INTERNATIONAL JOURNAL OF FILM AND MEDIA ARTS (2023) Vol. 8, Nº. 3 pp. 89-107

ijfma.ulusofona.pt
doi: 10.24140/ijfma.v8.n3.06

SONIC ALCHEMIST: AI'S ROLE IN AMPLIFYING CREATIVITY IN FILM SOUND DESIGN

VYTIS PURONAS
LITHUANIAN ACADEMY OF MUSIC AND THEATRE (LITHUANIA)

Vytis Puronas is a film sound designer, re-recording mixer, and composer. Associate Professor at the Lithuanian Academy of Music and Theatre, he has taught and presented in Australia, the United States, Singapore, and the UK. Creator of 'Sonic Alchemist' software, his extensive filmography has earned him three Silver Crane nominations.

Corresponding Author
Vytis Puronas
vytis.puronas@lmta.lt
Lithuanian Academy of Music and Theatre
Gedimino pr. 42
01110 Vilnius
Lithuania

Schedule for publication
Published online: 29th December 2023

Abstract

Defying mainstream views of AI as tools that mimic or replace human input, this paper presents Sonic Alchemist: an avant-garde AI software taking a fresh approach to sound design. Built on the Beginner's Mind principles, it empowers users to defy familiar paths and habits, fostering collaboration between human and machine. Unique features like automatic sound effects matching and a weight-based sound selection system further enhance this partnership. User feedback and case studies highlight Sonic Alchemist's capacity to evoke new insights and unexpected solutions, reinforcing its role as a creative ally. Analysis of industry reception and user responses during testing underscores its potential to transform sound design in visual media. This study underlines the enduring significance of human originality, even in the AI era, portraying Sonic Alchemist as an inspiration tool and a champion of human creativity. It asserts that the willingness to unlearn and rediscover is key in creative innovation, even when assisted by advanced AI tools.

Keywords: Sound Design, Artificial Intelligence, Creativity, Chance Operations, Unpredictability, Post-Production, Beginner's Mind

> 'Accidents and mistakes are probably the main "technique" that drives innovation in every field. The trick is being able to recognize the potential benefits from events that are completely unanticipated.'
>
> Randy Thom (Designing Sound, 2011)

Introduction

A shared thread among those passionate about their craft is an obsession that fuels an unending quest for perfection. This fixation drives us, morphing us from novices, mere wannabes, into seasoned know-it-alls. Over time, we become experts. Yet, this continuous strive for perfection, while seemingly a path to success, can inadvertently lead us into an unsuspecting trap — the comfort zone.

Within this comfort zone, tasks such as cutting dialogue, selecting sound effects, performing foley, and mixing become automatic, almost second nature. Despite each project bringing its own set of rules and expectations, some aspects of our work and certain actions, even our creative reactions, begin to seem predictable and repeatable.

We start to recognize certain pitfalls, including distractions caused by our digital audio workstations, familiar paths and presets, deep menu-diving applications that are both a blessing and a curse in our daily lives as sound designers. We see ourselves taking familiar routes from point A to point B, repeating the same keywords on every project to search our ever-expanding sound libraries.

We react to a vehicle in the scene by searching for a "pass by" sound effect. We react to a character falling by typing in "body fall". Some of our routine practices involve templates with enabled high pass and low pass filters. The tried-and-true reverb plugins with familiar presets become our go-to resources.

The words of Shunryu Suzuki, a renowned Sōtō Zen monk and teacher, resonate with this narrative. A statement often quoted at the School of Sound symposium in London, he said: "In the beginner's mind there are many possibilities; in the expert's there are few".

Patterns and comfort zones are inherently interconnected. As individuals settle into their comfort zones, they may develop a reliance on patterns, which can subsequently hinder the pursuit of innovation. In professional environments, the reinforcement of patterns may be further exacerbated by a system that rewards consistency and adherence to established practices, leading to risk aversion and entrenched views.

This work is aimed at offering a fresh perspective on sound design practices and a roadmap to innovative thinking, with the intent of encouraging sound designers to leap out of their comfort zones and rediscover the beginner's mind, where countless possibilities await. By doing so, this paper hopes to reframe the narrative around expertise, reminding us that while mastery is valuable, it should not lead to creative stagnation. In this light, AI is reimagined as a partner in creativity rather than a threat, serving to elevate rather than replace human involvement in the field of sound design.

