

Stato dell'arte nella sintesi di texture sonore

Diemo Schwarz*

(Traduzione ricevuta il 23 novembre 2015¹)

Sommario

La sintesi di *texture sonore* quali i suoni della pioggia, del vento, o di una folla, trova importanti applicazioni nel cinema e in varie forme di creatività multimediale, come giochi e installazioni. Tuttavia, nonostante ben definiti requisiti di naturalezza e di flessibilità, nessun metodo automatico si è ancora affermato. Dopo un chiarimento terminologico e relativo ai contesti d'utilizzazione della sintesi di suoni di questo tipo, daremo qui uno sguardo d'insieme sui molti metodi esistenti e sulle ancora poche implementazioni software disponibili, classificandoli in base ai differenti modelli di sintesi sottostanti, che siano di sintesi sottrattiva o additiva, granulare, concatenativa [*corpus-based concatenative synthesis*], e ancora sintesi mediante wavelet o mediante modelli fisici. Forniremo inoltre una rassegna dei metodi di analisi usati ai fini di questo tipo di sintesi, basati su criteri di segmentazione, sul rilevamento di proprietà statistiche, o sull'analisi del timbro e delle transizioni.

1 Introduzione

La sintesi di *texture sonore* quali i suoni della pioggia, del vento, o di una folla, trova importanti applicazioni nel cinema e in varie forme di creatività multimediale, come giochi e installazioni. Con *texture sonora* intendiamo generalmente una sonorità composta da molti micro-eventi, ma le cui caratteristiche restano stabili ad una scala di tempi più ampia, come avviene appunto nel suono della pioggia, del fuoco, del vento, nei suoni prodotti da giochi d'acqua, dal traffico automobilistico, dal movimento di masse di persone. Dobbiamo distinguere questo concetto rispetto a quello di *soundscape*, che invece indica l'insieme dei suoni che compongono una scena, alcuni dei quali a loro volta possono essere di tipo textureale.

Ci sono innumerevoli metodi per sintetizzare suoni del genere, secondo approcci molto differenti tra loro. Cercheremo di darne una classificazione e di riassumere lo stato dell'arte in questo settore. Cominceremo con una definizione dei termini e degli obiettivi d'utilizzo (sezioni 1.1, 1.2 e 1.3), per passare poi alla descrizione di metodi per la sintesi e l'analisi di texture sonore (sezioni 2 e 3) e dei primi software esistenti (sezione 4). Infine (sezioni 5 e 6), metteremo in evidenza le ricerche che appaiono più notevoli nel quadro di questa nostra ricostruzione complessiva.

*Real-Time Music Interaction team, IRCAM, Centre Pompidou – Parigi.

¹Il presente testo è la traduzione, a cura di Agostino Di Scipio, dell'articolo del 2011 «State of the Art in Sound Texture Synthesis», reperibile online al seguente url http://recherche.ircam.fr/pub/dafx11/Papers/30_e.pdf.

1.1 Definizione di texture sonora

Una prima definizione approfondita è quella avanzata da Saint-Arnaud [75] e Saint-Arnaud e Popat [76] a margine di vari esperimenti sulla percezione e generazione di texture sonore. È possibile riassumerla con la seguente analogia visiva:

Una texture sonora è come la trama di una carta da parati [wallpaper]: localmente può avere o una struttura particolare o un andamento casuale [randomness], ma nell'insieme queste caratteristiche locali rimangono costanti a scala globale.

Questa definizione è illustrata in Fig. 1. Si può arrivare dunque alla seguente definizione operativa:

1. le texture sonore sono formate da elementi sonori di base, o atomi;
2. questi elementi atomici si susseguono secondo un pattern di livello superiore che può essere periodico o casuale, o di entrambi i tipi;
3. le caratteristiche a livello generale rimangono le stesse per periodi di tempo estesi (ciò implica che non vi può essere un messaggio complesso);
4. il pattern di livello superiore dev'essere esposto nella sua completezza nel giro di pochi secondi ('tempo di attenzione');
5. una certa casualità a livello superiore è accettabile nella misura in cui gli elementi sonori all'interno del 'tempo di attenzione' sono in numero sufficiente ad evidenziare le proprietà casuali.

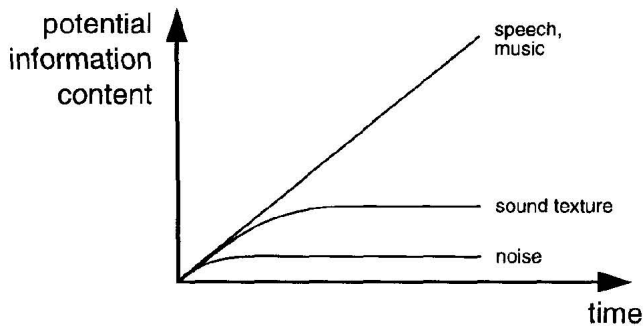


Figura 1: Contenuto di informazione potenziale di una texture sonora in funzione del tempo (da Saint-Arnaud e Popat [76]).

1.1.1 Ciò che una texture sonora *non* è

Per chiarire il concetto, può essere d'aiuto provare a dare una definizione negativa. Escludiamo dunque quanto segue:



Figura 2: Esempi di texture visive, sia naturali sia sintetiche, ‘orientate all’oscillazione’ (da Peyré [71]).

Suoni da contatto nell’interazione tra oggetti Impatti, frizioni, rotolamenti: tali suoni, pur trattati in certi lavori dedicati alla sintesi della texture sonora [1, 15, 49, 66, 89], contraddicono le proprietà di struttura tipo ‘carta da parati’ che abbiamo sopra indicato.

Paesaggi sonori [*sound scapes*] Sebbene siano talvolta trattati come texture, anche perché in effetti ne contengono sempre, i paesaggi sonori comprendono eventi sonori ricchi di informazione (come spiegheremo nella sezione 1.1.2).

Sound design La letteratura in questo campo spesso propone metodi utili per la sintesi di texture sonore; tuttavia il sound design in generale è un ambito più ampio di elaborazione creativa, volto a generare suoni da interazione e paesaggi sonori, oltre che texture [10, 12, 59–62].

In certi esempi di composizione e di performance musicale, la locuzione *texture sonora* [*sound texture*] viene usata per indicare materiale sonoro *non-tonico* e *non-percussivo*, o materiale musicale *non-armonico* e *non-ritmico*.

Si veda inoltre Strobl [86] per un’indagine sul termine *texture* in contesti diversi da quelli relativi al suono, come nei tessuti, in tipografia, in gastronomia.

1.1.2 Paesaggi sonori

Poiché le texture sonore costituiscono una parte vitale del paesaggio sonoro, è utile dare qui una breve introduzione alla classificazione e generazione automatica di paesaggi sonori. Inoltre, a sua volta chi si occupa di paesaggi sonori inevitabilmente guarda anche alla sintesi e all’organizzazione di texture sonore.

I primi tentativi in vista di una definizione e classificazione dei paesaggi sonori sono stati quelli di Raymond Murray Schafer [77], che distingue livelli diversi – come tonica [keynote], segnale [signal] e impronta sonora [sound mark] – e propone una tassonomia di referenti semantici inclusiva di attributi di tipo socio-culturale e presi dall’acustica ecologica.

In una prospettiva di ecologia acustica, invece, Gavin Gaver [37] propone una classificazione in base agli attributi fisici del suono e al tipo di interazione tra materiali.

Attualmente molteplici sono le ricerche sul paesaggio sonoro che rivestono interesse per la sintesi della texture sonora [8, 9, 34, 59–62, 90, 91].

1.2 Sintesi della texture: tentativi di classificazione esistenti

Come punto di partenza, consideriamo il tentativo di definizione di texture sonora proposto da Strobl et al. [87]. I metodi che vi sono presi in esame, appaiono divisi in due gruppi:

Metodi provenienti dalla grafica computazionale Trasferire nel contesto della sintesi del suono metodi usati per la generazione di texture visive [23, 65, 68]. Si veda in Fig. 2 qualche esempio di immagine texturale.

Metodi provenienti dalla musica informatica Applicare le tecniche di sintesi del suono usate nella *computer music* e nella sintesi del parlato [4, 6, 17, 43, 44, 95].

In una nuova rassegna degli strumenti usati nell'ambito generale del sound design e della composizione musicale [59], si propone una classificazione identica a quella che vedremo più avanti (nella sezione 2): vi si afferma che classi differenti di suoni richiedono differenti strumenti di lavoro, e viene proposta una lista di corrispondenze possibili tra differenti sonorità e metodi di sintesi che si prestano meglio ad evocarle.

