Audio Engineering Society

Convention Paper

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

Morphological sound description: computational model and usability evaluation

Julien Ricard and Perfecto Herrera

IUA-MTG, Pompeu Fabra University, Barcelona, 08003, Spain
jricard, pherrera@iua.upf.es

ABSTRACT

Sound samples metadata are usually limited to a source label and several related textual labels. In the context of sound retrieval this makes the retrieval of sounds having no identifiable source (‘abstract sounds’) a hard task. We propose a description framework focusing on intrinsic perceptual sound qualities, based on Schaeffer’s research on *sound objects,* which could be used to represent and retrieve abstract sounds and to refine traditional search by source for non-abstract sounds. We show that some perceptual labels can be automatically extracted with good performance, avoiding the time-consuming manual labelling task, and that the resulting representation is evaluated as useful and usable by a pool of users.

1. INTRODUCTION

Sound databases are used by many sound professionals or amateur for music composition or movie postproduction. Rapid development of information storage technology allows building larger and larger databases, making the retrieval of a specific sound a hard task.

Typical sound retrieval commercial systems are usually source-centred, which means that retrieval is based on using the proper keyword or selecting the proper category that defines or specifies a sound source. In that context, sounds having no identifiable source (‘abstract sounds’) can hardly be retrieved. Moreover, labels used for this type of sounds (electronic, weird, FX…) are often not consistent and vary from one company to another. Some systems also offer search by keywords or categories related to emotion (scary, happy, sad...), but they are only applicable to specific sounds. More complex systems allows retrieving sounds using perceptual or acoustical attributes such as pitch, loudness and timbre features, by similarity or by specifying classes based on these features [1,2]. Some systems offer visual representation of audio features for helping retrieving a sound [3], [4].

Since sound retrieval has a strong commercial potential, a lot of research has been done in order to improve current systems. In [5] Zhang describes a system using features based on timbre and rhythm, and Hidden Markov Model for classifying sounds into semantic classes such as *rain*, *dog bark* or *applause*. Slaney [6] describes a system which links points in a semantic space to points in an acoustic space using a mixture of probabilistic experts. This system allows describing automatically a sound by words or to retrieve a sound using natural language. In real world application, these systems would require thousands of models, and complete manual labelling of very specific sound (e.g. ‘Water, Lap, Rapid, Trickle’) is very time-consuming.. Moreover, they are still source-based and are then poorly adapted to abstract sounds.

An alternative sound representation, based on ecological perception, is proposed by Casey in [7]. He used ICA for decomposing time-frequency distribution in order to characterize structural invariants (such as material acoustical properties) and transformational invariants (acoustical properties of action such as bouncing, creaking or rolling) of sound objects. Similar methods are used in the Sounding Object project [8]. Separating the *material* from the *form* greatly reduce the number of models required for describing the sound world, but this number remains very large and this model seems poorly adapted to abstract sounds, which can have an infinite number of matter and form.

The representation we will describe follows the same idea by further decomposing matter and form in perceptual dimensions. It is based on work done by Pierre Schaeffer[1] on sound object and focus on intrinsic perceptual sound qualities. He defined several simple and independent dimensions such as the shape of the amplitude envelope or the 'pitchness' (whether a sound is rather pitched or noisy) which are not related to any 'external' properties (causal, semantic or subjective) and that can then be applied to any sounds (abstract or not). Each dimension is divided into typical cases using words (e.g. for the amplitude envelope dimension: unvarying, impulse, crescendo...) that could provide simple and intuitive search criteria for abstract sounds retrieval. This representation could also be used for refining a traditional search by source for non-abstract sounds.

[1] French composer and writer, who created one of the first research groups in music technology, the Groupe de Recherche Musicale, within the French national radio.