Chapter 1: The Power of Randomness in Art and Sound Design

The Dance of Chance

Chance operations have a long-standing history in the realm of artistic creativity, particularly in poetry. By relying on random processes, such as the rolling of dice or computer algorithms, chance operations allow artists to explore new creative possibilities beyond their conscious intentions and ego – concepts prominently embraced by the Dadaist movement. One of the key figures of the Dada movement, Tristan Tzara, articulated this concept in his "Dada Manifesto on Feeble & Bitter Love" (1920). In this provocative text, Tzara presented a method for composing Dadaist poetry by embracing randomness as a powerful muse.

> TO MAKE A DADAIST POEM
> Take a newspaper.
> Take some scissors.
> Choose from this paper an article of the length you want to make your poem.
> Cut out the article.
> Next carefully cut out each of the words that makes up this article and put them all in a bag.
> Shake gently.
> Next take out each cutting one after the other.
> Copy conscientiously in the order in which they left the bag.
> The poem will resemble you.
> And there you are – an infinitely original author of charming sensibility, even though unappreciated by the vulgar herd.

In contemporary art, the use of chance operations has been popularized by the international avant-garde group Fluxus and poet-composer John Cage, amongst others. Their works highlight the potential for chance operations to generate innovative and thought-provoking creative outputs.

Janet Sobel, an often overlooked but significant figure in the abstract expressionist movement, played a pioneering role in the exploration of chance and spontaneity in her art. She developed her distinctive "drip painting" technique, which involved allowing paint to drip, flow, and splash across the canvas. Sobel's approach was one of embracing the unpredictable, as she relinquished direct control over the paint's movement. This resulted in abstract compositions marked by a sense of dynamic energy and unanticipated patterns. Sobel's work embodied a connection to nature's forces and a celebration of the unexpected interplay between intention and accident.

Jackson Pollock further refined and popularized the "drip painting" technique. His method involved placing the canvas on the floor and dripping, pouring, and flinging paint from above, enabling him to work in a state of flow and instinct. By distancing himself from the canvas and using unconventional tools like sticks and brushes, Pollock unleashed the power of chance and gravity. The paint seemed to dance across the canvas, resulting in intricate webs of colour and texture. His works, like Sobel's, exemplify the fusion of intention and randomness, revealing an unparalleled expression of raw emotion and vitality.

Human cognition possesses an innate ability to perceive patterns in seeming chaos, much like identifying shapes in cloud formations or discerning an image in the swirls of a coffee cup. This speaks to our capacity for gestalt thinking, a psychological phenomenon where we naturally organize our perceptions into coherent wholes, spotting relationships and meaning where none might objectively exist (Wertheimer, M., 1923).

The Role of Pattern-breaking

In the field of artistic creation, the drive for innovation often mandates a divergence from established norms and the pursuit of fresh pathways. This chapter delves into the role of pattern-breaking in cultivating creativity and fostering innovation within the sphere of sound design, scrutinizing the challenges that stem from entrenched behaviours, and analyzing strategies that aid in carving out new solutions.

Creativity is intrinsically linked to our ability to challenge and reconfigure self-imposed boundaries (Eagleman, D. & Brandt, A., 2017). Within sound design, patterns can often manifest as recurring tendencies and familiar strategies, potentially serving as obstacles to the creative process. Therefore, the practice of pattern-breaking emerges as critical, given its power to unveil new prospects and stimulate progressive thinking in sound design.

Patterns and comfort zones, often subconsciously woven into the fabric of our working methods, can act as silent inhibitors of creativity. Sound designers, while honing their craft, may inadvertently slip into a repetition of tried-and-tested methods, gradually establishing a familiar routine. The danger lies not in the efficiency of the routine but in the blind spots it creates, subtly stiffening the impulse for exploration and innovation. Recognizing these habits and the comfort zones they create is the first step towards pattern-breaking.

Escaping from these self-constructed confines requires deliberate effort and strategic planning. Some of the potent strategies that can be employed within the field of sound design to break patterns include:

Pattern-breaking commences with the identification and introspection of one's own habits and the methodology that has morphed into the standard approach. This conscious process of self-awareness can offer crucial insights into potential areas ripe for transformation.

Deliberately diverging from familiar techniques and embracing novel methods allows sound designers to challenge their entrenched patterns and widen their creative scope.

In the intricate journey of crafting sounds, it is imperative not to settle for obvious solutions. Drawing from Sawyer's (2006) work, it is proposed that the first few ideas that come to mind should be purposefully set aside, clearing the path for less conventional but possibly more inventive solutions. This approach can assist sound designers in dodging imitation, thereby facilitating truly original and innovative outcomes.

The act of pattern-breaking is, in essence, a call to disrupt comfort zones and embrace the uncertainty of uncharted territories. For sound designers, it represents an exciting opportunity to infuse their work with a renewed sense of creativity and originality.