In un articolo di Filatriau e Arfib [32] vengono passati in rassegna vari algoritmi per la sintesi della texture sonora dal punto di vista dei controlli e dei dispositivi di controllo. Ciò rende opportuno ricollegarsi ai diversi contesti di utilizzo pratico, come faremo ora.

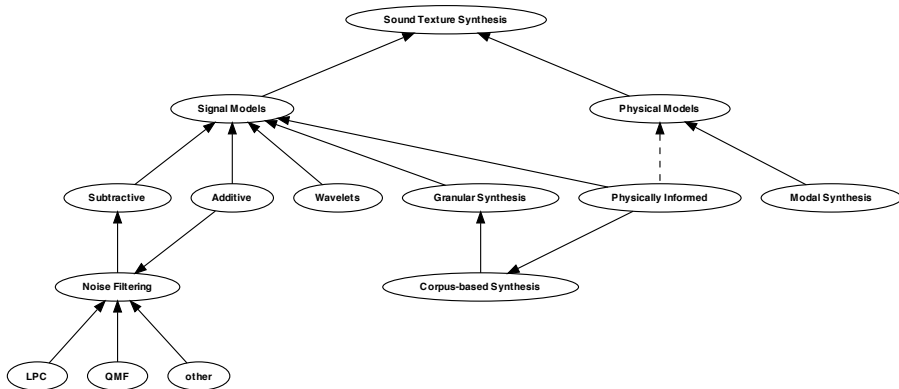


Figura 3: Classificazione gerarchica di metodi di sintesi della texture sonora.

1.3 Usi e significati differenti

È importante notare che nella letteratura su questo argomento esiste un certo rischio di confusione in rapporto al preciso significato della locuzione texture sonora, e che ciò dipende dalle intenzioni d'uso. Possiamo distinguere due destinazioni d'uso più frequenti:

Sintesi di texture espressive Qui lo scopo è la generazione interattiva del suono nel contesto della composizione e della performance musicale, o nel contesto di altre forme di arte sonora, comunque nel senso di uno strumento musicale digitale [Digital Musical Instrument, DMI] a fini espressivi. Il termine texture sonora viene allora spesso adoperato per distinguere il materiale così generato da suoni di tipo tonico o percussivo, cioè per denotare in modo generico un materiale caratterizzato nel timbro più che nell'altezza o nel ritmo.

Alcuni dei metodi impiegati a tale scopo possono dar luogo anche a texture evocative di sonorità naturali, come osserva Di Scipio [17]. Tuttavia, in contesti del genere non sono stati compiuti approfondimenti sullo spazio parametrico adoperabile in tal senso, e rimane quindi all'utente (al musicista) forzare i processi verso sonorità di tipo naturale. Ciò vale anche per uno dei lavori di ricerca già citati [32] e per altri che ne sono scaturiti [33].

Sintesi di texture naturali Qui lo scopo è di sintetizzare suoni texturali, sia prodotti dall'uomo sia prodotti dall'ambiente circostante, facenti parte del paesaggio sonoro complessivo, che tornino utili in contesti applicativi quali il cinema, i videogiochi, o altri contesti di produzione audio-visiva. Ci si sforza dunque di conseguire un certo grado di realismo (come nel rendering di immagini texturali fotorealistiche), sebbene poi in molti casi sia sufficiente conseguire una *sintesi texturale credibile* di tipo puramente simbolico o impressionistico, talché la texture ottenuta riesca a portare l'informazione ambientale desiderata anche senza effetti davvero realistici, come per esempio nella progettazione urbanistica. La maggior parte degli esempi che discuteremo qui mira a questo tipo di risultato.

2 Classificazione dei metodi di sintesi

Proporremo ora una classificazione dei metodi esistenti per la sintesi della texture sonora. Sembra piuttosto sensato dividere i differenti approcci in base ai processi di sintesi adoperati (ed in base ai processi di analisi, se effettivamente applicati):

- filtraggio di rumore (sezione 2.1) e sintesi additiva con sinusoidi (sezione 2.2.);
- modellazione fisica (sezione 2.3) e modelli di segnale informati in senso analogo;
- rappresentazione e sintesi mediante wavelet (sezione 2.4);
- sintesi granulare (sezione 2.5) e sue estensioni del tipo 'sintesi concatenativa' fondata su estese basi di dati (sezione 2.6);
- metodi non-standard, come mappe frattali o caotiche (sezione 2.7).

La Fig. 3 offre una visione d'insieme di queste diverse classi, e le loro relazioni interne. Altre possibilità classificatorie potrebbero riguardare altri aspetti: fino a che punto un metodo dipenda da un modello specifico, o fino a che punto è guidato da dati; quali sono le potenzialità

per la generazione del suono in tempo reale; se viene valutata l'attendibilità del metodo in base a test di ascolto. Alcuni di questi aspetti sono discussi nella sezione 5.

2.1 Sintesi sottrattiva

Il filtraggio di rumore è il metodo 'classico' per la sintesi di texture sonore, spesso usato a partire da una modellazione specifica della sorgente sonora.

In base alla particolare definizione operativa di cui abbiamo parlato sopra, Saint-Arnaud e Popat [76] costruiscono un primo modello di sintesi che adopera rumore filtrato mediante un sistema a sei bande con filtri in quadratura a specchio [6-band Quadrature Mirror filtered noise].

In altri contributi [4, 95] viene applicata l'analisi mediante predizione lineare a cascata sia nel dominio del tempo che della frequenza [cascaded time and frequency domain linear prediction, CTFLP], in particolare per risintetizzare il rumore sullo sfondo di eventi rilevati in primo piano (vedremo più avanti, nella sezione 3.1.3, quali tecniche di rilevamento sono utilizzate). Gli eventi sono sequenziati mediante una distribuzione di Poisson. Centrale, in questi contributi, è la volontà di ridurre la quantità di dati da trasmettere con codifica a basso bitrate.

In McDermott, Oxenham e Simoncelli [54] vengono applicati al filtraggio di rumore dati risultanti da un'analisi statistica (vi torneremo in 3.2), limitandone l'uso a texture piuttosto statiche e dall'altezza indeterminata, come nel caso della pioggia e altri suoni d'acqua, o come in suoni di fuoco.

In Peltola et al. [70] si studia il suono ottenuto con differenti modi di battere le mani, usando filtri regolati sulla base dell'analisi di suoni del genere, e combinando i vari risultati ottenuti per farne l'applauso di una folla, secondo un modello statistico con vari 'gradi di entusiasmo' e vari modi di occupare lo spazio da parte della folla.

La ben nota e apprezzata serie di esempi proposta da Andy Farnell [31], disponibile anche online², fornisce parecchi suoni di carattere texturale e le rispettive *patches* di sintesi scritte in PURE DATA, a partire da oscillatori e filtri attentamente regolati secondo una comprensione intuitiva del fenomeno simulato, come si può vedere in questa citazione:

“Qual è la natura [del suono] della pioggia? Cosa lo causa?” Secondo certe filosofie solipsistiche a buon mercato, la pioggia “mi cade sempre addosso”, ma ciò non ci fornisce alcun aiuto. Consideriamo invece che si tratta di particelle quasi sferiche di acqua, con diametro tra 1 e 3 mm, che si muovono a velocità costante impattando infine su materiali diversi con un flusso tipico di circa 200 al secondo per metro quadrato. Ogni goccia perviene alla stessa velocità terminale, non vi sono gocce più veloci e gocce meno veloci. Sono tutte approssimativamente delle stesse dimensioni, essendo causate dalle stesse condizioni atmosferiche di partenza, dunque non ci sono gocce più pesanti e meno pesanti. Infine, la forma di una goccia di pioggia non è quella di una lacrima, come si ritiene normalmente, ma quella di una sfera quasi perfetta. Il fattore che impedisce il formarsi di un suono completamente uniforme, e che determina quindi una certa varietà di altezza tra goccia e goccia, sta nel materiale su cui avviene l'impatto. Certe volte si tratta di foglie d'albero, certe volte di un pavimento, o di un tetto, certe altre volte ancora di una pozzanghera d'acqua formata da gocce precedenti³.

²http://obiwannabe.co.uk/tutorials/html/tutorials_main.html

³[La citazione è tratta dalla pagina web http://www.obiwannabe.co.uk/tutorials/html/tutorial_rain.html, che corrisponde, con leggere varianti, a Farnell [31], p. 441.

2.2 Sintesi additiva con sinusoidi e rumore

Il rumore filtrato viene spesso considerato complementare al concetto di ‘oscillatore’ quale usato nella sintesi mediante somma di parziali sinusoidali.