2. SCHAEFFER’S TYPO-MORPHOLOGY

In his *Traité des objets musicaux* (Treatise on musical object) [9] (synthesised and commented by Chion in [10]), Pierre Schaeffer proposes a generalization of what is usually heard as musical sounds (typically notes generated by traditional musical instruments) by considering all kind of *sound objects,* ignoring their origin (electronic sounds, noise, environmental sounds, loops…), for which the traditional musical sound representation source/pitch/duration is way too limited. After some listening experiments, he proposed a general sound classification (*typology*) according to some *morphological* criteria, in order to build a *solfège of musical objects*, foundation of musique concrète[2].

2.1. Sound object

In his research for a generalised solfège, Schaeffer defined the *sound object* as the element of study. The sound object is the correlate of a *reduced listening*, during which sounds are listened to for their intrinsic perceptual qualities, independently from their meaning or their origin. In a sound stream, any entity perceived as having its own internal properties and rules is considered as a sound object. In a piece of music, for example, a sequence of notes can be perceived as a single entity, a musical phrase, as well as the succession of smaller sound objects, i.e. the notes themselves[3]. The sound object is the result of a particular intention, for which any sound is listened to the same way, providing a good basis for general sound description.

2.2. Typo-morphology

In order to describe and classify sound objects, Schaeffer had to found criteria for comparing them in the context of reduced listening. Starting from rough classification and description, he built a sound *typo-morphology*, in which sounds objects are classified according to their morphological attributes into types. Typology refers to the general type (classification) of

[2] Musique concrète starts from concrete sound material arranged in such ways that some music emerges from it, as opposed to musique abstraite, which starts from an abstract representation, the score, and is played later.

[3] For a study on perceptual sound organization, see Bregman’s Auditory Scene Analysis [11].

sounds while morphology refers to their characteristics (description). The building of this typo-morphology is based on the pair of criteria *shape/matter*. The sound matter is defined by Schaeffer as *what we would hear if we could freeze the sound* (then mainly –but not only- related to spectral distribution), while the shape is related to the time evolution of this matter. These criteria were studied by listening to sounds with fixed matter, allowing studying the form, and sounds with fixed form, allowing studying the matter. Varying sounds, in which both the form and the matter vary, are also studied through the *variation* criteria.

By refining the rough sound description and classification obtained through these two criteria, Schaeffer defined seven morphological criteria related to different perceptual dimensions emerging from reduced listening:

Matter criteria:

- *Mass*: related to the perception of the 'pitchness' of a sound, and then to its spectral distribution. Schaeffer defines four types of mass: *pitched* (fixed mass and identifiable pitch), *complex* (fixed mass and non-identifiable pitch), *varying* (*pitched-varying* or *complex-varying*, for small or organized variation) and *nondescript* (excessive and unpredictable variation).

- *Harmonic timbre*: *the more or less diffuse halo associated to the mass and more generally what allows describing it*. We interpreted this definition as a finer characterisation of the mass, often described by analogy to vision: bright/mat, round/sharp…

- *Grain*: defined as the microstructure of sound matter, such as the rubbing of a bow. Even though it has a temporal dimension, is a matter criterion. It is divided into three types: *resonance* grain, for non-sustained sounds (e.g. cymbal resonance), *rubbing* grain, for sustained sounds (e.g. bow or breath sounds) and *iteration* grain, for iterative sounds (e.g. drum roll).

Shape criteria:

- *Dynamics*: Schaeffer distinguished seven types of energy temporal evolution (none, low, shaped, impulse, cyclic, iterated and accumulated).

- *Allure*: amplitude or frequency modulation. Three types: mechanical (very regular), lively ("flexible periodicity, revealing a living being") and natural (unpredictable).

Variation criteria:

- *Melodic profile*: variation of the whole mass (fixed mass, typically a melody played by a traditional instrument). Nine types, according to the variation type ("imperfect stability", continuous or discontinuous variation) and speed (slow, medium or fast) and four classes, according to the profile shape (crescendo, delta, inverse delta, decrescendo).

- *Mass profile*: variation within the mass. Three types of typical mass, according to the variation types described above, and four classes representing typical thickness evolution (from thin to thick, from thick to thin, thickening followed by thinning and thinning followed by thickening).