Card-Based Creativity

Over the years, researchers and practitioners have employed numerous tools and techniques designed to stimulate creative thinking, sparking fresh ideas and solutions. These methods span across the spectrum from random lyric generators and modular synthesizers to paper-based and software options. This chapter examines the development and application of a particular type of creative tool—a card-based system created to overcome creative hurdles and inspire innovative solutions.

A pioneering card-based tool designed to stimulate creativity is a deck developed by Brian Eno and Peter Schmidt in the 1970s. Oblique Strategies epitomizes the power of randomness and ambiguity in stimulating creative thinking. The deck comprises "over one hundred worthwhile dilemmas," each card prompting the user to think about their creative problem from a different angle. A user selects a random card from the deck and follows its instructions, simulating the experience of obtaining advice from a creative collaborator (Eno & Schmidt, 1975). The instructions can be straightforward, metaphorical, or even seemingly irrelevant.

There are several guiding principles behind card-based creative tools like Oblique Strategies:

Provocation: The cards are meant to provoke thought, pushing the user out of their comfort zone and challenging them to think in new and unconventional ways.

Randomness: The process of drawing a card introduces an element of randomness that can help to disrupt routine thinking patterns and stimulate fresh ideas.

Flexibility: These tools are not prescriptive. They are open to interpretation, allowing the user to apply them in ways that make sense for their specific creative dilemma.

Simplicity: The beauty of these tools lies in their simplicity. They are easy to use and require no special skills or knowledge, making them accessible to anyone seeking to stimulate their creativity.

It's the unpredictability and thought-provoking nature of these prompts that makes them valuable. They disrupt linear thinking, challenge assumptions, and stimulate lateral thinking.

The Accidental Synchronicity Technique in Sound Design

One technique that can foster innovation in sound design involves the use of random sound effects in conjunction with a film sequence. This is achieved by selecting random sound effects and freely synchronising them with the visual material – not necessarily mimicking the timing of on-screen action, thereby disrupting the monotony of familiar and predictable patterns.

The term 'random', in this context, signifies an open selection devoid of any preconceived relevance to the scene at hand. Whether the sound effect is as seemingly irrelevant as a "cat's purr" during a car chase, or a "ship horn" during a fight scene, the key is to resist dismissing any sound as irrelevant prematurely.

The placement of these sounds on the timeline can vary. They can overlap, or they can be isolated. The essential aspect here is to delay making judgments and decisions until later.

The next step of this experimental process involves careful listening and critical evaluation. The aim here is to identify any accidentally synchronous sounds that could potentially add a unique flavour to the scene. Amusingly, from an assortment of about 20 random sounds, typically 2 or 3 will 'click' into place — either aligning perfectly with the scene or fitting elsewhere within the sequence.

Now, attention turns to the sounds that did not initially appeal. This stage allows for creative exploration, as various processing techniques can be applied to these sound effects. These sounds can be stretched, reduced to a whisper, or amplified beyond recognition, thereby transforming their initial impact.

The outcome of this technique is a set of 'anchor' sounds, providing a unique sonic foundation from which to construct the rest of the sequence. This approach encourages a break away from the confines of the comfort zone, fostering a sound design process that is fresh, exciting, and innovative. Randomness and exploration in sound design can help rekindle the playful, experimental nature of the craft, freeing practitioners from the constraints of past successes and habitual tendencies.

This methodology has proven to be remarkably effective in educational settings and has been explored in film schools such as the Lithuanian Academy of Music and Theatre in Vilnius and Griffith Film School in Brisbane. It led to surprising, and delightful, outcomes. It is crucial to highlight that during these explorative exercises, the technique did not eclipse the individual aesthetics, tastes, or characters of the participants. On the contrary, it provided a unique platform that enabled them to express and demonstrate their highly distinct stylistic choices. This occurred even when all sound designers were working with the same restricted set of sounds.

Chapter 2: Coding the Art of Sound Design

The Human-Driven Sound Design Process

For the purposes of this investigation, the sound design process has been distilled into four primary stages: Spotting, Generating, Assembling, and Evaluating. Each stage represents a distinct aspect of the sound design process, from the initial identification of objects and movement in the picture and story to the final evaluation and validation of creative decisions.