Nel progetto QCITY⁴ si effettua una simulazione, non in tempo reale, di suoni del traffico, mediante una rappresentazione mista di sinusoidi e di rumore calibrati secondo misure ottenute dalle varie fasi di funzionamento di un motore, dal tubo di scarico, da vari elementi di smorzamento, ecc. Il modello permette di simulare diverse densità di traffico, differenti velocità di scorrimento, diverse tipologie di veicoli e di asfalto, l’assorbimento acustico di pareti, ecc. [39]. Il calcolo degli esempi sonori può richiedere ore.

In varie circostanze Charles Verron ha proposto metodi informati alla modellazione fisica [92–94] a partire da una classificazione tratta da Gaver [37] che distingue tre tipologie diverse di suoni ambientali: liquidi, solidi e aereodinamici. I metodi proposti per l’impatto tra solidi, per i suoni del vento, e per getti, gocce e onde d’acqua, si basano su cinque atomi sonori empiricamente definiti e parametrizzati: ‘impatto modale’, ‘impatto di rumore’, ‘impatto chirp’ [a frequenza variabile], ‘rumore a banda stretta’, ‘rumore a banda larga’. Ogni modello ha da due a quattro parametri di basso livello (con l’eccezione di ben trentadue ampiezze di banda per il rumore a banda larga).

Verron poi cerca scrupolosamente le corrispondenze tra parametri di livello superiore – come forza del vento, intensità di pioggia, dimensione dell’onda marina – con i controlli parametrici di livello inferiore e le rispettive distribuzioni probabilistiche di densità.

Il processo di sintesi, nel caso di Verron, sfrutta il metodo FFT-1 [74] esteso in modo tale da includere una codifica spaziale, o meglio binaurale, già a livello della trasformata veloce di Fourier (FFT), operando poi, in fase di risintesi, uno stadio di trasformata inversa (IFFT) per ciascuno dei canali in uscita⁵.

2.3 Modellazione fisica

La modellazione fisica può essere applicata alla sintesi della texture sonora con la limitazione che un diverso modello va specificato per ciascuna classe di suoni – per esempio suoni da frizione e da rotolamento, rumori di macchinari, suoni aerodinamici [64, 65]; quest’ultima possibilità implica l’estrazione del suono dell’impatto e la valutazione percettiva del realismo conseguito con la sintesi di suoni da rotolamento [si veda anche [48]]. Spesso si usano modelli di risonanza modale [20] nei quali i modi [di vibrazione] sintetizzabili senza sforzo [computazionale] sono precalcolati a partire da simulazioni [computazionalmente] costose del comportamento del corpo rigido.

Altri metodi di sintesi basati su segnali sono descritti come fisicamente informati, cioè ispirati per lo più intuitivamente a modelli fisici [13, 14], nel senso che essi controllano modelli di segnale mediante un modello fisico del comportamento della sorgente sonora. Si vedano anche Cook [12], Picard, Tsingos e Faure [72] e Verron [93] e il libro di testo di Menzies [56], su cui torneremo più avanti.

La sintesi di suoni liquidi descritta in Doel [21] è una combinazione tra un modello di segnale ispirato a modelli fisici, per singole bolle sonore – basato su Mallock [52] – ed un

⁴<http://qcity.eu/dissemination.html>

⁵Al sito <http://www.charlesverron.com/thesis/> si possono trovare vari esempi sonori binaurali (a volte leggermente artificiali, all’ascolto) e un video che illustra i controlli sui parametri di livello superiore, valutando anche le differenze per sorgenti sonore puntiformi e sorgenti sonore spazialmente estese.

modello empirico, fenomenologico, del comportamento statistico delle bolle d'aria, ed offre una grande varietà di risultati, da singole gocce d'acqua alla pioggia, da bolle d'aria nell'acqua a flussi d'acqua e torrenti.

Un esempio estremo è la sintesi di suoni di liquidi mediante simulazione di fluidi [63], dove l'informazione per il controllo della sintesi viene derivata dalle relazioni armoniche della sfera nelle individuali bolle d'aria (fino a quindicimila bolle)⁶.

2.4 Wavelets

Il primo esempio di sintesi di texture sonore basato sulla scomposizione multiscale del segnale mediante coefficienti wavelet è dovuto ai contributi di El-Yaniv, Werman e Dubnov [27] e di Dubnov et al. [23], ripresi poi in Kersten e Purwins [46]. Qui, la rappresentazione multiscale del segnale viene ricampionata riorganizzando l'ordine di ramificazione dei coefficienti. Ogni percorso attraverso l'albero di ramificazioni viene usato per sintetizzare, invertendo la trasformata wavelet, solo una piccola porzione di suono.

Questi contributi traggono ispirazione dall'analisi e sintesi di immagini texturali, e cercano di modellare sia le dipendenze temporali sia le dipendenze gerarchiche relative alle varie scale di rappresentazioni in uso. Il lavoro di Kersten e Purwins si colloca ad uno stadio di approfondimento ancora precoce in cui il suono complessivo della texture è già riconoscibile come tale (stando a test quantitativi svolti in tal senso), ma la struttura risultante sembra troppo sottile e sgranata, perché effettivamente le dipendenze sequenziali della texture in esame non vengono modellate, per cui gran parte del dettaglio nella struttura fine del segnale si perde. Ciò viola la caratteristica di autocorrelazione che invece si è dimostrato essere importante sia nelle texture sonore sia in quelle visive [30, 54].

In Miner e Caudell [57] viene usata l'analisi wavelet per ricostruire un suono-modello avente carattere complessivamente stocastico. Le parametrizzazioni dell'analisi wavelet restituiscono una certa varietà di risultati sonori tutti leggermente diversi dal suono considerato, a partire da un unico insieme di modelli dinamici.

Un altro approccio basato su wavelet è quello di Kokaram e O'Regan [47] e O'Regan e Kokaram [67], a sua volta sviluppato a partire da un algoritmo proposto in Efros e Leung [24] per la sintesi di immagini texturali. Viene effettuata una sintesi multicanale che adopera dimensioni piuttosto grandi di segmenti di segnale, adeguati alle più diverse sorgenti (parole, grida di neonato, accelerazione di autoveicoli, colpi di tamburo). I risultati ottenuti in fase di risintesi sono convincenti e spesso privi di artefatti.

2.5 Sintesi granulare

Nella sintesi granulare si usano frammenti di un suono iniziale, che può essere anche un suono registrato, e possibilmente un modello statistico di (ri)composizione di tali 'grani sonori' nel tempo. L'esplorazione di questo metodo ai fini della sintesi di texture sonore è stata affrontata da vari ricercatori: Bascou e Pottier [6], Dubnov et al. [23], El-Yaniv, Werman e Dubnov [27], Fröjd e Horner [36], Hoskinson [43], Hoskinson e Pai [44] e Parker e Behm [68]. La dimensione ottimale del grano dipende dalla scala temporale tipica della texture considerata: se sufficientemente grande, la distribuzione di eventi temporali nel breve periodo

⁶<http://gamma.cs.unc.edu/SoundingLiquids>

è preservata all'interno del grano stesso, pur permettendo di creare una struttura non ripetitiva su un periodo di tempo più esteso.

In Lu, Wenyin e Zhang [50] vengono ricombinati segmenti molto brevi di segnale, anche con trasposizioni, seguendo un modello fondato sulle probabilità di transizione (si veda sezione 3.4), cercando esplicitamente di prevenire ripetizioni sul tempo breve. La segmentazione sottostante, come vedremo, segue un criterio di 'punteggio di novità' [novelty score] nella forma di una matrice di similarità.

In Strobl [86] vengono studiati accuratamente metodi già introdotti in Hoskinson [43] e Hoskinson e Pai [44] e in Lu, Wenyin e Zhang [50], migliorandone i parametri e la fase di risintesi, e ottenendone texture percettivamente identiche a quelle prese a modello, provando poi anche a farne delle ibridazioni [86, capitolo 4]. Il lavoro di Gerda Strobl offre inoltre un'implementazione dei metodi di Lu, Wenyin e Zhang [50] nella forma di una patch PURE DATA in tempo reale⁷.

2.6 Sintesi concatenativa

Questa nuova forma di sintesi [78, 79] può essere vista come un'estensione della sintesi granulare, cioè come una sintesi granulare operante a partire da un database [11, 82, 84, 85]. Viene spesso qualificata come *corpus-based* nel senso che implica la navigazione e la selezione di pezzettini di segnale conservati in un database molto esteso, cioè in un *corpus* di dati. Ogni grano è posizionato secondo il proprio carattere sonoro in base a determinati descrittori audio, a loro volta estratti in rapporto all'altezza, all'intensità, alla brillantezza, e altri meta-dati di livello più generale attribuiti al suono che si intende risintetizzare. Ciò permette di esplorare interattivamente una grande quantità di dati sonori e di creare delle traiettorie al suo interno, producendo nuove evoluzioni timbriche pur preservando le sottigliezze nella microstruttura del segnale originale – il che è davvero importante per ottenere texture sonore convincenti.