Schaeffer typo-morphology provides a description scheme that could be of great interest for sound retrieval issues. His attempt to describe sound according to intrinsic properties, rather than relating it to the external information it transmits, allows describing any type of sounds. It could be the basis of a general sound description scheme, and is of particular interest for abstract sounds, i.e. sounds that have no identifiable real or imagined origin. They generally have been generated by synthesis or transformation and are typically labelled in audio libraries as effect, electronic, weird, noise, atmosphere… Moreover, the decomposition of perceptual dimensions of sounds could allow searching for a particular quality of sound. This idea is enforced by the increasing efficiency of sound transformation tools, which permit modifying separately a single quality of a sound. Ideal pitch shifting and time stretching algorithms could allow, for example, using a single musical note to create an entire musical phrase[4]. However, if we want to apply Schaeffer's approach to practical sound retrieval systems, some computational counterpart of the *typo-morphology* presented above should be found.

3. COMPUTATIONAL MORPHOLOGICAL DESCRIPTION

There has been very little research on automatic morphological description. The main work done in that

[4] An overview of content-based transformation is given in [12].

area was the ECRINS project on sound samples audio content description [13]. A morphological description scheme is defined, in which sounds are described according to the following descriptors: dynamical profile (amplitude evolution), melodic profile (pitch evolution), pitch, spectral distribution, and sound location. Simple classes are automatically detected and the description can be refined manually by the user. The classification allowed by the automatic description is quite limited and the description scheme only includes pitched sounds. It seems possible to complete it by further analysis and by adding new description criteria.

In an earlier paper [14], we showed that a few specific low-level descriptors allow classifying sounds in a simplified 3-dimensional morphological representation (mass, mass profile and dynamics) with good performance. We have been continuing this work by adding some classes to the dynamics dimension and investigating the melodic profile dimension.

For usability's sake, we have changed the name of the morphological criteria. From now on, we will talk about dynamic profile instead of dynamics, pitchness instead of mass and pitch profile instead of melodic profile.

3.1. Current description scheme

The current morphological description scheme is the following:

Morphological Criteria	**Types**
Dynamic profile (7 types)	*Unvarying* *Varing*: -*Impulse* -*Iterative* (several transients) -*Crescendo* -*Decrescendo* -*Delta* (Crescendo-Decrescendo) -*Other*
Pitchness (3 types)	*Pitched* (one predominant pitch) *Complex* (simultaneous pitched components or simultaneous or sequential pitched and noisy components) *Noisy*
Pitchness profile (2 types)	*Unvarying* *Varying* (e.g. from noisy to pitched)
Pitch profile (3 types)	*Unvarying* *Varying*: -*Continuous* (e.g. siren) -*Stepped* (e.g. piano phrase)

Table 1 Current morphological description scheme.

3.2. System description

The system block diagram is shown in figure 1.

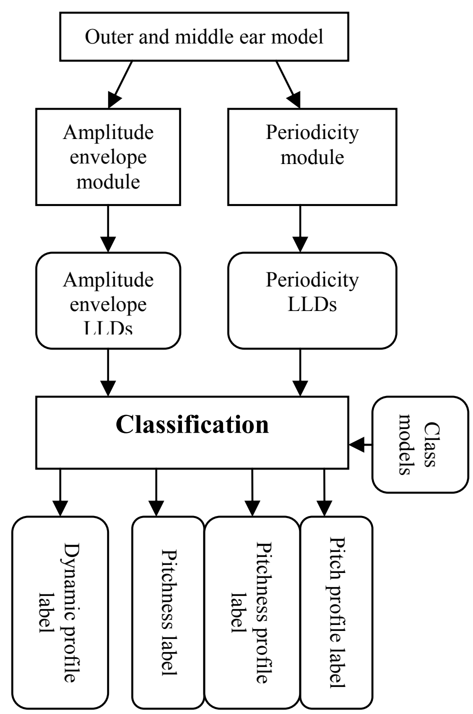

Figure 1 System block diagram.

