State of the Art in Sound Texture Synthesis

Diemo Schwarz

STATE OF THE ART IN SOUND TEXTURE SYNTHESIS

Diemo Schwarz

Real-Time Music Interaction Team
IRCAM–CNRS–UPMC, UMR STMS
Paris, France
diemo.schwarz@ircam.fr

ABSTRACT

The synthesis of sound textures, such as rain, wind, or crowds, is an important application for cinema, multimedia creation, games and installations. However, despite the clearly defined requirements of naturalness and flexibility, no automatic method has yet found widespread use. After clarifying the definition, terminology, and usages of sound texture synthesis, we will give an overview of the many existing methods and approaches, and the few available software implementations, and classify them by the synthesis model they are based on, such as subtractive or additive synthesis, granular synthesis, corpus-based concatenative synthesis, wavelets, or physical modeling. Additionally, an overview is given over analysis methods used for sound texture synthesis, such as segmentation, statistical modeling, timbral analysis, and modeling of transitions.

1. INTRODUCTION

The synthesis of sound textures is an important application for cinema, multimedia creation, games and installations. Sound textures are generally understood as sound that is composed of many micro-events, but whose features are stable on a larger time-scale, such as rain, fire, wind, water, traffic noise, or crowd sounds. We must distinguish this from the notion of *soundscape*, which describes the sum of sounds that compose a scene, some components of which could be sound textures.

There are a plethora of methods for sound texture synthesis based on very different approaches that we will try to classify in this state-of-the-art article. We'll start by a definition of the terminology and usages (sections 1.1– 1.3), before giving an overview of the existing methods for synthesis and analysis of sound textures (sections 2 and 3), and some links to the first available software products (section 4). Finally, the discussion (section 5) and conclusion (section 6) also point out some especially noteworthy articles that represent the current state of the art.

1.1. Definition of Sound Texture

An early thorough definition, and experiments on the perception and generation of sound textures were given by Saint-Arnaud [74] and Saint-Arnaud and Popat [75], summarised in the following visual analogy:

> *A sound texture is like wallpaper: it can have local structure and randomness, but the characteristics of the fine structure must remain constant on the large scale.*

Figure 1 illustrates this statement. They culminate in the following working definition:

1. Sound textures are formed of basic sound elements, or atoms;
2. atoms occur according to a higher-level pattern, which can be periodic, random, or both;
3. the high-level characteristics must remain the same over long time periods (which implies that there can be no complex message);
4. the high-level pattern must be completely exposed within a few seconds ("attention span");
5. high-level randomness is also acceptable, as long as there are enough occurrences within the attention span to make a good example of the random properties.

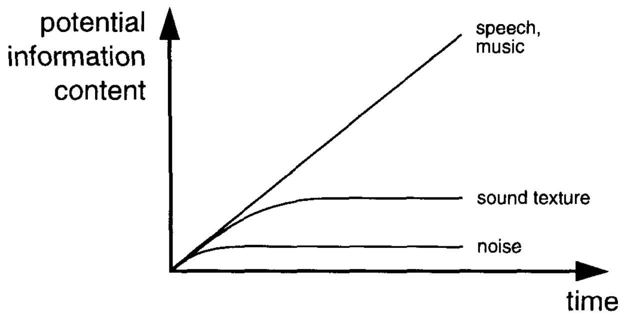

Figure 1: Potential information content of a sound texture vs. time (from Saint-Arnaud and Popat [75]).

1.1.1. What Sound Texture is Not

Attempting a negative definition might help to clarify the concept. We exclude from sound textures the following:

Contact sounds from interaction with objects, such as impact, friction, rolling sounds, treated in many works close to sound texture synthesis [1, 15, 48, 65, 87]. These sounds violate the "wallpaper" property.

Sound scapes are often treated together with sound textures, since they always contain sound textures. However, sound scapes also comprise information-rich event-type sounds, as further explained in section 1.1.2.

Figure 2: Examples of natural and synthesised oriented oscillating patterns from Peyré [70].

Sound design is the wider context of creating interaction sounds, sound scapes, and sound textures. Literature in the field often contains useful methods for sound texture design [10, 14, 58–61].

In some cases of music composition or performance, *sound texture* is used to mean *non-tonal, non-percussive* sound material, or *non-harmonic, non-rhythmic* musical material.

See also Strobl [84] for an investigation of the term *texture* outside of sound, such as in textiles, typography, gastronomy.

1.1.2. Sound Scapes

Because sound textures constitute a vital part of sound scapes, it is useful to present here a very brief introduction to the classification and automatic generation of *sound scapes*. Also, the literature about sound scapes is inevitably concerned about the synthesis and organisation of sound textures.

The first attempts at definition and classification of sound scapes have been by Murray Schafer [76], who distinguishes *keynote*, *signal*, and *soundmark* layers in a soundscape, and proposes a referential taxonomy incorporating socio-cultural attributes and ecological acoustics.

Gaver [36], coming from the point of view of acoustic ecology, organises sounds according to their physical attributes and interaction of materials.

Current work related to sound scapes are frequent [8, 9, 33, 58–61, 88, 89].

1.2. Existing Attempts at Classification of Texture Synthesis

As a starting point, Strobl et al. [85] provide an attempt at a definition of sound texture, and an overview of work until 2006. They divide the reviewed methods into two groups:

Methods from computer graphics Transfer of computer graphics methods for visual texture synthesis applied to sound synthesis [22, 64, 67]. See figure 2 for examples of textured images.

Methods from computer music Synthesis methods from computer music or speech synthesis applied to sound texture synthesis [4, 7, 17, 42, 43, 94].

A newer survey of tools in the larger field of sound design and composition [58] propose the same classification by synthesis method as elaborated in section 2 below. The article makes a point that different classes of sound require different tools (*"A full toolbox means the whole world need not look like a nail!"*) and gives a list of possible matches between different types of sound and the sound synthesis methods on which they work well.

In an article by Filatriau and Arfib [31], texture synthesis algorithm are reviewed from the point of view of gesture-controlled instruments, which makes it worthwile to point out the different usage contexts of sound textures in the following section.

1.3. Different Usages and Significations

It is important to note that there is a possible confusion in the literature about the precise signification of the term *sound texture* that is dependent on the intended usage. We can distinguish two frequently occuring usages:

Expressive texture synthesis Here, the aim is to interactively generate sound for music composition, performance, or sound art, very often as an expressive digital musical instrument (DMI). *Sound texture* is then often meant to distinguish the generated sound material from tonal and percussive sound, i.e. sound texture is anything that is predominantly defined by timbre rather than by pitch or rhythm.

