 27th - September 1st, 2000.

[Peeters 00] Peeters, G. *Timbre Toolbox*. Set of Matlab routines computing timbral descriptors developed at Ircam.

[R project] The R Project for Statistical Computing. http://www.r-project.org/

[Susini & McAdams 00] Susini, P., McAdams, S. *L'approche psychoacoustique de la qualité sonore*. internal report.