Finney [34] utilizza un corpus di registrazioni non strutturate disponibili nel database collaborativo *freesound*⁸, come materiale di partenza per generare eventi e texture sonore nel contesto di una forma più generale di sintesi del paesaggio sonoro (si vedano i test qualitativi discussi più avanti, nella sezione 4). I campioni prelevati vengono segmentati (come vedremo in 3.1.2) e poi filtrati con filtri passa-basso e passa-alto. I residui reputati non conformi secondo le tecniche usate da Finney vengono rimossi dallo spettro. In fase di sintesi viene scelto un insieme di cinque segmenti reputati più vicini a quelli analizzati (come viene fatta questa scelta non viene spiegato esplicitamente).

In Schwarz e Schnell [84] si segnala che i metodi esistenti per la sintesi della texture sonora spesso prendono in considerazione l'intera estensione di una certa registrazione, di cui conservano le proprietà complessive cercando di evitare eventuali artefatti. Tuttavia, tali metodi in genere mancano di controllabilità in rapporto ai risultati da conseguire. Si propongono allora due nuovi metodi basati sulla modellazione statistica dei descrittori audio applicati a registrazioni di suoni texturali, usando istogrammi e modelli di mistura Gaussiani. I quali possono poi essere interpolati per governare la transizione della texture tra differenti registrazioni-obiettivo (per esempio, da pioggia leggera a pioggia pesante). I descrittori audio dei suoni registrati vengono derivati statisticamente a partire dal campionamento della trasformata inversa per controllare infine il processo di sintesi del suono in uscita; i controlli

⁷Si vedano gli esempi in http://www.netsoc.tcd.ie/~dee/STS_EUSIPCO.html

⁸<https://freesound.org/>

però possono essere ottenuti anche navigando interattivamente nello spazio dei descrittori stessi⁹ (nella sezione 4 parleremo dell'applicazione CATART che è servita nei test interattivi per la valutazione di queste procedure). Per meglio coprire lo spazio dei descrittori relativi al suono da conseguire, in Schwarz e Schnell [84] il corpus di dati viene allargato automaticamente con varianti dei suoni originali ottenute dalle trasformazioni di volta in volta applicate, ma aggiungendo al corpus di dati [non i suoni ma] solo i relativi descrittori e parametri di trasformazione. In Masurelle [53] è stata condotta una prima validazione percettiva dei descrittori usati per suoni di vento, pioggia, e onde del mare, con test soggettivi ispirati a precedenti studi della percezione di sonorità ambientali [58, 88].

Il lavoro esposto in Picard, Tsingos e Faure [72] presenta aspetti simili alla sintesi concatenativa, ma usa una procedura di selezione di grani guidata da un modello fisico.

In Dobashi, Yamamoto e Nishita [18, 19] si impiega una sintesi concatenativa che sfrutta a sua volta un database ispirato a modelli fisici di suoni aerodinamici, come i suoni del vento o quelli del movimento di spade. Questi autori precalcolano i dati di emissione sonora aerodinamica di sorgenti puntiformi mediante una simulazione della turbolenza [d'aria] computazionalmente onerosa, per differenti angoli e differenti velocità di movimento; dopo di che, consultando i dati precalcolati e sommandone i contributi parziali, possono generare interattivamente il suono di un oggetto che si muove nell'aria in modo complesso.

2.7 Metodi di sintesi non-standard

I metodi di sintesi non-standard, per esempio basati su frattali e mappe caotiche prodotte mediante iterazione di funzioni nonlineari, sono usati per la generazione di texture a scopi prevalentemente espressivi da Di Scipio [17] e Filatriau, Arfib e Couturier [33], specialmente quando i processi di sintesi sono controllati con dispositivi di ingresso di tipo gestuale [3, 32].

3 Metodi di analisi delle texture sonore

L'analisi di proprietà texturali del suono può riguardare il grado di segmentazione del suono (sezione 3.1), l'andamento delle sue qualità statistiche (sezione 3.2), in particolare per quanto riguarda il timbro (sezione 3.3), o gli stati di transizione tipici della sorgente sonora presa in esame (sezione 3.4).

3.1 Segmentazione

3.1.1 Rilevamento dell'attacco

In O'Modhrain e Essl [66] viene descritto un metodo di analisi granulare che gli autori chiamano *granificazione*, che essi applicano a suoni nati dall'interazione di molti elementi parziali (pietre in una scatola, cereali in una busta), e con tali dati di analisi controllano in modo espressivo un metodo di sintesi granulare (siamo quindi nel caso descritto nella sezione 1.3): usando un algoritmo di rilevamento dell'attacco basato su criteri di soglia, entro un tempo di riattivazione limitato, essi derivano i tempi di attacco, il livello sonoro (il primo picco successivo all'attacco vero e proprio) e alcuni contenuti spettrali (conto degli attraversamenti dello zero da parte del segnale in una finestra di 100 campioni successiva all'attacco rilevato).

⁹Si vedano gli esempi in http://imtr.ircam.fr/imtr/Sound_Texture_Synthesis

Tali parametri vanno a controllare il tempo di attivazione del sintetizzatore granulare, il suo livello sonoro e vari fattori di trasposizione – si veda anche Essl e O’Modhrain [29].

In Lee, Depalle e Scavone [49] si effettua una stima degli eventi di contatto in una sequenza di suoni da rotolamento, sottoponendo il segnale ad un filtro passa-alto e applicando una soglia di energia. I segmenti rilevati a loro volta vengono modellati, in fase di risintesi, mediante filtri LPC su varie bande di frequenza.

3.1.2 Rilevamento di modifiche spettrali

Come si è anticipato, Lu, Wenyin e Zhang [50] segmentano il segnale in base ad un ‘punteggio di novità’, assegnato seguendo i coefficienti di analisi indicati come MFCC [*Mel-Frequency Cepstral Coefficients*]. Vedremo più avanti (sezione 3.4) che ciò è utile anche a valutare le probabilità di transizione [tra configurazioni spettrali diverse]. La procedura di analisi di questi autori è stata ripresa e migliorata in Strobl [86].

Finney [34] a sua volta usa un metodo di segmentazione di registrazioni ambientali che sfrutta un *criterio di informazione bayesiano* (BIC) per valutare i coefficienti MFCC [2], ma introduce anche una relazione tra la lunghezza dei segmenti rilevati ed il tipo di suono preso in esame: la lunghezza del segmento dovrà avere una certa corrispondenza con la lunghezza tipica dell’evento sonoro complessivo.

3.1.3 Segmentazione LPC

In Kauppinen e Roth [45], il segnale viene suddiviso in segmenti o frames localmente stazionari, mediante un treno di impulsi LPC [*linear predictive coding*, codifica a predizione lineare]. Gli autori ottengono la lunghezza ottimale del frame come misura statistica a seguito della predizione sul breve e sul lungo periodo. La soglia di picco è automaticamente adattata mediante un filtro mediano della derivata dello spettro.

In modo analogo, in Zhu e Wyse [95] vengono rilevati gli eventi sonori in primo piano [rispetto ad una texture continua sullo sfondo] mediante algoritmi LPC nel dominio della frequenza: rimuovendo gli eventi rilevati, è possibile quindi isolare il ‘frastuono’ del rumore di fondo. (Nella sezione 2.1 abbiamo accennato alla sintesi sottrattiva usata in questa ricerca per modellare il rumore così isolato).

3.1.4 Wavelets

In Hoskinson [43] e Hoskinson e Pai [44] la sorgente sonora viene segmentata in *grani naturali*, definiti in base ai *minima* nelle variazioni di energia all’interno delle prime 6 bande wavelet, dove il suono risulta di solito maggiormente stabile.

3.1.5 Analisi di componenti atomiche

Altri metodi effettuano l’analisi del suono di cui si intende modellare la texture in termini di componenti di evento e di componenti spettrali, per poi darne una ricombinazione statistica in sede di sintesi [50, 51, 60, 62, 68]. Si tratta di metodi ispirati alla modellazione di suoni da impatto, come in Ahmad, Hacıhabiboglu e Kondoz [1].

In Bascou [5] e Bascou e Pottier [7] il suono viene scomposto in atomi tempo-frequenza e si procede alla valutazione della corrispondenza tra questi atomi ed un dizionario di grani ‘caratteristici’ costruito empiricamente dagli autori stessi a partire dal suono da scomporre.