The methods employed for expressive texture generation can give rise to naturally sounding textures, as noted by

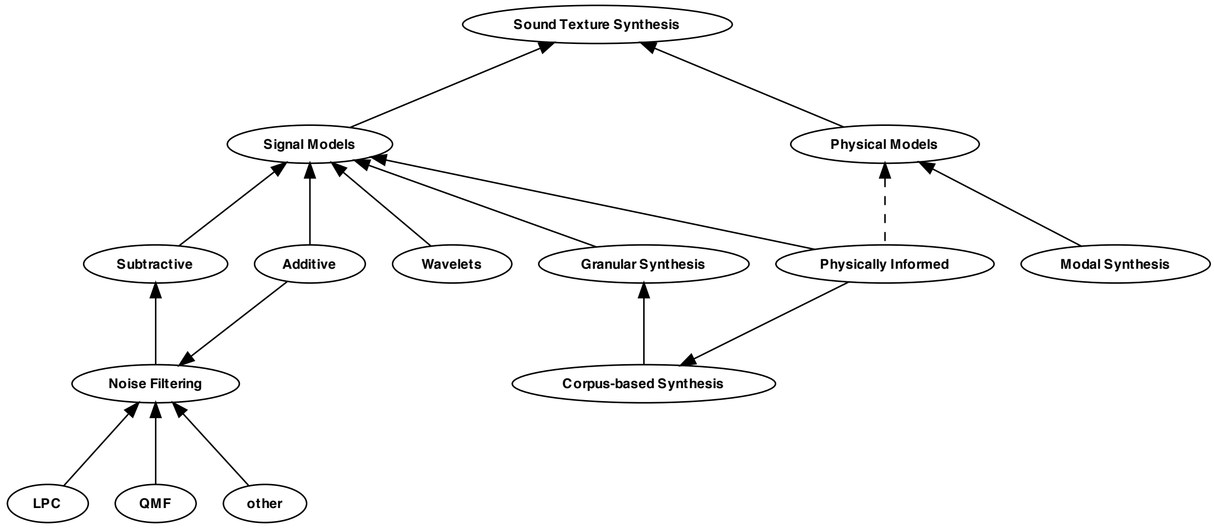

Figure 3: Classification hierarchy of sound texture synthesis methods. Dashed arrows represent use of information.

Di Scipio [17], but no systematic research on the usable parameter space has been carried out, and it is up to the user (or player) to constrain herself to the natural sounding part.

This strand of sound texture is pursued in the already mentioned review in [31], and their follow-up work [32].

Natural texture resynthesis tries to synthesise environmental or human textural sound as part of a larger soundscape, amongst others for audio–visual creation like cinema or games. Often, a certain degree of realism is striven for (like in photorealistic texture image rendering), but for most applications, either symbolic or impressionistic *credible texture synthesis* is actually sufficient, in that the textures convey the desired ambience or information, e.g. in simulations for urbanistic planning. All but a few examples of the work described in the present article is aimed this usage.

2. CLASSIFICATION OF SYNTHESIS METHODS

In this section, we will propose a classification of the existing methods of sound texture synthesis. It seems most appropriate to divide the different approaches by the synthesis methods (and analysis methods, if applicable) they employ:

- Noise filtering (section 2.1) and additive sinusoidal synthesis (section 2.2)
- Physical modeling (section 2.3) and physically-informed signal models
- Wavelet representation and resynthesis (section 2.4)
- Granular synthesis (section 2.5) and its content-based extension corpus-based concatenative synthesis (section 2.6)
- Non-standard synthesis methods, such as fractal or chaotic maps (section 2.7)

Figure 3 gives an overview over the classes and their relationships. Other possible aspects for classification are the degree of dependency on a model, the degree to which the method is data-driven, the real-time capabilities, and if the method has been formally evaluated in listening tests. Some of these aspects will be discussed in section 5.

2.1. Subtractive Synthesis

Noise filtering is the “classic” synthesis method for sound textures, often based on specific modeling of the source sounds.

Based on their working definition listed in section 1.1, Saint-Arnaud and Popat [75] build one of the first analysis–synthesis models for texture synthesis, based on 6-band Quadrature Mirror filtered noise.

Athineos and Ellis [4] and Zhu and Wyse [94] apply cascaded time and frequency domain linear prediction (CTFLP) analysis and resynthesis by noise filtering. The latter resynthesise the background din and the previously detect foreground events (see section 3.1.3) by applying the time and frequency domain LPC coefficients to noise frames with subsequent overlap–add synthesis. The events are sequenced by a Poisson distribution. The focus here is data reduction for transmission by low-bitrate coding.

McDermott et al. [53] apply statistical analysis (see section 3.2) to noise filtering synthesis, restrained to unpitched static textures like rain, fire, water only.

Peltola et al. [69] synthesises different characters of hand-clapping sounds by filters tuned to recordings of claps and combines them into a crowd by statistical modeling of different levels of enthousiasm and flocking behaviour of the crowd.

The venerable collection by Farnell [30], also available online[1], gives many sound and PUREDATA patch examples for synthesising

[1] http://obiwannabe.co.uk/tutorials/html/tutorials_main.html

various sound textures and sound effects by oscillators and filters, carefully tuned according to insights into the phenomenon to be simulated, as in this quote about rain:

> *"What is the nature of rain? What does it do?" According to the lyrics of certain shoegazing philosophies it's "Always falling on me", but that is quite unhelpful. Instead consider that it is nearly spherical particles of water of approximately 1-3mm in diameter moving at constant velocity impacting with materials unknown at a typical flux of 200 per second per meter squared. All raindrops have already attained terminal velocity, so there are no fast or slow ones. All raindrops are roughly the same size, a factor determined by their formation at precipitation under nominally uniform conditions, so there are no big or small raindrops to speak of. Finally raindrops are not "tear" shaped as is commonly held, they are in fact near perfect spheres. The factor which prevents rain being a uniform sound and gives rain its diverse range of pitches and impact noises is what it hits. Sometimes it falls on leaves, sometimes on the pavement, or on the tin roof, or into a puddle of rainwater.*

2.2. Additive Sinusoidal + Noise Synthesis

Filtered noise is often complemented by oscillators in the additive sinusoidal partials synthesis method.

In the QCITY project[2], the non-real time simulation of traffic noise is based on a sinusoids+noise sound representation, calibrated according to measurements of motor states, exhaust pipe type, damping effects. It allows to simulate different traffic densities, speeds, types of vehicles, tarmacs, damping walls, etc. [38]. The calculation of sound examples can take hours.

Verron [93] proposes in his PhD thesis and in other publications [91, 92], 7 physically-informed models from Gaver's [36] 3 larger classes of environmental sounds: liquids, solids, aerodynamic sounds. The models for impacting solids, wind, gushes, fire, water drops, rain, and waves are based on 5 empirically defined and parameterised sound atoms: modal impact, noise impact, chirp impact, narrow band noise, wide band noise. Each model has 2–4 low-level parameters (with the exception of 32 band amplitudes for wide band noise).