We considered each sound sample as a single sound object, i.e. no segmentation was performed. The audio signal is filtered by a model of the outer and middle ear [15] and some amplitude envelope and periodicity low-level descriptors (LLDs) are then computed by two independent modules.

3.2.1. Amplitude envelope module

The amplitude envelope is estimated by applying a simple one-pole low-pass filter with a time constant of 5ms on the full-wave rectified waveform. The descriptors are computed from the log-amplitude envelope (for rough modelling of loudness perception) and were chosen intuitively, according to specificity of each class. Example of 'intuitive' choice: if one wants to classify sounds according to three types of amplitude envelope shape, unvarying, crescendo and decrescendo, one might assume that some amplitude variance measure and some 'balance' measure (i.e. whether the centre of gravity of the amplitude envelope is rather at the middle, at the left or at the right) should be used. Initial intuitive selection of descriptors was filtered by the classification algorithm.

3.2.2. Periodicity module

Pitchness and pitch are computed in the time domain by autocorrelation analysis. Short term autocorrelation is computed in 10 ERB-spaced frequency bands from 100 to 10000Hz and summed over frequency.

In periodic signals, autocorrelation peaks occur at integer multiples of the periods. The amplitude of the peaks is a measure of the pitchness of a sound. Pitchness is computed from pitch salience, defined by Slaney [16] as the ratio of the magnitude of the highest autocorrelation peak to the magnitude of the 0-lag peak. We used the mean of short-term pitch salience weighted by the amplitude of the frame for determining the global pitchness of the signal. Highly pitched sounds will have a pitch salience close to 1 (pure tone), unpitched sounds will have a pitch salience close to 0 (white noise). The pitchness profile is computed from pitch salience variance.

The pitch is computed from the highest peak in the autocorrelation for pitched sounds only. In order to achieve better pitch estimation, autocorrelation peaks are tracked over time and some post processing correct local errors. The log pitch variance as well as some pitch 'step' (discontinuity followed by flat pitch) detection are then used for pitch profile classification.

All the LLDs used for morphological classification are listed in Appendix.

3.2.3. Evaluation of the models

In order to evaluate our system, we built a small database (around 200 sound samples for each class) in which sounds were manually labeled according to the classes defined earlier. Classes of each morphological criterion were modeled by simple C4.5 decision tree (the other classifiers tried did not perform significantly better and decision trees give a clear view on how features are related to classes) and tested by 10-folds cross-validation.

Amplitude envelope were modeled for each defined class (one against all), excepted for the *Varying-Other* class, and the models were tested on the whole database, including those labeled as *Varying-Other*. Classification performance was very good (around 98%) for the *Unvarying* and *Varying-Crescendo* classes, as expected because of the specificity of these classes, around 94% for *Varying-Impulse* and *Varying-Decrescendo* classes (most misclassified instances in one class was labeled as belonging to the other) and around 88% for the *Varying-Delta* class. The poor performance, compared to the other classes, of *Varying-Delta* amplitude envelope classification might be explained by the high variability of this class (close to crescendo, symmetric or close to decrescendo).

Pitchness discrimination between *Noisy* and *Pitched* sounds is close to 100%. Some errors occur when the pitch is too low to be detected by our current analysis settings (basically a too short analysis frame size): no peak is detected in the periodogram and the sound is then classified as *Noisy*. This will be solved by using an adaptive analysis frame size. The main problem is the classification of sounds having *Complex* pitchness (simultaneous pitched components, or simultaneous or sequential pitched and noisy components): only 50% are correctly classified, the rest being classified as noisy or pitched, in the same proportion. Two issues here: whereas the *Noisy*/*Pitched* sounds discrimination is quite obvious, the perceptual boundary between *Noisy* and *Complex* and that between *Complex* and *Pitched* classes is not clear. Moreover, measuring the global pitchness of *Complex* sounds might not be meaningful and it would probably be much better to analyze each component separately and to perform the classification according to their number and their type (e.g. 'sequence of pitched and noisy components' or 'simultaneous pitched components').