3.2 Analisi di proprietà statistiche

Dubnov et al. [23] e El-Yaniv, Werman e Dubnov [27] applicano alla sintesi di texture sonore l'algoritmo basato sulle catene di Markov di El-Yaniv, Fine e Tishby [26], costruendo il modello statistico discreto di una sequenza di percorsi nello spazio di rappresentazione wavelet del segnale (ne abbiamo accennato nella sezione 2.4).

In Zhu e Wyse [95], una volta che gli eventi in primo piano sono separati dal rumore di fondo, se ne valuta la densità con una segmentazione mediante LPC (si veda 3.1.3). Masurelle [53], a partire da esperimenti precedenti di O'Modhrain e Essl [66], ha sviluppato un metodo semplice per stimare la densità di eventi di impatto, che risulta applicabile per esempio nel caso di suoni della pioggia. Anche Doel [21] cita varie ricerche di analisi statistica di questo tipo di suoni.

In McDermott, Oxenham e Simoncelli [54] si fa uso di un'analisi statistica ispirata a ricerche di neurofisiologia e riguardante la curtosi [cioè lo spessore variabile delle curve di distribuzione] di energia in determinate bande di frequenza. I dati ricavati in questo modo sono applicati alla sintesi mediante filtraggio di rumore (il metodo è stato poi applicato anche alla classificazione di suoni ambientali, si veda Ellis, Zeng e McDermott [28]).

3.2.1 Analisi non orientata alla sintesi

Certi autori mirano all'analisi e alla classificazione di texture sonore senza necessariamente cercarvi implicazioni rilevanti per una eventuale fase di sintesi. È il caso di Dubnov e Tishby [22], che usano spettri di ordine superiore per la sola classificazione di suoni ambientali, e di Desainte-Catherine e Hanna [16] che propongono dei descrittori statistici validi genericamente per sonorità ricche di rumore.

In Grill [38], parlando di una particolare installazione sonora, si descrive un sistema che riproduce campioni presi da un corpo di dati, valutati in tempo reale, corrispondenti ai suoni casuali prodotti dai visitatori. Se la fase di sintesi è estremamente semplice, il metodo di *matching* [ricerca delle corrispondenze] è notevole per l'uso di *patterns di fluttuazione*, cioè dello spettro di modulazione per tutte le bande di frequenza (misurate in *barks*) di un segmento di texture lungo tre secondi. Il vettore di dati risultante è costituito da ben 744 elementi, ma viene ridotto a ventiquattro [il numero di bande critiche, misurate in *barks*], prima di operare il matching finale.

3.3 Analisi delle qualità timbriche

In Hanna et al. [42] si fa notare che nel contesto del recupero di informazioni musicali [*Music Information Retrieval*, MIR] si è lavorato poco sulle caratteristiche specifiche di suoni ricchi di rumore, e si propone una classificazione in quattro sotto-classi: 'rumore colorato', 'rumore pseudo-periodico', 'rumore impulsivo' (pioggia, applausi) e 'rumore con componenti sinusoidali' (vento, rumori di strada, uccelli). Dato un segnale, questi autori ne rilevano le transizioni fra queste quattro categorie, con un insieme di misure probabilistiche bayesiane. L'idea può essere generalizzata in un modello di rappresentazione del suono basato su configurazioni stocastiche di sinusoidi [40, 41].

Soltanto i metodi di sintesi concatenativa basati su un corpus di dati provano a caratterizzare i suoni di una sorgente sonora in termini di descrittori audio percettivamente significativi [25, 78, 79, 82, 84, 85].

3.4 Clustering e modellizzazione delle transizioni

Saint-Arnaud [75] sfrutta un *modello di probabilità basato su gruppi* [Cluster based probability model], descritto anche in Popat e Picard [73], per raggruppare secondo varie medie statistiche un certo numero di atomi sonori relativi ad un segnale in ingresso. Gli atomi sono intesi come dati di ampiezza relativi al segnale in uscita da filtri, e sono misurati in rapporto al segmento o frame corrente ma anche in rapporto ad un intorno temporale [*neighbourhood mask*] dell'andamento passato del segnale, cercando così di codificare le transizioni tipiche che accadono nella texture sonora considerata. Nel suo lavoro, Saint-Arnaud propone una classificazione delle texture sonore ed effettua vari esperimenti, poi estesi in Saint-Arnaud e Popat [76], per costruire un modello di analisi basato sul filtraggio di rumore.

Come si è già segnalato, in Lu, Wenyin e Zhang [50] le probabilità di transizione vengono modellizzate con una matrice di similarità basata su coefficienti MFCC; in Hoskinson [43] e Hoskinson e Pai [44] vengono invece modellizzate in base alla continuità rilevata all'interno di un flusso di *grani naturali* – cioè, come si è già visto, in segmenti del segnale ricavati mediante analisi wavelet. Entrambi i metodi sono stati approfonditi e migliorati in Strobl [86].

4 Software disponibile

Rari sono i prodotti software, gratuiti o disponibili a pagamento, destinati alla sintesi di texture sonore, specialmente se consideriamo certi particolari tipi di texture. Il solo prodotto commerciale di cui personalmente sono a conoscenza è il 'simulatore di folle' CROWD CHAMBER¹⁰, che prende un file audio e lo elabora con varie tecniche di pitch-shifting e time-stretching. In realtà non si tratta di un sintetizzatore di texture, ma di un processore di segnale che aggiunge un 'effetto folla' a sonorità vocali preesistenti. Gli esempi non sono molto convincenti.

Finney [34] presenta un sistema di sintesi del paesaggio sonoro basato su tecniche concatenative di campioni preesistenti (si veda 2.6 e 3.1.2). Una parte del suo sistema è integrata al software GOOGLE STREET VIEW¹¹. È notevole la sua capacità di simulare il rumore del traffico a partire da singoli campioni del passaggio di automobili, classificati per tipo di auto e per velocità, e ricombinati probabilisticamente in funzione dell'ora e del numero di strade laterali (vi è anche la simulazione della trama luminosa dei fari delle auto, realizzata mediante clustering). In Finney e Janer [35] troviamo una valutazione di questo sistema, particolarmente in rapporto alla qualità immersiva dei paesaggi sonori così generati, mediante test soggettivi di ascolto con otto partecipanti. È interessante che, in questa procedura di valutazione, i paesaggi sintetici risultino sistematicamente più credibili della semplice registrazione dei luoghi particolari presi in considerazione.

Il sistema PHYA [55, 56] offre vari strumenti di lavoro, concepiti come processi di filtraggio e di risonanza ispirati a modelli fisici, al fine di sintetizzare suoni di impatto, di collisione e suoni di diverse superfici con cui si simula di entrare in contatto¹².

Il sistema CATART, che è un mio personale contributo [82], effettua una sintesi concatenativa basata su corpus di dati in tempo reale, con controlli interattivi, implementata in

¹⁰http://www.quikquak.com/Prod_CrowdChamber.html

¹¹http://www.w3.org/2010/06/w3car/soundscape_generation.pdf

¹²<http://www.tech.dmu.ac.uk/~dylan/z/phya/>

MAX/MSP (e librerie aggiuntive FTM&Co)¹³. È un sistema disponibile gratuitamente¹⁴ che permette di navigare una proiezione bidimensionale (o anche pluridimensionale) dello spazio dei descrittori di un database di segmenti sonori, col mouse o altro controller gestuale. Si tratta effettivamente di una sorta di estensione della sintesi granulare con accesso diretto, basato su contenuti, alle caratteristiche sonore desiderate; ciò rende possibile la creazione di texture sonore in evoluzione dinamica secondo controlli precisi delle variazioni timbriche, conservando intatta la coerenza della micro-struttura del suono, almeno fintanto che i grani sono di durata abbastanza estesa (come si è detto nella sezione 2.6). Una possibilità aggiuntiva è quella di aumentare la densità texturale, elevando la velocità di attivazione dei segmenti sonori indipendentemente dalla loro lunghezza, e quindi stratificando più unità di generazione del suono, il che risulta di grande efficacia nel caso di suoni texturali come quelli della pioggia, del vento, dell'acqua o di folle di persone¹⁵.

In CATART i descrittori sono calcolati da una struttura di analisi modulare [83]. Accanto a descrittori usuali, come frequenza fondamentale, periodicità, intensità, ci sono descrittori spettrali: 'centroide spettrale', 'acutezza' [*sharpness*], 'piattezza' [*flatness*], 'energia a media e alta frequenza', 'contenuto di frequenza elevata', 'pendenza spettrale' (espressa come coefficiente di autocorrelazione del primo ordine), ed 'energia complessiva'. Dettagli in merito si possono trovare in Peeters [69] e Schwarz [80].