Verron then painstakingly maps high-level control parameters like *wind force and coldness, rain intensity, ocean wave size* to the low-level atom parameters and density distribution.

The synthesis uses the FFT^{-1} method [73] that is extended to include spatial encoding into the construction of the FFT, and then one IFFT stage per output channel.[3]

2.3. Physical Modeling

Physical modeling can be applied to sound texture synthesis, with the drawback that a model must be specifically developed for each class of sounds to synthesise (e.g. friction, rolling, machine noises, bubbles, aerodynamic sounds) [63, 64], the latter adding an extraction of the impact impulse sound and a perceptual evaluation of the realism of synthesised rolling sounds (see also Lagrange et al. [47]). Often, modal resonance models are used [90], where the inexpensively synthesisable modes are precalculated from expensive rigid body simulations.

Other signal-based synthesis methods are often *physically-informed* [12, 13] in that they control signal models by the output of a physical model that captures the behaviour of the sound source. See, e.g. Cook [14], Verron [93] (also described in section 2.2), Picard et al. [71], or the comprehensive toolbox by Menzies [55] (see section 4 for its implementation).

The synthesis of liquid sounds described by Doel [20] is a combination of a physically informed sinusoidal signal model for single bubble sounds (going back to [51]), and an empirical phenomenological model for bubble statistics, resulting in a great sound variety ranging from drops, rain, air bubbles in water to streams and torrents.

An extreme example is the synthesis of sounds of liquids by fluid simulations [62], deriving sound control information from the spherical harmonics of individually simulated bubbles (up to 15000).[4]

2.4. Wavelets

The multiscale decomposition of a signal into a wavelet coefficient tree has been first applied to sound texture synthesis by El-Yaniv et al. [26] and Dubnov et al. [22], and been reconsidered by Kersten and Purwins [45].[5]

Here, the multiscale wavelet tree signal and structure representation is resampled by reorganising the order of paths down the tree structure. Each path then resynthesises a short bit of signal by the inverse wavelet transform.

These approaches take inspiration from image texture analysis and synthesis and try to model temporal dependencies as well as hierarchic dependencies between different levels of the multi-level tree representation they use. Kersten and Purwins's work is in an early stage where the overall sound of the textures is recognisable (as shown by a quantitative evaluation experiment), but the resulting structure seems too fine-grained, because the sequence constraints of the original textures are actually not modeled, such that the fine temporal structure gets lost. This violates the *autocorrelation feature* found important for audio and image textures by McDermott et al. [53] and Fan and Xia [29].

An model-based approach using wavelets for modeling stochastic-based sounds is pursued by Miner and Caudell [56]. Parameterizations of the wavelet models yield a variety of related sounds from a small set of dynamic models.

Another wavelet-based approach is by Kokaram and O'Regan [46, 66], based on Efros and Leung's algorithm for image texture synthesis [23]. Their multichannel synthesis achieves a large segment size well adapted to the source (words, baby cries, gear shifts,

[2] http://qcity.eu/dissemination.html

[3] Binaural sound examples (sometimes slightly artificial sounding) and one video illustrating the high-level control parameters and the difference between point and extended spatial sources can be found on http://www.charlesverron.com/thesis/.

[4] http://gamma.cs.unc.edu/SoundingLiquids

[5] Sound examples are available at http://mtg.upf.edu/people/skersten?p=Sound%20Texture%20Modeling

drum beats) and thus a convincing and mostly artefact-free resynthesis.[6]

2.5. Granular Synthesis

Granular synthesis uses snippets of an original recording, and possibly a statistical model of the (re)composition of the grains [7, 22, 26, 35, 42, 43, 67]. The optimal grain size is dependent on the typical time-scale of the texture. If chosen sufficiently long, the short-term micro-event distribution is preserved within a grain, while still allowing to create a non-repetitive long-term structure.

Lu et al. [50] recombine short segments, possibly with transposition, according to a model of transition probabilities (see section 3.4). They explicitly forbid short backward transitions to avoid repetition. The segmentation is based on a *novelty score* on MFCCs, in the form of a similarity matrix.

Strobl [84] studies the methods by Hoskinson and Pai [42, 43] and Lu et al. [50] in great detail, improves the parameters and resynthesis to obtain "perceptually perfect segments of input textures", and tries a hybridisation between them [84, chapter 4]. She then implements Lu et al.'s method in an interactive real-time PUREDATA patch.

2.6. Corpus-based Synthesis

Corpus-based concatenative synthesis can be seen as a content-based extension of granular synthesis [78, 79]. It is a new approach to sound texture synthesis [11, 80, 82, 83]. Corpus-based concatenative synthesis makes it possible to create sound by selecting snippets of a large database of pre-recorded audio (the corpus) by navigating through a space where each snippet is placed according to its sonic character in terms of audio descriptors, which are characteristics extracted from the source sounds such as pitch, loudness, and brilliance, or higher level meta-data attributed to them. This allows one to explore a corpus of sounds interactively or by composing paths in the space, and to create novel timbral evolutions while keeping the fine details of the original sound, which is especially important for convincing sound textures.

Finney [33] uses a corpus of unstructured recordings from the *freesound* collaborative sound database[7] as base material for sample-based sound events and background textures in a comprehensive sound scape synthesis application (see section 4, also for evaluation by a subjective listening test). The recordings for sound texture synthesis are segmented by MFCC+BIC (see section 3.1.2) and high- and low-pass filtered to their typical frequency ranges. Spectral outliers (outside one standard deviation unit around the mean MFCC) are removed. Synthesis then chooses randomly out of a cluster of the 5 segments the MFCCs of which are closest. How the cluster is chosen is not explicitly stated.

Schwarz and Schnell [80] observe that existing methods for sound texture synthesis are often concerned with the extension of a given recording, while keeping its overall properties and avoiding artefacts. However, they generally lack controllability of the resulting sound texture. They propose two corpus-based methods of statistical modeling of the audio descriptor distribution of texture recordings using histograms and Gaussian mixture models. The models can be interpolated to steer the evolution of the sound texture between different target recordings (e.g. from light to heavy rain). Target descriptor values are stochastically drawn from the statistic models by inverse transform sampling to control corpus-based concatenative synthesis for the final sound generation, that can also be controlled interactively by navigation through the descriptor space.[8] See also section 4 for the freely available CATART application that served as testbed for interactive sound texture synthesis. To better cover the target descriptor space, they expand the corpus by automatically generating variants of the source sounds with transformations applied, and storing only the resulting descriptors and the transformation parameters in the corpus. A first attempt of perceptual validation of the used descriptors for wind, rain, and wave textures has been carried out by subject tests [52], based on studies on the perception of environmental sound [57, 86].