Pitchness profile classification performance is 70%, which is not very good for 2-classes discrimination. The LLDs used (pitch salience variance and flux) are probably not appropriate or sufficient. It works very well for sounds having a clearly unvarying (e.g. white noise or pure sinusoid) or a clearly varying pitchness (e.g. sequence of white noise and pure sinusoids), but many noisy sounds that have a very varying pitch salience are perceived as having an unvarying pitchness, i.e. as being an unvarying noise. Smoothing the pitch salience did not change the performance.

Pitch profile classification error rates are tightly related to automatic pitch estimation performance. Sounds having an unvarying pitch are correctly classified for 90% of the instances. Varying sounds are further classified as *continuous* or *stepped* with a performance of 85%. When the pitch is correctly estimated, pitch profile classification is almost perfect.

3.2.4. Prototype sound search engine

In order to test the usability of morphological representation for sound retrieval, we integrated it to a prototype system developed for the AudioClas project [17], which makes use of two sound representation levels (perceptual and semantic) for managing large sound databases. Sound effects were automatically labelled according to the description scheme described above, and morphological criteria were added to the search interface.

4. USEFULNESS AND USABILITY EVALUATION

A very important issue when dealing with sound representation is the evaluation of its usability in a given context. We evaluated the usability of morphological representation for sound retrieval for abstract and non-abstract sounds through an online questionnaire, including sound examples and tests on the prototype application described above.

4.1. Material, procedure and subjects

After an introduction on morphological description and our system (including sound examples), users were asked to label 10 sounds according to our current description scheme (table 1) in order to measure the percentage of agreement and to confirm our assumption that morphological description is essentially listener-independent.

In the second part, users had to listen to three abstract sounds and to retrieve them in a database of 100 abstract sounds using the prototype sound search engine. They could use traditional keywords and/or morphological classes. They were then asked whether they think that morphological representation was useful as search criteria for such sounds and if yes, whether it was useful as a main representation or as a complementary representation to traditional source-related keyword description.

The third part consisted in the same test for non-abstract sounds.

In the final part users were asked more general questions on the understandability of the current dimensions and their names and on the completeness of the description of each dimension (i.e. are there enough classes to describe each dimension?). Some comments were also gathered during informal discussions.

The questionnaire was answered by researchers in music technology, musicians and sound technicians. 14 people answered the listening test, from who 10 answered the whole questionnaire. Most of the subjects (13) had medium or high musical training. Half of them used sound databases for professional purpose. Half of the subjects had someone working on the project sat beside.

4.2. Results

In order to measure how much listener-dependent was each dimension, we calculated a percentage of agreement, given by the number of sounds classified as the system did divided by the total number of answers (14 subjects*10 sounds) multiplied by 100. The results for the listening test are the following:

- 71% agreement for the amplitude envelope type. Typical disagreements were between following classes: Delta / Crescendo or Decrescendo, Impulse / Decrescendo and Iterative / Any class. The two first disagreement types are due to the unclear boundaries between these classes. A sound having a Delta shaped amplitude envelope with a Crescendo part much shorter than the Decrescendo part could be perceived as globally Decrescendo. Disagreements between the Iterative class and the other classes are all due to the fact that some people perceive the global envelope shape rather than each smaller entities a sound is made of (e.g. increasingly strong knocks on a door can be perceived as Iterative or Crescendo).

- 78% agreement for the pitchness type. As expected, disagreements existed mainly between Pitched / Complex and Complex / Noisy classes. Once again disagreements are due to the unclear boundaries between the classes. Only one sound, a bouncing ping pong ball (sequence of very short impulses) was classified as both Noisy (11 subjects) and Pitched (3 subjects).

