SEMI-AUTOMATIC AMBIANCE GENERATION

P. Cano, L. Fabig, F. Gouyon, M. Koppenberger, A. Loscos[†] *A. Barbosa*[† ‡]

[†] Institut Universitari de l'Audiovisual
Universitat Pompeu Fabra, Barcelona, Spain
pcano@iua.upf.es

[‡] Dept. Som e Imagem
Universidade Católica, Porto, Portugal
abarbosa@porto.ucp.pt

ABSTRACT

Ambiances are background recordings used in audiovisual productions to make listeners feel they are in places like a pub or a farm. Accessing to commercially available atmosphere libraries is a convenient alternative to sending teams to record ambiances yet they limit the creation in different ways. First, they are already mixed, which reduces the flexibility to add, remove individual sounds or change its panning. Secondly, the number of ambient libraries is limited. We propose a semi-automatic system for ambiance generation. The system creates ambiances on demand given text queries by fetching relevant sounds from a large sound effect database and importing them into a sequencer multitrack project. Ambiances of diverse nature can be created easily. Several controls are provided to the users to refine the type of samples and the sound arrangement.

1. INTRODUCTION

The audio component of audiovisual productions has long been regarded as of minor importance. Nevertheless, in the last years and especially after productions such as Apocalypse Now (1979), its importance has been acknowledged. Audio is gaining interest for its evocative and overall immersive experience for the audiences. Audio has an immense power, even when accompanying coarsely-drawn cartoons, for creating the illusion of reality.

Traditionally, from the film production process point of view, sound is broken into a series of layers: dialog, music and sound effects—from now SFX [1]. SFX can be broken further into *hard SFX* (car doors opening and closing, and other foreground sound material) and *foley* (sound made by humans, e.g: footsteps) on the one hand, and *ambiances* on the other hand. Ambiances—also known as atmospheres—are the background recordings which identify scenes aurally. They make the listener really feel like they are in places like an airport, a church, a subway station, or the jungle. Ambiances have two components: The *ambient loop*, which is a long, streaming, stereo recording, and *specifics* or *stingers*, which are separate, short elements (e.g: dog barks,car horns, etc) that trigger randomly to break up repetition [2].

Sound engineers need to access sound libraries for their video and film productions, multimedia and audio-visual presentations, web sites, computer games and music. Access to libraries is a convenient alternative to sending a team to record a particular ambiances (consider for instance "a Rain forest" or "a Vesuvian eruption". However, the approach has some drawbacks:

1. Accessing the right ambiances is not easy due to the information retrieval models, currently based mainly on keyword search [3].

2. The number of libraries is large but limited. Everybody has access to the same content although sound designers can use them as starting point and make them unrecognizable and unique.

3. The ambiances offered by SFX library providers are already mixed. There may be SFX in the mix that the sound engineer does not want in that position of may be does not want at all. It is a hassle to fix it.

In this context, we present a system for the automatic generation of ambiances. In short, the system works as follows: the user specifies his need with a standard textual query, e.g: "farm ambiance". The ambiance is created on-the-fly combining SFX related to the query. For example,the query "farm ambiance" may return "chicken", "tractors", "footsteps on mud" or "cowbells" sounds. A subset of retrieved sounds is randomly chosen. After listening to the ambiance, the user may decide to refine the query—e.g: to remove the "cowbells" and add more "chickens"—, ask another random ambiance—with a "shuffle-type" option—or decide that the ambiance is good enough to start working with. The system outputs the individual SFX samples in a multitrack project.

The intended goals of the approach can be summarized as follows:

Enhance creativity: Sound engineers have access to a huge ever-changing variety of ambiances instead of a fix set of ambiances. The combination of individual SFX provides a substantially larger number of ambiances.

Enhance productivity: Engineers can have several possible sonifications in a short time.

Enhance flexibility: Having different SFX of the ambiance separately in a MultiTrack gives more flexibility to the ambiance specification process, some sounds–a bird singing in a forest ambiance–can be removed or their location in the time line changed. It also allows for spatialization using 5.1.

Enhance quality: With a very low overhead—basically clicking on a "shuffle" button and adjusting some sliders, sound engineers can obtain several ambiance templates. Hence, the production cycle reduces. The producers can give their feedback faster and their opinions be incorporated earlier in the production improving the overall quality.

2. SYSTEM DESCRIPTION

The system is based on a concept-based SFX search engine developed within the AudioClas project (www.audioclas.org). The objectives of the project were to go beyond current professional SFX provider information retrieval model, based on keyword-matching, mainly through two approaches [4]:

Semantically-enhanced management of SFX using a general ontology, WordNet [5][1].

Content-based audio technologies which allow automatic generation of perceptual meta data (such as prominent pitch, dynamics, beat, noisiness).