The work by Picard et al. [71] (section 2.3) has a corpus-based aspect in that it uses grain selection driven by a physics engine.

Dobashi et al. [18, 19] employ physically informed corpus-based synthesis for the synthesis of aerodynamic sound such as from wind or swords. They precompute a corpus of the aerodynamic sound emissions of point sources by computationally expensive turbulence simulation for different speeds and angles, and can then interactively generate the sound of a complex moving object by lookup and summation.

2.7. Non-standard Synthesis Methods

Non-standard synthesis methods, such as fractal synthesis or chaotic maps, generated by iterating nonlinear functions, are used most often for expressive texture synthesis [17, 32], especially when controlled by gestural input devices [3, 31].

3. ANALYSIS METHODS FOR SOUND TEXTURES

Methods that analyse the properties of sound textures are concerned with segmentation (section 3.1), the analysis of statistical properties (section 3.2) or timbral qualities (section 3.3), or the modeling of the sound source's typical state transitions (section 3.4).

3.1. Segmentation of Source Sounds

3.1.1. Onset Detection

O'Modhrain and Essl [65] describe a granular analysis method they call *grainification* of the interaction sounds with actual grains (pebbles in a box, starch in a bag), in order to expressively control granular synthesis (this falls under the use case of expressive texture synthesis in section 1.3): By threshold-based attack detection with a retrigger limit time, they derive the grain attack times, volume (by picking the first peak after the attack), and spectral content (by counting the zero-crossings in a 100 sample window after the attack). These parameters control a granular synthesizer's trigger, gain and transposition. See also Essl and O'Modhrain [28].

Lee et al. [48] estimate contact events for segmentation of rolling sounds on a high-pass filtered signal, on which an energy threshold

[6] Sound examples are available at http://www.netsoc.tcd.ie/~dee/STS_EUSIPCO.html.

[7] http://www.freesound.org/

[8] Sound examples can be heard on http://imtr.ircam.fr/imtr/Sound_Texture_Synthesis.

is applied. Segments are then modeled by LPC filters on several bands for resynthesis.

3.1.2. Spectral Change Detection

Lu et al. [50] segment sounds based on a *novelty score* on MFCCs, in the form of a similarity matrix. This also serves to model transition probabilities (see section 3.4). The analysis has been improved upon by Strobl [84].

Finney [33] (see also sections 2.6 and 4) uses a method of segmenting environmental recordings using the *Bayesian Information Criterion* (BIC) on MFCCs [2], while enforcing a minimum segment length dependent on the type of sounds: The segment length should correspond to the typical event length.

3.1.3. LPC Segmentation

Kauppinen and Roth [44] segment sound into locally stationary frames by LPC pulse segmentation, obtaining the optimal frame length by a stationarity measure from a long vs. short term prediction. The peak threshold is automatically adapted by the median filter of the spectrum derivative.

Similarly, Zhu and Wyse [94] detect foreground events by frequency domain linear predictive coding (FDLPC), which are then removed to leave only the 'din' (the background sound). See also section 2.1 for their corresponding subtractive synthesis method.

3.1.4. Wavelets

Hoskinson and Pai [42, 43] (see also section 2.5) segment the source sounds into *natural grains*, which are defined by the *minima* of the energy changes in the first 6 wavelet bands, i.e. where the sound is the most stable.

3.1.5. Analysis into Atomic Components

Other methods [49, 50, 59, 60, 67] use an analysis of a target sound in terms of event and spectral components for their statistical recombination. They are linked to the modelisation of impact sounds by wavelets by Ahmad et al. [1].

Bascou [5], Bascou and Pottier [6] decompose a sound by *Matching Pursuit* into time-frequency atoms from a dictionary manually built from "characteristic" grains of the sound to decompose.

3.2. Analysis of Statistical Properties

Dubnov et al. [22], El-Yaniv et al. [26] apply El-Yaniv et al.'s [25] Markovian unsupervised clustering algorithm to sound textures, thereby constructing a discrete statistical model of a sequence of paths through a wavelet representation of the signal (see section 2.4).

Zhu and Wyse [94] estimate the density of foreground events, singled out of the texture by LPC segmentation (see section 3.1.3). Masurelle [52] developed a simple density estimation of impact events based on O'Modhrain and Essl [65], applicable e.g. to rain. For the same specific case, Doel [20] cites many works about the statistics of rain.

McDermott et al. [53] (see section 2.4) propose a neurophysically motivated statistical analysis of the kurtosis of energy in subbands, and apply these statistics to noise filtering synthesis (later also applied to classification of environmental sound [27]).

3.2.1. Analysis not for Synthesis

There is work concerned with analysis and classification of sound textures, which is not relevant for synthesis, like Dubnov and Tishby [21], who use higher-order spectra for classification of environmental sounds, or by Desainte-Catherine and Hanna [16], who propose statistical descriptors for noisy sounds.

In the recent work by Grill [37] for an interactive sound installation, a corpus-based synthesis system plays back samples of soundscapes matching the participants' noises. While the synthesis part is very simple, the matching part is noteworthy for its use of *fluctuation patterns*, i.e. the modulation spectrum for all bark bands of a 3 second segment of texture. This 744-element feature vector was then reduced to 24 principal components prior to matching.

3.3. Analysis of Timbral Qualities

Hanna et al. [40] note that, in the MIR domain, there is little work about audio features specifically for noisy sounds. They propose classification into the 4 sub-classes coloured, pseudo-periodic, impulsive noise (rain, applause), and noise with sinusoids (wind, street soundscape, birds). They then detect the transitions between these classes using a Bayesian framework. This work is generalised to a sound representation model based on stochastic sinusoids [39, 41].

Only corpus-based concatenative synthesis methods try to characterise the sonic contents of the source sounds by perceptually meaningful audio descriptors: [24, 78–80, 82, 83]

3.4. Clustering and Modeling of Transitions

Saint-Arnaud [74] builds clusters by k-means of their input sound atoms (filter band amplitudes) using the *Cluster based probability model* [72]. The amplitudes are measured at the current frame and in various places in the past signal (as defined by a *neighbourhood mask*) and thus encode the typical transitions occuring in the sound texture. Saint-Arnaud's master's thesis [74] focuses on classification of sound textures, also with perceptual experiments, while the later article [75] extends the model to analysis by noise filtering (section 2.1).

Lu et al. [50] model transition probabilities based on a similarity matrix on MFCC frames. Hoskinson and Pai [42, 43] also model transitions based on smoothness between their wavelet-segmented *natural grains* (section 3.1.4). Both methods have been studied in detail and improved upon by Strobl [84].