5 Discussione

Per quanto riguarda la dipendenza da uno specifico modello, possiamo vedere che i metodi presentati riflettono una dicotomia piuttosto netta, con approcci basati su regole, da una parte, e approcci basati su dati, dall'altra: i metodi che usano modelli fisici o modelli di basso livello del segnale (sezioni 2.2 e 2.3) partono quasi tutti da una modellazione molto specifica del processo di generazione della texture, con l'eccezione dei metodi basati su filtraggio di rumore e su analisi statistiche. I metodi che usano segmenti di segnale e coefficienti wavelet (sezioni 2.4, 2.5 e 2.6), essendo costitutivamente basati su dati, appaiono più generalmente applicabili a molte e differenti sonorità texturali, e sono molto più indipendenti da uno specifico modello di texture.

Inoltre, i modelli fisici non forniscono un legame diretto tra i loro parametri interni e le caratteristiche del suono risultante. Come nota Menzies [56],

in linea di principio, in ambiente virtuale un suono può essere riprodotto accuratamente attraverso una modellazione fisica accurata. Ma anche se ciò si realizzasse, non sarebbe sufficiente per chi, lavorando per esempio nel sound design e nella sonorizzazione cinematografica, ha bisogno di forgiare il suono secondo la propria immaginazione e secondo determinati riferimenti semantici: i modelli fisici espliciti sono spesso difficili da calibrare rispetto a comportamenti sonori desiderati, anche se controllati direttamente con parametri fisici.

In tal senso, i metodi ispirati alla modellazione fisica in modo più intuitivo risultano certamente più flessibili, ma possono ancora proporre parametri di sintesi non direttamente relativi a caratteristiche sonore effettivamente percepibili. Peraltro, i parametri usati nei

¹³<http://ftm.ircam.fr>

¹⁴<http://imtr.ircam.fr/imtr/CataRT>

¹⁵Cfr. nota 9.

modelli fisici e nei modelli di segnale possono andare bene per la simulazione di certe texture sonore, ma restano limitati ad un insieme molto ristretto di sfumature e, diversamente da metodi che usano segmenti sonori registrati (sintesi mediante wavelet e tecniche concatenative), tendono in ogni caso a non includere quegli aspetti del suono dovuti al contesto della sorgente sonora.

5.1 Percezione e interazione

Sono piuttosto rare, in generale, le ricerche che si occupano della percezione di texture sonore ambientali. Vi sono eccezioni [54, 58, 88], ma solo di recente si inizia a considerare seriamente una valutazione sistematica della qualità texturale ottenuta mediante sintesi, ricorrendo a test di ascolto formalmente organizzati [54, 58, 88]. Solo in Kokaram e O'Regan [47] e O'Regan e Kokaram [67], partendo da esempi precedenti [23, 27], si è pensato di definire sin dall'inizio un insieme condivisibile e comparabile di suoni-test per svolgere simili procedure di valutazione.

Infine, dobbiamo notare che in queste pagine ci siamo concentrati su modelli di sintesi e di analisi applicati a sonorità texturali dell'ambiente, trascurando le possibilità di controllo e di interazione che essi rendono praticabili. I controlli di tipo gestuale sembrano essere un approccio promettente in tal senso [3, 32, 53].

5.2 Letture utili

Pur avendo cercato, col nostro sguardo d'insieme, di accennare al maggior numero possibile di lavori di ricerca in questo settore, va detto che alcuni di essi rivestono un interesse superiore, e possono ben rappresentare quanto di più avanzato si sia fatto fino ad oggi. In particolare pensiamo a:

- Finney [34], per l'introduzione in questo contesto del concetto di paesaggio sonoro, per i metodi di segmentazione adoperati, e per la modellazione molto efficace dei rumori del traffico automobilistico;
- Verron [93] e Farnell [31], per l'accurata descrizione di metodi di sintesi di suoni ambientali ispirati alla modellazione fisica, ricca di intuizioni importanti su questo tipo di sonorità;
- Kokaram e O'Regan [47] e O'Regan e Kokaram [67] e Schwarz e Schnell [84] per aver conseguito quelli che per ora sembrano i risultati più convincenti.

6 Conclusioni

Come abbiamo visto, nonostante i problemi e i contesti d'uso siano ben delineati, nel corso degli ultimi anni la ricerca sulla sintesi della texture sonora non ha ancora condotto ad un metodo prevalente capace di soddisfare tutti i requisiti di realismo e di flessibilità. Effettivamente, poiché il requisito di realismo costituisce praticamente sempre la priorità principale, la flessibilità dei modelli e dei metodi implementati appare messa in secondo piano a favore di soluzioni pragmatiche, magari anche ricorrendo all'editing manuale in fase di post-produzione digitale di registrazioni di suoni texturali, o semplicemente mettendo in loop campioni preregistrati per applicazioni interattive quali i giochi elettronici.

In ogni caso, i risultati più avanzati conseguiti con la risintesi wavelet [47, 67] e con la sintesi granulare basata su descrittori audio [84] promettono un buon livello di applicabilità grazie alla qualità dei risultati sonori conseguiti.

7 Ringraziamenti

L'autore ringrazia Steven Kersten e tutti coloro che, avendo visionato la versione iniziale di questo articolo, hanno apportato correzioni, ulteriori riferimenti bibliografici e osservazioni pertinenti. Il lavoro è stato sostenuto in parte dalla Agence Nationale de la Recherche, in particolare nell'ambito del progetto Topophonie (ANR-09-CORD-022, <http://topophonie.com>).

Bibliografia

- [1] W. Ahmad, H. Hacihabiboglu e 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 e J. Hernando. *Xbic: Real-time cross probabilities measure for speaker segmentation*. Paper. Univ. California, Berkeley: ICSI (International Computer Science Institute), 2005. URL: http://www.xavieranguera.com/papers/techreport_xbic.pdf (visitato il 01/12/2015).
- [3] Daniel Arfib. «Gestural strategies for specific filtering processes». In: *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_strategies.pdf.
- [4] M. Athineos e D. P. W. Ellis. «Sound texture modelling with linear prediction in both time and frequency domains». In: *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on 5* (2003), V-648-51 vol.5.
- [5] C. Bascou. «Modélisation de sons bruités par la Synthèse Granulaire». Rapport de stage de DEA ATIAM. Université Aix-Marseille II, 2004.
- [6] C. Bascou e L. Pottier. «GMU, A Flexible Granular Synthesis Environment in Max/M-SP». In: *Sound and Music Computing Conference*. Citeseer. 2005.
- [7] C. Bascou e L. Pottier. «New sound decomposition method applied to Granular Synthesis». In: *Proc. ICMC*. Barcelona, Spain, 2005.
- [8] D. Birchfield, N. Mattar e H. Sundaram. «Design of a generative model for soundscape creation». In: *International Computer Music Conference, Barcelona, Spain*. Citeseer. 2005.
- [9] D. Birchfield et al. «Generative Soundscapes for Experiential Communication». In: *Society for Electro Acoustic Music in the United States, Muncie, IN* (2005).
- [10] P. Cano et al. «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*. Rapp. tecn. University of Cambridge, UK: Computer Laboratory, mag. 2004. URL: <http://www.cl.cam.ac.uk/users/mpc33/Cardle-Sound-Synthesis-techreport-2004-low-quality.pdf>.
- [12] P. R. Cook. «Din of an “iquity”: Analysis and synthesis of environmental sounds». In: *International Conference on Auditory Display (ICAD2007)*. 2007, pp. 167–172.
- [13] Perry R. Cook. «Physically informed sonic modeling (PhISM): Percussive synthesis». In: *Proceedings of the International Computer Music Conference (ICMC)* (1996). URL: <http://quod.lib.umich.edu/cgi/p/pod/dod-idx/physically-informed-sonic-modeling-phism-percussive.pdf?c=icmc;idno=bbp2372.1996.071>.
- [14] Perry R. Cook. «Physically informed sonic modeling (phism): Synthesis of percussive sounds». In: *Computer Music Journal* (1997). URL: <http://www.jstor.org/stable/3681012>.
- [15] Richard Corbett et al. «TimbreFields: 3D Interactive Sound Models for Real-Time Audio». In: *Presence: Teleoperators and Virtual Environments* 16.6 (2007), pp. 643–654. eprint: <http://www.mitpressjournals.org/doi/pdf/10.1162/pres.16.6.643>. URL: <http://www.mitpressjournals.org/doi/abs/10.1162/pres.16.6.643>.
- [16] M. Desainte-Catherine e P. Hanna. «Statistical Approach FOR Sound Modeling». In: *Digital Audio Effects (DAFx)*. Citeseer, 2000.
- [17] A. Di Scipio. «Synthesis of environmental sound textures by iterated nonlinear functions». In: *Digital Audio Effects (DAFx)*. 1999.
- [18] Yoshinori Dobashi, Tsuyoshi Yamamoto e Tomoyuki Nishita. «Real-time rendering of aerodynamic sound using sound textures based on computational fluid dynamics». In: *ACM Trans. Graph.* 22 (3 2003), pp. 732–740. URL: <http://doi.acm.org/10.1145/882262.882339>.
- [19] Yoshinori Dobashi, Tsuyoshi Yamamoto e Tomoyuki Nishita. «Synthesizing sound from turbulent field using sound textures for interactive fluid simulation». In: *Proc. of Eurographics*. 2004, pp. 539–546.
- [20] Kees van den Doel, Paul G. Kry e 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. New York, NY, USA: ACM, 2001, pp. 537–544. URL: <http://doi.acm.org/10.1145/383259.383322>.
- [21] Kees van den Doel. «Physically based models for liquid sounds». In: *ACM Trans. Appl. Percept.* 2 (4 2005), pp. 534–546. URL: <http://doi.acm.org/10.1145/1101530.1101554>.
- [22] S. Dubnov e N. Tishby. «Analysis of sound textures in musical and machine sounds by means of higher order statistical features». In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 5. IEEE, 1997, pp. 3845–3848. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=604726.