These two approaches are the building blocks of the semi-automatic ambiance generation. Current prototype uses 80.000 sounds from a major on-line SFX provider[2]. Sounds come with textual descriptions which have been disambiguated with the augmented WordNet ontology [3]. WordNet is a lexical database that, unlike standard dictionaries which index terms alphabetically, indexes concepts with relations among them.

```
thrush  (songbirds characteristically having ...)
 => oscine, oscine bird
  => passerine, passeriform bird
   => bird
    => vertebrate, craniate
     => chordate
      => animal, animate being...
       => organism, being
        => living thing, animate thing
         => object, physical object
          => entity, physical thing
```

[1] http://www.cogsci.princeton.edu/~wn/

[2] http://www.sound-effects-library.com

Accordingly, the sound "Thrush And Nightingale Various Calls" becomes labeled with the following set of concepts:

```
01234719%n thrush -- (songbirds having
 brownish upper plumage with a spotted breast)
01237641%n nightingale, Luscinia megarhynchos
  -- (European songbird noted for its melodious
 nocturnal song)
05680983%n birdcall, call, birdsong, song --
 (the characteristic sound produced by a bird)
```

The numbers before the definitions correspond to the unique identifiers, *offsets*, of the concepts, or synonym sets, *synsets* as referred in the WordNet literature [5].

There are two main functional blocks in the system. The first one retrieves the relevant sounds of the SFX Database and a second one organizes the sounds in a multitrack according to some heuristic rules (see figure 1).

3. SOUND SELECTION AND RETRIEVAL

The first step has been mining ambiance sounds to learn the type of sources used. We use a database of SFX that has been labeled with concepts rather than with words (see [3] for details). We are therefore able to study the co-occurrence of concepts in sounds. For example, the ambiance "Farm Ambiance Of Rooster And Hen With Wagtail In Background" has been converted to:

```
01466271%n hen, biddy -- (adult female chicken)
01206115%n wagtail -- (Old World bird having a very
 long tail that jerks up and down as it walks)
02893950%n farm -- (workplace consisting of farm
 buildings and cultivated land as a unit)
```

By mining this information we learn that farm is related to the concept hen and the concept wagtail. Moreover, there are relations encoded in WordNet, which knows that hen and chicken are related. Whenever a user asks for farm sounds we can retrieve a set of sounds where farm appears. Besides we can also search for the sounds of the related concepts, such as chicken. A random subset of the relevant sounds is forwarded to the subsequent block, the sound sequencing.

4. SOUND SEQUENCING

Besides the definition and selection of the suitable SFX, a significant part of the work of the sound designer is setting up parameters and time lines in a multitrack project, such as volumes or panoramic envelopes. This section details some of the rules used to mix all fetched tracks and compose the synthetic atmosphere. The SFX retrieval module returns mono and dry (no effect has been applied) tracks.

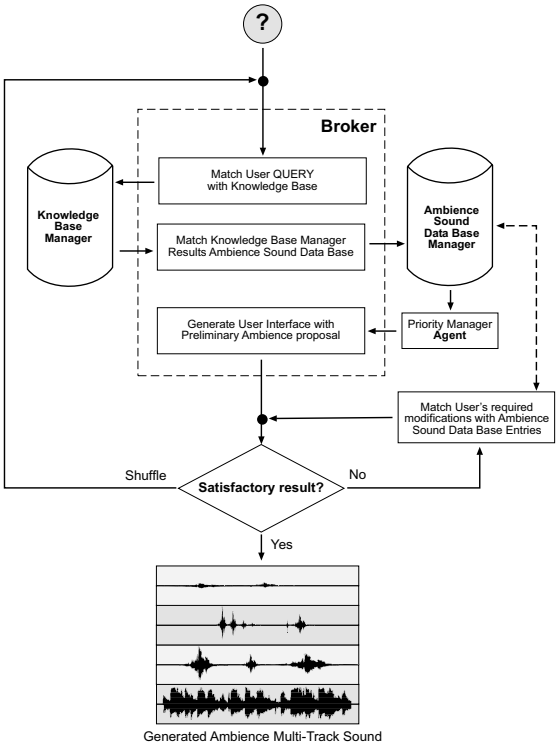

Figure 1: *Flow diagram of the system.*

Whenever available, the module differentiates between two types of tracks: long ambient tracks and several short isolated effects. One long track is selected to serve as a ambient loop on which the short sounds, or specifics, are added. With such picture of the workspace we hint some rules on how to place the tracks in the mix, how to adjust channel controls (gain, panning and equalization), and which effects (echo, reverb) can be applied to each track.

The systems automatically distributes the tracks along the mix, placing first the ambient loop and inserting sequentially the specifics, with a probabilistic criterion. This probabilistic criterion is based on the inverse of a frame-based energy computation. This means that the more energetic regions of the mix will have less probability to receive the following effect track. This process is depicted in figure 2.

It is a cunning feature to keep a certain degree of randomness. Again, a shuffle button can remix the atmosphere as many times as desired. Also, further models can take into account other parameters such as energy variation (in order to avoid two transients happening at the same time), such as spectrum centroid (in order to avoid as much as possible the frequency content overlap), or others.