4. AVAILABLE SOFTWARE

Freely or commercially available products for sound textures are very rare, and mostly specific to certain types of textures. The only commercial product the author is aware of is the crowd simulator CROWD CHAMBER[9], that takes a given sound file to multiply the sources, probably using PSOLA-based pitch shifting, time-stretching and filtering effects. That means, it is not actually a texture synthesiser, but an effects processor that adds a "crowd" effect on an existing voice recording. The provided example sounds are very unconvincing.

Finney [33] presents a full soundscape synthesis system based on concatenative and sample-based playback (see sections 2.6 and 3.1.2). The synthesis and interaction part is integrated into Google Street View.[10] Notable and related to sound textures is the precise modeling of traffic noise with single samples of passing cars, categorised by car type and speed, that are probabilistically recombined according to time of day, number of street lanes, and with traffic lights simulated by clustering. The evaluation also reported in [34] concentrates on the immersive quality of the generated sound scapes in a subjective listening test with 8 participants. Interestingly, the synthetic sound scapes rate consistently higher than actual recordings of the 6 proposed locations.

The PHYA framework [54, 55] is a toolbox of various physically motivated filter and resonator signal models for impact, collision, and surface sounds.[11]

The author's CATART system [83] for interactive real-time corpus-based concatenative synthesis is implemented in MAX/MSP with the extension libraries FTM&CO.[12] and is freely available.[13] It allows to navigate through a two- or more-dimensional projection of the descriptor space of a corpus of sound segments in real-time using the mouse or other gestural controllers, effectively extending granular synthesis by content-based direct access to specific sound characteristics. This makes it possible to recreate dynamic evolutions of sound textures with precise control over the resulting timbral variations, while keeping the micro-event structure intact, as soon as the segments are long enough, described in section 2.6. One additional transformation is the augmentation of the texture density by triggering at a faster rate than given by the segments' length, thus layering several units, which works very well for textures like rain, wind, water, or crowds.[8]

The descriptors are calculated within the CATART system by a modular analysis framework [81]. The used descriptors are: fundamental frequency, periodicity, loudness, and a number of spectral descriptors: spectral centroid, sharpness, flatness, high- and mid-frequency energy, high-frequency content, first-order autocorrelation coefficient (expressing spectral tilt), and energy. Details on the descriptors used can be found in [77] and [68].

5. DISCUSSION

Concerning the dependency on a specific model, we can see that the presented methods fall clearly on one of two sides of a strong dichotomy between rule-based and data-driven approaches: The methods using a low-level signal or physical model (sections 2.2–2.3) are almost all based on a very specific modeling of the sound texture generating process, except the first three methods using noise filtering by statistical modeling. The methods using segments of signal or wavelet coefficients (sections 2.4– 2.6) are, by their data-driven nature, more generally applicable to many different texture sounds, and far more independent from a specific texture model.

Also, physical models do not provide a direct link between their internal parameters, and the characteristics of the produced sound. As Menzies [55] notes:

> *In principle, sound in a virtual environment can be reproduced accurately through detailed physical modelling. Even if this were achieved, it is not enough for the Foley sound designer, who needs to be able to shape the sound according to their own imagination and reference sounds: explicit physical models are often difficult to calibrate to a desired sound behaviour although they are controlled directly by physical parameters.*

Physically-informed models allow more of this flexibility but still expose parameters of a synthesis model that might not relate directly to a percieved sound character. What's more, the physical and signal models' parameters might capture a certain variety of a simulated sound source, but will arguably be limited to a smaller range of nuances, and include to a lesser extent the context of a sound source, than the methods based on actual recordings (wavelets and corpus-based concatenative synthesis).

5.1. Perception and Interaction

General studies of the perception of environmental sound textures are rare, with the exception of [53, 57, 86], and systematic evaluation of the quality of the synthesised sound textures by formal listening tests is only beginning to be carried out in some of the presented work, e.g. [52, 63]. Only Kokaram and O'Regan [46, 66] have taken the initiative to start defining a common and comparable base of test sounds by adopting the examples from El-Yaniv et al. [26] and Dubnov et al. [22] as test cases.

Finally, this article concentrated mainly on the sound synthesis and analysis models applied to environmental texture synthesis, and less on the way how to control them, or the interactivity they afford. Gestural control seems here a promising approach for interactive generation of sound textures [3, 31, 52].

5.2. Recommended Reading

While this article strove to give a comprehensive overview of existing methods for sound texture synthesis and analysis, some of the work stands out, representing the state of the art in the field:

- Finney [33] for the introduction to sound scapes, the reference to the MFCC+BIC segmentation method, and the precise traffic modeling.
- Verron [93] and Farnell [30] for the detailed account of physically informed environmental sound synthesis that gives an insight about how these sounds work.
- Kokaram and O'Regan [46, 66] and Schwarz and Schnell [80] for the most convincing results so far.

[9] http://www.quikquak.com/Prod_CrowdChamber.html

[10] http://dev.mtg.upf.edu/soundscape/media/StreetView/streetViewSoundscaper2_0.html

[11] Available at http://www.zenprobe.com/phya/.

[12] http://ftm.ircam.fr

[13] http://imtr.ircam.fr/imtr/CataRT

6. CONCLUSION

We have seen that, despite the clearly defined problem and application context, the last 16 years of research into sound texture synthesis have not yet brought about a prevailing method that satisfies all requirements of realism and flexibility. Indeed, in practice, the former is always the top priority, so that the flexibility of automated synthesis methods is eschewed in favour of manual matching and editing of texture recordings in post-production, or simple triggering of looped samples for interactive applications such as games.

However, the latest state-of-the-art results in wavelet resynthesis [46, 66] and descriptor-based granular synthesis [80] promise practical applicability because of their convincing sound quality.

7. ACKNOWLEDGMENTS

The author would like to thank the anonymous reviewers and Stefan Kersten for their careful re-reading, additional references, and pertinent remarks.

The work presented here is partially funded by the *Agence Nationale de la Recherche* within the project *Topophonie*, ANR-09-CORD-022, http://topophonie.fr.

8. REFERENCES

[1] W. Ahmad, H. Hacıhabiboglu, and A.M. Kondoz. Analysis-Synthesis Model for Transient Impact Sounds by Stationary Wavelet Transform and Singular Value Decomposition. In *Proceedings of the International Computer Music Conference (ICMC)*, 2008.

[2] X Anguera and J Hernando. Xbic: Real-time cross probabilities measure for speaker segmentation, 2005.

[3] D. Arfib. Gestural strategies for specific filtering processes. *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, 2002. URL http://www2.hsu-hh.de/EWEB/ANT/dafx2002/papers/DAFX02_Arfib_Couturier_Kessous_gestural_stategies.pdf.