- Since pitch profile is (in our current system) only defined for pitched sounds, the result was calculated only for such sounds, which gives an agreement of 83%. Most disagreements happened in one sound made of two successive impulses having the same pitch (bike bell rings). Four subjects classified this sound as having a varying-stepped pitch profile, probably because of the iterative amplitude envelope type.

- 73% agreement for pitchness profile. Despite the weakness of the model (see the notes on the evaluation of the model for this dimension), this result is not as bad as expected. As for the other dimensions, sounds having clearly varying or clearly unvarying pitchness profile were correctly classified with much higher agreement that for ambiguous sounds.

In the test on abstract sounds retrieval, all (11 subjects) considered morphological labels as useful. Six though that they should be use on their own (i.e. as a primary representation) and 5 though that they should be used in combination with traditional labels. Comments and suggestions included the possibility to describe the amount of pitched components vs. noisy components in sound having complex pitchness, the addition of a dynamic profile sub-category for specifying the attack (smooth, steep…) and the need for timbre description.

In the test on non-abstract sounds retrieval, all (10 subjects) considered morphological labels as useful in combination with traditional labels. In addition to those described above, more comments and suggestions were done, including the addition of more amplitude envelope types (it was not specified what could be added) and some classes related to instrumental practice.

The last part dealt with the understandability and the completeness of each dimension. All found that the *pitchness* is understandable and 3 found that it should be further described. One subject suggested distinguishing harmonic and inharmonic sounds in the pitched class. Dynamic profile is well understood by all subjects, but one suggested that 'amplitude envelope shape' would be more comprehensible. Four subjects though that this dimension should be further described. Suggestions included adding tremolo and attack description as sub-dimensions, adding the possibility to combine the Iterative class with another one (e.g. Iterative-Crescendo) and adding a tool for drawing any envelope. Comments were done about the ambiguity between Delta and Crescendo or Decrescendo for some sounds. Pitchness profile is not well understood by 2 subjects. One pointed out that having three dimensions sharing the word *pitch* is misleading and the second commented that a sound could have two simultaneous unvarying and varying components. One subject also found that pitchness and pitchness profile were incompatible since if a sound is classified as noisy or pitched (this is not true for complex sounds), it should have an unvarying pitchness profile. Five subjects found that this dimension should be further described and suggested to add some typical profile, such as pitched to noisy. Pitch profile is not well understood by 2 subjects, one because of the use of the word *pitch* (see above) and the second because he did not understand well the class *Varying-Continuous*. Four subjects found that this dimension should be further described by adding some pitch contour classes (increasing, decreasing…).

Surprisingly, only one subject suggested adding one morphological dimension, timbre (with no more details), to the current scheme.

No correlation was found between the results and the musical training or the use of sound databases for professional purpose.

4.3. Discussion

These results show that morphological labels are useful for retrieving abstract sounds as a primary or complementary representation and for non-abstract sounds as a complementary representation. Some useful comments were done in order to improve the system.

The main problem seems to be that exclusive classes often lead to misclassification for ambiguous sounds. This could be solved by using probabilistic models or by using continuous scales instead of discrete classes when applicable (e.g. pitchness). It also seems sometimes difficult for users to perceptually separate the different morphological dimensions (see the bike bell rings example in results of the listening test for pitch profile). Since this way of listening (called *reduced listening* by Schaeffer) is unnatural, we assume that some more training would be sufficient to be able to focus on only one dimension. Suggestions are being realized (we are adding some pitch contour classes) or will be further analyzed. Analyzing each component in a complex objects, which would require some sound separation, and adding a tool for drawing any profile are not planned yet. Nevertheless, we think that more dimensions should definitely be added and we are currently investigating sound roughness.

5. FURTHER WORK AND GENERAL CONCLUSION

Our current system allows automatically labelling sounds according to a description scheme based on independent perceptual sound attributes called morphological criteria. A usability evaluation showed that users consider morphological representation as rather useful and usable in the context of sound effects retrieval, especially when dealing with abstract sounds. These results are encouraging and we are working on completing the current representation by investigating more morphological dimensions and classes. Apart the suggestions gathered during the evaluation, which we are further considered, we are currently working on roughness and further iterative sounds description.