Another important feature is the automatic adjustment of channel controls: gain, panning and equalization. Regarding the levels, these are set so that the track maximum levels are 3 dB above the long ambient mean level and that

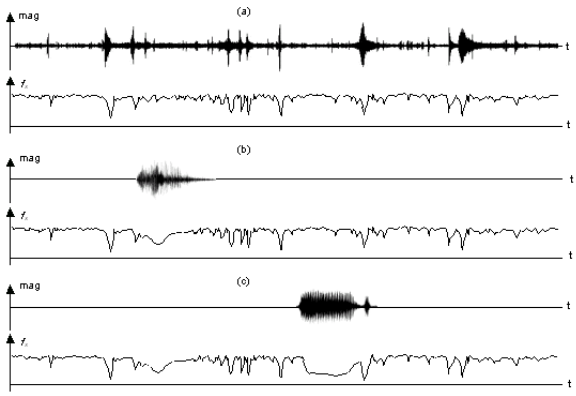

Figure 2: Mix example. a. long ambient sound and the corresponding probability density function. b and c SFX added and the corresponding recalculated probability density function

no saturation / clipping problems appear. Regarding the stereo panning the ambient sound is centered and the isolated tracks are panned one left one right along time in order to minimize time overlap. The amount of panning depends on how close are two consecutive tracks, the closer, the more panned. Equalizing is only applied to those tracks that overlap significantly in frequency domain with the ad-

jacent tracks or with the ambient loop sound. In these cases the effect track is 6-band equalized to flatten down to -12 dB the overlapping frequency region.

Finally, the strategy for the automation of the effects we propose is based on rules. These rules are mainly related with the context of the ambiance. Say we are reconstructing an office atmosphere, we will apply a medium room reverb to whatever effect track we drop to the mix; if we are reconstructing a mountain atmosphere, we can apply some echo to the tracks.

4.1. Integration in professional environments

The advent of high quality audio and spatialization surround setups (e.g: 5.1), first in the film industry, and more recently in home entertainment with DVD, offers the possibilities to create more engaging and immersive ambient sound. It is now possible to have ambient loops that take advantage of very low pitch sound (using subwoofers). It is possible to simulate movement in a tri-dimensional space or specific sound elements that pan in every direction we wish. On the other hand the complexity of setting up a multitrack project for a surround scenario increased a lot. It would be extremely useful for a sound designer to specify at a higher level which surround characteristics are desirable for the ambiance, so that the system can provide him a multitrack project file, and respective sound files, already configured to be integrated in his main project.

5. EXAMPLES

Let us now give critical comments on some typical examples on ambiance generation:

Some of the ambiances created had too many events in it. The "jungle" ambiance had plenty of tropical birds, elephants and monkeys and sounded more like a zoo than a jungle.

Some of the ambiances need greater detail in the specification. A "war" ambiance query returned war sounds of different epochs, e.g: bombs, machine guns, swords and laser guns.

The sex ambiance retrieved sounds produced by too many people to be realistic.

These experiences lead us to the conception of a refinement control to add/remove specific sound classes or another control for the density of specifics.

As a multitrack application, we have used the free editor Audacity[3]. In addition to common sound editing functionalities, Audacity allows to mix several tracks together and apply effects to tracks. Audacity allows to save multitrack sessions yet it does not read sessions created by external programs. We have therefore tweaked the application in order to load our automatically generated ambiance multitrack sessions.

[3] `http://audacity.sourceforge.net/`

6. CONCLUSIONS AND FUTURE WORK AND FUTURE WORK

We have presented a system for semi-automatic ambiance generation. The ambiances generated by textual query can be further refined by the user. The user controls the number of sounds that should be returned and can add and remove types of sounds, e.g: "more penguin sounds". Furthermore the ambiance is delivered to the user as a multitrack project, providing thus flexibility to fine tune the results. We plan to extend this work to semi-automatic sonifications of audiovisual productions given scripts (or briefings) and some information of the timing.

7. ACKNOWLEDGMENTS

We thank the staff from the Tape Gallery for all the support, discussion and feedback.

This work is partially funded by the AUDIOCLAS Project E! 2668 Eureka (http://www.audioclas.org). We thank the collaboration from Sylvain Le Groux, Julien Ricard and Nicolas Wack. We thank the review and feedback from Perfecto Herrera and José Lozano.

8. REFERENCES

[1] Robert L.Mott, *Sound Effects: Radio, TV, and Film*, Focal Press, 1990.

[2] Nick Peck, "Beyond the library: Applying film post-production techniques to game sound design," in *Proc. of Game Developers Conference*, San Jose CA, USA, 2001.

[3] P. Cano, M. Koppenberger, P. Herrera, O. Celma and V. Tarasov, "Sound effects taxonomy management in production environments," in *Proc. AES 25th Int. Conf.*, London, UK, 2004.

[4] P. Cano, M. Koppenberger, S. Le Groux, P. Herrera, and N.Wack, "Perceptual and semantic management of sound effects with a WordNet-based taxonomy," in *Proc. of the ICETE*, Setúbal, Portugal, 2004.

[5] George A. Miller, "WordNet: A lexical database for english," *Communications of the ACM*, pp. 39–45, November 1995.