[4] M. Athineos and D.P.W. Ellis. Sound texture modelling with linear prediction in both time and frequency domains. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 5:V–648–51 vol.5, April 2003. ISSN 1520-6149.

[5] C Bascou. Modélisation de sons bruités par la synthese granulaire. Rapport de stage de DEA ATIAM, Université Aix-Marseille II, 2004.

[6] C. Bascou and L. Pottier. New sound decomposition method applied to granular synthesis. In *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, 2005.

[7] C. Bascou and L. Pottier. GMU, a flexible granular synthesis environment in Max/MSP. In *Proceedings of the International Conference on Sound and Music Computing (SMC)*, 2005.

[8] D. Birchfield, N. Mattar, and H. Sundaram. Design of a generative model for soundscape creation. In *Proceedings of the International Computer Music Conference (ICMC)*, Barcelona, Spain, 2005.

[9] D. Birchfield, N. Mattar, H. Sundaram, A. Mani, and B. Shevade. Generative Soundscapes for Experiential Communication. *Society for Electro Acoustic Music in the United States (SEAMUS)*, 2005.

[10] P. Cano, L. Fabig, F. Gouyon, M. Koppenberger, A. Loscos, and A. Barbosa. Semi-automatic ambiance generation. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, Naples, Italy, 2004.

[11] Marc Cardle. Automated Sound Editing. Technical report, Computer Laboratory, University of Cambridge, UK, May 2004. URL http://www.cl.cam.ac.uk/users/mpc33/Cardle-Sound-Synthesis-techreport-2004-low-quality.pdf.

[12] P Cook. Physically informed sonic modeling (PhISM): Percussive synthesis. *Proceedings of the International Computer Music Conference (ICMC)*, 1996. URL http://scholar.google.com/scholar?hl=en&q=cook+phism&bav=on.2,or.r_gc.r_pw.&biw=1584&bih=783&um=1&ie=UTF-8&sa=N&tab=ws#3.

[13] PR Cook. Physically informed sonic modeling (phism): Synthesis of percussive sounds. *Computer Music Journal*, 1997. URL http://www.jstor.org/stable/3681012.

[14] P.R. Cook. Din of an"iquity": Analysis and synthesis of environmental sounds. In *Proceedings of the International Conference on Auditory Display (ICAD2007)*, pages 167–172, 2007.

[15] Richard Corbett, Kees van den Doel, John E. Lloyd, and Wolfgang Heidrich. Timbrefields: 3d interactive sound models for real-time audio. *Presence: Teleoperators and Virtual Environments*, 16(6):643–654, 2007. doi: 10.1162/pres.16.6.643. URL http://www.mitpressjournals.org/doi/abs/10.1162/pres.16.6.643.

[16] M. Desainte-Catherine and P. Hanna. Statistical approach for sound modeling. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, 2000.

[17] A. Di Scipio. Synthesis of environmental sound textures by iterated nonlinear functions. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, 1999.

[18] Yoshinori Dobashi, Tsuyoshi Yamamoto, and Tomoyuki Nishita. Real-time rendering of aerodynamic sound using sound textures based on computational fluid dynamics. *ACM Trans. Graph.*, 22:732–740, July 2003. ISSN 0730-0301. doi: http://doi.acm.org/10.1145/882262.882339. URL http://doi.acm.org/10.1145/882262.882339.

[19] Yoshinori Dobashi, Tsuyoshi Yamamoto, and Tomoyuki Nishita. Synthesizing sound from turbulent field using sound textures for interactive fluid simulation. In *Proc. of Eurographics*, pages 539–546, 2004.

[20] Kees van den Doel. Physically based models for liquid sounds. *ACM Trans. Appl. Percept.*, 2:534–546, October 2005. ISSN 1544-3558. doi: http://doi.acm.org/10.1145/1101530.1101554. URL http://doi.acm.org/10.1145/1101530.1101554.

[21] S. Dubnov and N. Tishby. Analysis of sound textures in musical and machine sounds by means of higher order statistical features. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*

(ICASSP), volume 5, pages 3845–3848. IEEE, 1997. ISBN 0818679190. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=604726.

[22] Shlomo Dubnov, Ziz Bar-Joseph, Ran El-Yaniv, Danny Lischinski, and Michael Werman. Synthesis of audio sound textures by learning and resampling of wavelet trees. *IEEE Computer Graphics and Applications*, 22(4):38–48, 2002.

[23] A.A. Efros and T.K. Leung. Texture synthesis by non-parametric sampling. In *Proceedings of the International Conference on Computer Vision*, volume 2, page 1033, 1999.

[24] Aaron Einbond, Diemo Schwarz, and Jean Bresson. Corpus-based transcription as an approach to the compositional control of timbre. In *Proceedings of the International Computer Music Conference (ICMC)*, Montreal, QC, Canada, 2009.

[25] R. El-Yaniv, S. Fine, and N. Tishby. Agnostic classification of Markovian sequences. In *Proceedings of the 1997 conference on Advances in neural information processing systems 10*, pages 465–471. MIT Press, 1998. ISBN 0262100762.

[26] Z.B.J.R. El-Yaniv, D.L.M. Werman, and S. Dubnov. Granular Synthesis of Sound Textures using Statistical Learning. In *Proceedings of the International Computer Music Conference (ICMC)*, 1999.

[27] D.P.W. Ellis, X. Zeng, and J.H. McDermott. Classifying soundtracks with audio texture features. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010. URL http://www.ee.columbia.edu/~{}dpwe/pubs/EllisZM11-texture.pdf.

[28] G. Essl and S. O'Modhrain. Scrubber: an interface for friction-induced sounds. In *Proceedings of the Conference for New Interfaces for Musical Expression (NIME)*, page 75. National University of Singapore, 2005.

[29] G. Fan and X.G. Xia. Wavelet-based texture analysis and synthesis using hidden Markov models. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, 50(1), 2003.

[30] Andy Farnell. *Designing Sound*. MIT Press, October 2010. ISBN 9780262014410. URL http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=12282.

[31] J.J. Filatriau and D. Arfib. Instrumental gestures and sonic textures. In *Proceedings of the International Conference on Sound and Music Computing (SMC)*, 2005.

[32] J.J. Filatriau, D. Arfib, and JM Couturier. Using visual textures for sonic textures production and control. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, 2006.

[33] N. Finney. Autonomous generation of soundscapes using unstructured sound databases. Master's thesis, MTG, IUA–UPF, Barcelona, Spain, 2009. URL static/media/Finney-Nathan-Master-Thesis-2009.pdf.

[34] N. Finney and J. Janer. Soundscape Generation for Virtual Environments using Community-Provided Audio Databases. In *W3C Workshop: Augmented Reality on the Web*, June 15 - 16, 2010 Barcelona, 2010. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.168.4252&rep=rep1&type=pdfhttp://www.w3.org/2010/06/w3car/.