6. ACKNOWLEDGEMENTS

This work was partially supported by the IST project CUIDADO and the EUREKA AudioClas project E! 2668. I sincerely thank people that answered the questionnaire for their useful comments and suggestions, Markus Koppenberger for the sound retrieval system interface and Pedro Cano and Nicolas Wack for their technical help.

7. APPENDIX: LOW-LEVEL DESCRIPTORS USED FOR MORPHOLOGICAL CLASSIFICATION

Specific low-level descriptors (LLDs) were used for each morphological dimension.

7.1. Dynamic profile

All LLDs were computed on the log-amplitude envelope:

- Flatness coefficient: ratio of the value above which lie 5% of the values to the value above which lie 80% of the values.
- Number of onsets, detected by looking for peaks above a threshold on the amplitude envelope derivative.
- Maximum amplitude time to total length ratio
- Derivative average after the maximum.
- Maximum derivative before the maximum.
- Temporal centroid to total length ratio.

7.2. Pitchness

We only used the average of the short-term pitch salience weighted by the short-term energy.

7.3. Pitchness profile

We used the pitchness flux and variance.

7.4. Pitch profile

All LLDs were computed on the log-pitch envelope. Some segmentation was performed by detecting peaks above a threshold in the pitch envelope acceleration. The following LLDs were then used:

- Global pitch envelope variance.
- Average of the log pitch variances computed in each segment.
- Average of the log pitch fluxes computed in each segment.

8. REFERENCES

[1] E. Wold, T. Blum, D. Keislar, and J. Wheaton. *Content-based classification, search and retrieval of audio.* IEEE Multimedia, vol. 3, no. 2, pp. 27–36, 1996

[2] http://www.musclefish.com

[3] http://www.findsounds.com

[4] G. Tzanetaki and P. Cook. *3D Graphics Tools for Isolated Sound Collections*. In Proc. Conference on Digital Audio Effects (DAFX), Verona, Italy (2000)

[5] T. Zhang and C.-C. Jay Kuo. *Classification and retrieval of sound effects in audiovisual data management*. In Proceedings of 33rd Asilomar Conference on Signals, Systems, and Computers (1999)

[6] M. Slaney. *Mixtures of probability experts for audio retrieval and indexing*. In Proc. IEEE International Conference on Multimedia and Expo (2002)

[7] M. A. Casey. *Auditory Group Theory with Applications to Statistical Basis Methods for Structured Audio*. PhD thesis, M.I.T. (1998)

[8] http://www.soundobject.org

[9] P. Schaeffer. *Traité des* objets *musicaux*. Paris: Seuil (1966)

[10] M. Chion. *Guide des objets sonores* Buchet/Chastel, Paris (1983)

[11] A. S. Bregman. A*uditory Scene Analysis.* MIT Press, Cambridge, MA (1990)

[12] X. Amatriain, J. Bonada, A. Loscos, J. Arcos and V. Verfaille. *Content-based Transformations.* Journal of New Music Research Vol.32 .1 (2003)

[13] Y. Geslin, P. Mullon, and M. Jacob. *Ecrins : an audiocontent description environment for sound samples*. In Nordahl, M. (Ed.), Proc. of ICMC 2002, Göteborg, Sweden.

[14] J. Ricard and P. Herrera. *Using morphological description for generic sound retrieval*. Proceedings of ISMIR 2003. Baltimore, Maryland, USA

[15] P. Kabal. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality.* TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University, May 2002 (updated Dec. 2003)

[16] M. Slaney. *Auditory toolbox: a Matlab toolbox for auditory modelling work.* Apple Computer technical report #45 (1994)

[17] P. Cano, M. Koppenberger, S. Le Groux, P. Herrera and N. Wack. *Perceptual and semantic management of sound effect*. Submitted to the 2004 IEEE International Conference on Multimedia an Expo (ICME 2004)