- [23] Shlomo Dubnov et al. «Synthesis of Audio sound Textures by Learning and Resampling of Wavelet Trees». In: *IEEE Computer Graphics and Applications* 22.4 (2002), pp. 38–48.
- [24] A. A. Efros e T. K. Leung. «Texture Synthesis by Non-parametric Sampling». In: *International Conference on Computer Vision*. Vol. 2. Citeseer. 1999, p. 1033.
- [25] Aaron Einbond, Diemo Schwarz e 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.
- [26] R. El-Yaniv, S. Fine e N. Tishby. «Agnostic classification of Markovian sequences». In: *Proceedings of the 1997 conference on Advances in neural information processing systems 10*. MIT Press. 1998, pp. 465–471.
- [27] Z. B. J. R. El-Yaniv, D. L. M. Werman e S. Dubnov. «Granular Synthesis of Sound Textures using Statistical Learning». In: *Proc. ICMC*. 1999.
- [28] Daniel P. W. Ellis, Xiaohong Zeng e Josh H. McDermott. «Classifying Soundtracks with Audio Texture Features». In: *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>.
- [29] G. Essl e S. O'Modhrain. «Scrubber: an interface for friction-induced sounds». In: *2005 conference on New interfaces for musical expression*. National University of Singapore. 2005, p. 75.
- [30] G. Fan e X. G. Xia. «Wavelet-based texture analysis and synthesis using hidden Markov models». In: *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 50.1 (2003).
- [31] Andy Farnell. *Designing Sound*. MIT Press, ott. 2010. URL: <http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=12282>.
- [32] J. J. Filatriau e D. Arfib. «Instrumental gestures and sonic textures». In: *International Conference on Sound and Music Computing (SMC)*. 2005.
- [33] J. J. Filatriau, D. Arfib e J. M. Couturier. «Using visual textures for sonic textures production and control». In: *Digital Audio Effects (DAFx)*. 2006.
- [34] N. Finney. «Autonomous generation of soundscapes using unstructured sound databases». MSc Thesis. UPF, 2009. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.168.4139&rep=rep1&type=pdf>.
- [35] N. Finney e 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=pdf>, <http://www.w3.org/2010/06/w3car/>.
- [36] M. Fröjd e A. Horner. «Sound texture synthesis using an overlap-add/granular synthesis approach». In: *Journal of the Audio Engineering Society* 57.1/2 (2009), pp. 29–37. URL: <http://www.aes.org/e-lib/browse.cfm?elib=14805>.

- [37] William W. Gaver. «How do we hear in the world? Explorations in ecological acoustics». In: *Ecological psychology* (1993). URL: <http://www.tandfonline.com/doi/abs/10.1207/s15326969eco0504_2>, <http://www.cog.brown.edu/courses/cg195/pdf_files/fall07/Gaver-howdowehear.pdf>.
- [38] T. Grill. «Re-texturing the sonic environment». In: *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*. ACM, 2010, pp. 1–7. URL: <<http://portal.acm.org/citation.cfm?id=1859805>>, <<http://grrrr.org/research/publications/paper-am-2010/>>.
- [39] S. Guidati e Head Acoustics GmbH. «Auralisation and psychoacoustic evaluation of traffic noise scenarios». In: *Journal of the Acoustical Society of America* 123.5 (2008), p. 3027.
- [40] P. Hanna. «Statistical modelling of noisy sounds : spectral density, analysis, musical transformations and synthesis». Tesi di dott. Laboratoire de Recherche en Informatique de Bordeaux (LaBRI), 2003. URL: <http://dept-info.labri.fr/~hanna/phd.html>.
- [41] Pierre Hanna e Myriam Desainte-Catherine. «A Statistical and Spectral Model for Representing Noisy Sounds with Short-Time Sinusoids». In: *EURASIP Journal on Advances in Signal Processing* 2005.12 (2005), pp. 1794–1806. URL: <http://www.hindawi.com/journals/asp/2005/182056.abs.html>.
- [42] Pierre Hanna et al. «Audio features for noisy sound segmentation». In: *Proceedings of the 5th International Symposium on Music Information Retrieval (ISMIR'04)*. 2004, pp. 120–123. URL: <http://www.ee.columbia.edu/~dpwe/ismir2004/CRFILES/paper184.pdf>.
- [43] R. Hoskinson. «Manipulation and Resynthesis of Environmental Sounds with Natural Wavelet Grains». Tesi di dott. The University of British Columbia, 2002. URL: https://www.cs.ubc.ca/grads/resources/thesis/May02/Reynald_Hoskinson.pdf.
- [44] Reynald Hoskinson e Dinesh Pai. «Manipulation and Resynthesis with Natural Grains». In: *International Computer Music Conference (ICMC)*. Havana, Cuba, set. 2001, pp. 338–341.
- [45] Ismo Kauppinen e Kari Roth. «An Adaptive Technique for Modeling Audio Signals». In: *Digital Audio Effects (DAFx)*. Limerick, Ireland, dic. 2001. URL: <http://www.csis.ul.ie/dafx01/proceedings/papers/kauppinen.pdf>.
- [46] Stefan Kersten e Hendrik Purwins. «Sound Texture Synthesis with Hidden Markov Tree Models in the Wavelet Domain». In: *International Conference on Sound and Music Computing (SMC)*. Barcelona, Spain, lug. 2010.
- [47] Anil Kokaram e Deirdre O'Regan. «Wavelet Based High Resolution Sound Texture Synthesis». In: *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*. Giu. 2007. URL: <http://www.aes.org/e-lib/browse.cfm?elib=13952>.
- [48] M. Lagrange et al. «Objective quality measurement of the excitation of impact sounds in a source/filter model». In: *Acoustical Society of America Journal* 123 (2008), p. 3746.