[35] M. Fröjd and A. Horner. Sound texture synthesis using an overlap-add/granular synthesis approach. *Journal of the Audio Engineering Society*, 57(1/2):29–37, 2009. URL http://www.aes.org/e-lib/browse.cfm?elib=14805.

[36] WW Gaver. How do we hear in the world? Explorations in ecological acoustics. *Ecological psychology*, 1993.

[37] T. Grill. Re-texturing the sonic environment. In *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*, pages 1–7. ACM, 2010. URL http://portal.acm.org/citation.cfm?id=1859805.

[38] S. Guidati and Head Acoustics GmbH. Auralisation and psychoacoustic evaluation of traffic noise scenarios. *Journal of the Acoustical Society of America*, 123(5):3027, 2008.

[39] P. Hanna. *Statistical modelling of noisy sounds : spectral density, analysis, musical transformtions and synthesis*. PhD thesis, Laboratoire de Recherche en Informatique de Bordeaux (LaBRI), 2003. URL http://dept-info.labri.fr/~hanna/phd.html.

[40] P. Hanna, N. Louis, M. Desainte-Catherine, and J. Benois-Pineau. Audio features for noisy sound segmentation. In *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR'04)*, pages 120–123, 2004.

[41] Pierre Hanna and Myriam Desainte-Catherine. A Statistical and Spectral Model for Representing Noisy Sounds with Short-Time Sinusoids. *EURASIP Journal on Advances in Signal Processing*, 2005(12):1794–1806, 2005. ISSN 1687-6172. doi: 10.1155/ASP.2005.1794. URL http://www.hindawi.com/journals/asp/2005/182056.abs.html.

[42] R. Hoskinson. *Manipulation and Resynthesis of Environmental Sounds with Natural Wavelet Grains*. PhD thesis, The University of British Columbia, 2002. URL https://www.cs.ubc.ca/grads/resources/thesis/May02/Reynald_Hoskinson.pdf.

[43] Reynald Hoskinson and Dinesh Pai. Manipulation and resynthesis with natural grains. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 338–341, Havana, Cuba, September 2001.

[44] I. Kauppinen and K. Roth. An Adaptive Technique for Modeling Audio Signals. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, 2001.

[45] Stefan Kersten and Hendrik Purwins. Sound texture synthesis with hidden markov tree models in the wavelet domain. In *Proceedings of the International Conference on Sound and Music Computing (SMC)*, Barcelona, Spain, July 2010.

[46] Anil Kokaram and Deirdre O'Regan. Wavelet based high resolution sound texture synthesis. In *Proceedings of the Audio Engineering Society Conference*, 6 2007. URL http://www.aes.org/e-lib/browse.cfm?elib=13952.

[47] M. Lagrange, B.L. Giordano, P. Depalle, and S. McAdams. Objective quality measurement of the excitation of impact sounds in a source/filter model. *Acoustical Society of America Journal*, 123:3746, 2008.

[48] Jung Suk Lee, Philippe Depalle, and Gary Scavone. Analysis / synthesis of rolling sounds using a source filter approach. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, Graz, Austria, September 2010.

[49] L. Lu, S. Li, L. Wen-yin, H.J. Zhang, and Y. Mao. Audio textures. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 2, 2002. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.4586&rep=rep1&type=pdf.

[50] L. Lu, L. Wenyin, and H.J. Zhang. Audio textures: Theory and applications. *IEEE Transactions on Speech and Audio Processing*, 12(2):156–167, 2004. ISSN 1063-6676. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1284343.

[51] A. Mallock. Sounds produced by drops falling on water. *Proc. R. Soc.*, 95:138—143, 1919.

[52] Aymeric Masurelle. Gestural control of environmental texture synthesis. Rapport de stage de DEA ATIAM, Ircam–Centre Pompidou, Université Paris VI, 2011.

[53] J.H. McDermott, A.J. Oxenham, and E.P. Simoncelli. Sound texture synthesis via filter statistics. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, October 18–21 2009.

[54] Dylan Menzies. Phya and vfoley, physically motivated audio for virtual environments. *Proceedings of the Audio Engineering Society Conference*, 2010. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.9848&rep=rep1&type=pdf.

[55] Dylan Menzies. Physically Motivated Environmental Sound Synthesis for Virtual Worlds. *EURASIP Journal on Audio, Speech, and Music Processing*, 2011. URL http://www.hindawi.com/journals/asmp/2010/137878.html.

[56] Nadine E. Miner and Thomas P. Caudell. Using wavelets to synthesize stochastic-based sounds for immersive virtual environments. *ACM Trans. Appl. Percept.*, 2:521–528, October 2005. ISSN 1544-3558. doi: http://doi.acm.org/10.1145/1101530.1101552. URL http://doi.acm.org/10.1145/1101530.1101552.

[57] Nicolas Misdariis, Antoine Minard, Patrick Susini, Guillaume Lemaitre, Stephen McAdams, and Etienne Parizet. Environmental sound perception: Metadescription and modeling based on independent primary studies. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010. URL http://articles.ircam.fr/textes/Misdariis10b/.

[58] A. Misra and P.R. Cook. Toward synthesized environments: A survey of analysis and synthesis methods for sound designers and composers. In *Proceedings of the International Computer Music Conference (ICMC)*, 2009.

[59] A. Misra, P.R. Cook, and G. Wang. A new paradigm for sound design. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, 2006.

[60] A. Misra, P.R. Cook, and G. Wang. Tapestrea: Sound scene modeling by example. In *ACM SIGGRAPH 2006 Sketches*, page 177. ACM, 2006. ISBN 1595933646.

[61] A. Misra, G. Wang, and P. Cook. Musical Tapestry: Recomposing Natural Sounds†. *Journal of New Music Research*, 36(4):241–250, 2007. ISSN 0929-8215.

[62] William Moss, Hengchin (Yero) Yeh, Jeong-Mo Hong, Ming C. Lin, and Dinesh Manocha. Sounding Liquids: Automatic Sound Synthesis from Fluid Simulation. *ACM Transactions on Graphics*, 28(4):1–12, August 2009. ISSN 07300301. doi: 10.1145/1559755.1559763. URL http://portal.acm.org/citation.cfm?doid=1559755.1559763.

[63] E. Murphy, M. Lagrange, G. Scavone, P. Depalle, and C. Guastavino. Perceptual Evaluation of a Real-time Synthesis Technique for Rolling Sounds. In *Conference on Enactive Interfaces*, Pisa, Italy, 2008.

[64] J.F. O'Brien, C. Shen, and C.M. Gatchalian. Synthesizing sounds from rigid-body simulations. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 175–181. ACM New York, NY, USA, 2002.