- [49] Jung Suk Lee, Philippe Depalle e 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, set. 2010.
- [50] L. Lu, L. Wenyin e H.J. Zhang. «Audio textures: Theory and applications». In: *IEEE Transactions on Speech and Audio Processing* 12.2 (2004), pp. 156–167. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1284343.
- [51] L. Lu et al. «Audio textures». In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Vol. 2. 2002. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.4586&rep=rep1&type=pdf>.
- [52] A. Mallock. «Sounds Produced by Drops Falling on Water». In: *Proceedings of the Royal Society of London* 95.667 (1918), pp. 138–143. URL: <https://ia600502.us.archive.org/15/items/philtrans00768948/00768948.pdf>.
- [53] Aymeric Masurelle. «Gestural control of environmental texture synthesis». Rapport de stage de DEA ATIAM. Ircam–Centre Pompidou, Université Paris VI, 2011.
- [54] J. H. McDermott, A. J. Oxenham e 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, 2009.
- [55] Dylan Menzies. «Phya and vfoley, physically motivated audio for virtual environments». In: *Proceedings of the Audio Engineering Society Conference* (2010). URL: <http://www.aes.org/e-lib/browse.cfm?elib=15171>.
- [56] Dylan Menzies. «Physically Motivated Environmental Sound Synthesis for Virtual Worlds». In: *EURASIP Journal on Audio, Speech, and Music Processing* (2011). URL: <http://www.asmp.eurasipjournals.com/content/2010/1/137878>.
- [57] Nadine E. Miner e Thomas P. Caudell. «Using wavelets to synthesize stochastic-based sounds for immersive virtual environments». In: *ACM Trans. Appl. Percept.* 2 (4 2005), pp. 521–528. URL: <http://doi.acm.org/10.1145/1101530.1101552>.
- [58] Nicolas Misdariis et al. «Environmental Sound Perception: Metadescription and Modeling Based on Independent Primary Studies». In: *EURASIP Journal on Audio, Speech, and Music Processing* (2010). URL: <http://articles.ircam.fr/textes/Misdariis10b/>.
- [59] A. Misra e P. R. Cook. «Toward synthesized environments: A survey of analysis and synthesis methods for sound designers and composers». In: *Proc. ICMC*. 2009.
- [60] A. Misra, P. R. Cook e G. Wang. «A New Paradigm for Sound Design». In: *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*. 2006.
- [61] A. Misra, G. Wang e Perry Cook. «Musical Tapestry: Re-composing Natural Sounds». In: *Journal of New Music Research* 36.4 (2007), pp. 241–250.
- [62] Ananya Misra, Perry R. Cook e Ge Wang. «Tapestrea: Sound scene modeling by example». In: *ACM SIGGRAPH 2006 Sketches*. ACM. 2006, p. 177. URL: <http://dl.acm.org/citation.cfm?id=1180070>, http://soundlab.cs.princeton.edu/publications/taps_sketch.pdf.

- [63] William Moss et al. «Sounding Liquids: Automatic Sound Synthesis from Fluid Simulation». In: *ACM Transactions on Graphics* 28.4 (ago. 2009), pp. 1–12. URL: <http://portal.acm.org/citation.cfm?doid=1559755.1559763>.
- [64] E. Murphy et al. «Perceptual Evaluation of a Real-time Synthesis Technique for Rolling Sounds». In: *Conference on Enactive Interfaces*. Pisa, Italy, 2008.
- [65] J. F. O’Brien, C. Shen e C. M. Gatchalian. «Synthesizing sounds from rigid-body simulations». In: *2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*. ACM New York, NY, USA, 2002, pp. 175–181.
- [66] S. O’Modhrain e G. Essl. «PebbleBox and CrumbleBag: tactile interfaces for granular synthesis». In: *2004 conference on New interfaces for musical expression*. National University of Singapore, 2004, p. 79.
- [67] D. O’Regan e 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>.
- [68] J. R. Parker e B. Behm. «Creating audio textures by example: tiling and stitching». In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* 4 (2004), iv–317–iv–320 vol.4.
- [69] Geoffroy Peeters. *A large set of audio features for sound description (similarity and classification) in the Cuidado project*. Rapp. tecn. version 1.0. Paris, France: Ircam – Centre Pompidou, apr. 2004. URL: http://www.ircam.fr/anasy/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf.
- [70] Leevi Peltola et al. «Synthesis of hand clapping sounds». In: *Audio, Speech, and Language Processing, IEEE Transactions on* 15.3 (mar. 2007), pp. 1021–1029. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4100694.
- [71] Gabriel Peyré. *Oriented Patterns Synthesis*. 2007.
- [72] Cecile Picard, Nicolas Tsingos e François Faure. «Retargetting Example Sounds to Interactive Physics-Driven Animations». Anglais. In: *n AES 35th International Conference, Audio in Games*. London Royaume-Uni, 2009. URL: <http://hal.inria.fr/inria-00394469/en/>.
- [73] K. Popat e R. W. Picard. «Cluster-based probability model and its application to image and texture processing.» In: *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* 6.2 (gen. 1997), pp. 268–84. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=551697.
- [74] Xavier Rodet e Phillipe Depalle. «A new additive synthesis method using inverse Fourier transform and spectral envelopes». In: *International Computer Music Conference (ICMC)*. Ott. 1992.
- [75] Nicolas Saint-Arnaud. «Classification of Sound Textures». Tesi di dott. Universite Laval, Quebec, MIT, 1995.
- [76] Nicolas Saint-Arnaud e Kris Popat. «Analysis and synthesis of sound textures». In: *Readings in Computational Auditory Scene Analysis*. 1995, pp. 125–131.

- [77] R. Murray Schafer. *The Soundscape*. Destiny Books, 1993, p. 320. URL: <http://www.amazon.com/Soundscape-R-Murray-Schafer/dp/0892814551>.
- [78] Diemo Schwarz. «Concatenative Sound Synthesis: The Early Years». In: *Journal of New Music Research* 35.1 (mar. 2006). A cura di Adam T. Lindsay. Special Issue on Audio Mosaicing, pp. 3–22.
- [79] Diemo Schwarz. «Corpus-Based Concatenative Synthesis. Assembling sounds by content-based selection of units from large sound databases». In: *IEEE Signal Processing Magazine* 24.2 (mar. 2007). A cura di Shih-Fu Chang. Special Section: Signal Processing for Sound Synthesis, pp. 92–104.
- [80] Diemo Schwarz. «Data-Driven Concatenative Sound Synthesis». Thèse de doctorat. Paris: Université Paris 6 – Pierre et Marie Curie, 2004. URL: <http://mediatheque.ircam.fr/articles/textes/Schwarz04a>.
- [81] Diemo Schwarz. «State of the Art in Sound Texture Synthesis». In: *Proc. of the 14th International Conference on Digital Audio Effects (DAFx-11)*. Paris, set. 2011, pp. 221–231. URL: http://recherche.ircam.fr/pub/dafx11/Papers/30_e.pdf (visitato il 23/11/2015).
- [82] Diemo Schwarz, Roland Cahen e Sam Britton. «Principles and Applications of Interactive Corpus-Based Concatenative Synthesis». In: *Journées d'Informatique Musicale (JIM)*. GMEA, Albi, France, mar. 2008. URL: <http://mediatheque.ircam.fr/articles/textes/Schwarz08a/index.pdf>.
- [83] Diemo Schwarz e Norbert Schnell. «A Modular Sound Descriptor Analysis Framework for Relaxed-Real-Time Applications». In: *Proc. ICMC*. New York, NY, 2010.
- [84] Diemo Schwarz e Norbert Schnell. «Descriptor-based Sound Texture Sampling». In: *International Conference on Sound and Music Computing (SMC)*. Barcelona, Spain, lug. 2010, pp. 510–515. URL: <http://articles.ircam.fr/textes/Schwarz10a/index.pdf>.
- [85] Diemo Schwarz et al. «Real-Time Corpus-Based Concatenative Synthesis with CataRT». In: *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*. Montreal, Canada, set. 2006, pp. 279–282.
- [86] G. Strobl. «Parametric Sound Texture Generator». MSc Thesis. Universität für Musik und darstellende Kunst, Graz; Technische Universität Graz, 2007. URL: <http://en.scientificcommons.org/43580321>.
- [87] G. Strobl et al. «Sound texture modeling: A survey». In: *Sound and Music Computing Conference*. 2006.
- [88] Patrick Susini et al. «Characterizing the sound quality of air-conditioning noise». In: *Applied Acoustics* 65-8 (2004), pp. 763–790. URL: <http://articles.ircam.fr/textes/Susini04b/>.
- [89] N. Tsingos, E. Gallo e G. Drettakis. «Perceptual audio rendering of complex virtual environments». In: *ACM SIGGRAPH 2004 Papers*. ACM, 2004, pp. 249–258.
- [90] A. Valle, V. Lombardo e M. Schirosa. «Simulating the Soundscape through an Analysis/Resynthesis Methodology». In: *Auditory Display* (2010), pp. 330–357.

- [91] Andrea Valle, Vincenzo Lombardo e Mattia Schirosa. «A Graph-based System for the Dynamic Generation of Soundscapes». In: *15th International Conference on Auditory Display*. Copenhagen, 2009, pp. 217–224.
- [92] C. Verron et al. «Spatialized Synthesis of Noisy Environmental Sounds». In: *Auditory Display*. Springer-Verlag, 2010, pp. 392–407. URL: <http://www.springerlink.com/index/J3T5177W11376R84.pdf>.
- [93] Charles Verron. «Synthèse immersive de sons d’environnement». PhD Thesis. Université Aix-Marseille I, 2010.
- [94] Charles Verron et al. «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>>, <<https://hal.archives-ouvertes.fr/hal-00546833>>.
- [95] X. Zhu e L. Wyse. «Sound texture modeling and time-frequency LPC». In: *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx)*. Vol. 4. 2004.