[65] S. O'Modhrain and G. Essl. PebbleBox and CrumbleBag: tactile interfaces for granular synthesis. In *Proceedings of the Conference for New Interfaces for Musical Expression (NIME)*, page 79. National University of Singapore, 2004.

[66] D. O'Regan and A. Kokaram. Multi-resolution sound texture synthesis using the dual-tree complex wavelet transform. In *Proc. 2007 European Signal Processing Conference (EUSIPCO)*, 2007. URL http://www.eurasip.org/Proceedings/Eusipco/Eusipco2007/Papers/A3L-B03.pdf.

[67] J.R. Parker and B. Behm. Creating audio textures by example: tiling and stitching. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 4:iv–317–iv–320 vol.4, May 2004. ISSN 1520-6149. doi: 10.1109/ICASSP.2004.1326827.

[68] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the Cuidado project. Technical Report version 1.0, Ircam – Centre Pompidou, Paris, France, April 2004. URL http://www.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf.

[69] Leevi Peltola, Cumhur Erkut, P.R. Cook, and Vesa Valimaki. Synthesis of hand clapping sounds. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(3):1021–1029, March 2007. ISSN 1558-7916. doi: 10.1109/TASL.2006.885924. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4100694.

[70] Gabriel Peyré. Oriented patterns synthesis. Technical report, Unité Mixte de Recherche du C.N.R.S. No. 7534, 2007. URL http://www.ceremade.dauphine.fr/preprints/consult/cmd_by_years.php.

[71] Cecile Picard, Nicolas Tsingos, and François Faure. Retargetting Example Sounds to Interactive Physics-Driven Animations. In *AES 35th International Conference, Audio in Games*, London, UK, 2009.

[72] K. Popat and R. W. Picard. Cluster-based probability model and its application to image and texture processing. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 6(2):268–84, January 1997. ISSN 1057-7149. doi: 10.1109/83.551697. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=551697.

[73] Xavier Rodet and Phillipe Depalle. A new additive synthesis method using inverse Fourier transform and spectral envelopes. In *Proceedings of the International Computer Music Conference (ICMC)*, October 1992.

[74] Nicolas Saint-Arnaud. *Classification of Sound Textures*. PhD thesis, Universite Laval, Quebec, MIT, 1995.

[75] Nicolas Saint-Arnaud and Kris Popat. Analysis and synthesis of sound textures. In *in Readings in Computational Auditory Scene Analysis*, pages 125–131, 1995.

[76] R. Murray Schafer. *The Soundscape*. Destiny Books, 1993. ISBN 0892814551. URL http://www.amazon.com/Soundscape-R-Murray-Schafer/dp/0892814551.

[77] Diemo Schwarz. *Data-Driven Concatenative Sound Synthesis*. Thèse de doctorat, Université Paris 6 – Pierre et Marie Curie, Paris, 2004. URL http://mediatheque.ircam.fr/articles/textes/Schwarz04a.

[78] Diemo Schwarz. Concatenative sound synthesis: The early years. *Journal of New Music Research*, 35(1):3–22, March 2006. Special Issue on Audio Mosaicing.

[79] Diemo Schwarz. Corpus-based concatenative synthesis. *IEEE Signal Processing Magazine*, 24(2):92–104, March 2007. Special Section: Signal Processing for Sound Synthesis.

[80] Diemo Schwarz and Norbert Schnell. Descriptor-based sound texture sampling. In *Proceedings of the International Conference on Sound and Music Computing (SMC)*, pages 510–515, Barcelona, Spain, July 2010.

[81] Diemo Schwarz and Norbert Schnell. A modular sound descriptor analysis framework for relaxed-real-time applications. In *Proc. ICMC*, New York, NY, 2010.

[82] Diemo Schwarz, Grégory Beller, Bruno Verbrugghe, and Sam Britton. Real-Time Corpus-Based Concatenative Synthesis with CataRT. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, pages 279–282, Montreal, Canada, September 2006.

[83] Diemo Schwarz, Roland Cahen, and Sam Britton. Principles and applications of interactive corpus-based concatenative synthesis. In *Journées d'Informatique Musicale (JIM)*, GMEA, Albi, France, March 2008. URL http://mediatheque.ircam.fr/articles/textes/Schwarz08a/index.pdf.

[84] G. Strobl. Parametric Sound Texture Generator. Msc thesis, Universität für Musik und darstellende Kunst, Graz; Technische Universitäåt Graz, 2007. URL http://en.scientificcommons.org/43580321.

[85] G. Strobl, G. Eckel, and D. Rocchesso. Sound texture modeling: A survey. In *Proceedings of the International Conference on Sound and Music Computing (SMC)*, 2006.

[86] Patrick Susini, Stephen McAdams, Suzanne Winsberg, Yvan Perry, Sandrine Vieillard, and Xavier Rodet. Characterizing the sound quality of air-conditioning noise. *Applied Acoustics*, 65-8:763–790, 2004. URL http://articles.ircam.fr/textes/Susini04b/.

[87] N. Tsingos, E. Gallo, and G. Drettakis. Perceptual audio rendering of complex virtual environments. In *ACM SIGGRAPH 2004 Papers*, pages 249–258. ACM, 2004.

[88] A. Valle, V. Lombardo, and M. Schirosa. Simulating the Soundscape through an Analysis/Resynthesis Methodology. *Auditory Display*, pages 330–357, 2010.

[89] Andrea Valle, Vincenzo Lombardo, and Mattia Schirosa. A graph-based system for the dynamic generation of soundscapes. In *Proceedings of the 15th International Conference on Auditory Display*, pages 217–224, Copenhagen, 18–21 May 2009.

[90] Kees van den Doel, Paul G. Kry, and Dinesh K. Pai. Foleyautomatic: physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 537–544, New York, NY, USA, 2001. ACM. ISBN 1-58113-374-X. doi: http://doi.acm.org/10.1145/383259.383322. URL http://doi.acm.org/10.1145/383259.383322.

[91] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. *Spatialized Synthesis of Noisy Environmental Sounds*, pages 392–407. Springer-Verlag, 2010. URL http://www.springerlink.com/index/J3T5177W11376R84.pdf.

[92] C. Verron, M. Aramaki, R. Kronland-Martinet, and G. Pallone. Contrôle intuitif d'un synthétiseur d'environnements sonores spatialisés. In *10eme Congres Français d'Acoustique*, 2010. URL http://cfa.sfa.asso.fr/cd1/data/articles/000404.pdf.

[93] Charles Verron. *Synthèse immersive de sons d'environnement*. Phd thesis, Université Aix-Marseille I, 2010.

[94] X. Zhu and L. Wyse. Sound texture modeling and time-frequency LPC. In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*, volume 4, 2004.