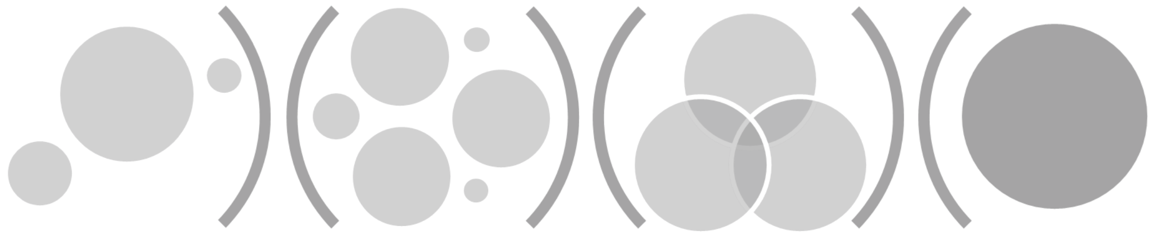

Fig. 1 The Four-Stage Human-Driven Sound Design Process: Spot, Generate, Assemble and Evaluate.

Stage 1. Spotting: Identification of Objects and Movement

The Spotting stage involves the early analysis of a script and a spotting session with the director at the beginning of the post-production process or, ideally, participation prior to production. This stage requires reading the story, identifying characters, their roles, and ultimate goals, as well as tracking their trajectories within the narrative. The purpose of this stage is to discern the system in which these characters function, including when they appear and disappear, and the reasons behind these events.

Stage 2. Generating: Gathering Sound Effects and Defining Language

In the Generating stage, a unique sonic language tailored to the specific requirements and expectations of the story is developed. This involves establishing a system with its own internal logic, gathering sound effects through field recordings, synthesis, sampling, or extraction from sound libraries. As the sonic language evolves, its grammar and rules are refined to suit the narrative. Each story demands a distinct sonic vocabulary, necessitating a unique approach for every project.

Stage 3. Assembling: Layering and Adjusting Sounds

During the Assembling stage, a cohesive sequence that aligns with the picture is produced by layering, adjusting, and blending sounds. This stage enables a deeper understanding of the relationships between sonic elements, the audio-visual relationship, and the various characteristics of the sound, such as synchronicity, weight, size, and depth. The diegetic and non-diegetic, background and foreground elements of the sonic system are established.

Stage 4. Evaluating: Interpretation and Validation of Creative Decisions

The Evaluating stage involves nuanced decision-making, a process that draws heavily upon a sound designer's personal aesthetics, creative intuition, and subjective judgement - elements that are inherently human and cannot be fully replicated by an algorithm. The sonic language has now taken shape, with the alphabet forming words and sentences that contribute to the overall narrative. This stage, often considered the most enjoyable part for the artist, incorporates not just technical expertise but the human ability to empathize, intuit, and create in a way that resonates with other human experiences. The ability to switch roles between the creator and the audience allows sound designers to validate their creative decisions and understand the significance of the sound story.

Considering this, it is indeed beneficial to revise the algorithmic approach to sound design, reducing it to three stages: Analysis, Production, and Assembly. These stages involve the identification of key elements, generation of sound assets, and creation of the sonic structure - tasks that an algorithm can handle effectively and efficiently.

The final stage, Evaluating, should then be handled by the human sound designer. Here, they can listen to the assembled sound design, evaluate the creative choices made by the algorithm, and apply their unique artistic perspective. They can approve elements that meet the creative goals, modify those that need adjustments, or even replace parts entirely.

This approach allows the algorithm to do the heavy lifting of sound identification, production, and assembly, while the artist is left with the creative and rewarding task of shaping and refining the sound to meet the emotional and narrative needs of the project. In doing so, we create a symbiotic relationship between technology and creativity, where each plays to their strengths, enhancing the overall quality and efficiency of the sound design process.

A Three-Stage Algorithmic Approach to Sound Design

Stage 1: Analysis. Translating Visual Cues into Sound Maps

In the first stage of the algorithmic process, the artificial intelligence (AI) system deploys techniques such as computer vision and machine learning to analyse the imported video footage. The goal is to identify significant visual cues and actions, from dramatic scene shifts to subtle character movements, which will require synchronizing sound effects. To provide context and refine the spotting of potential sync points, Natural Language Processing (NLP) can also be used to parse scene descriptions or relevant keywords.

Stage 2: Production. Crafting the Sonic Palette

Armed with the sync points identified, the AI algorithm proceeds to generate the corresponding sound effects. It employs a two-pronged approach: sourcing sound samples from an existing library and synthesizing new sounds when required. Machine learning models categorize the sound library's contents based on their sonic characteristics, enabling quick and efficient matching with the identified sync points. Furthermore, the AI system can manipulate these existing sounds—altering, combining, or morphing them—to yield unique audio effects, a process akin to 'breeding' new species of sonic materials. The AI also can adjust parameters such as volume, pitch, and duration to create sound effects that perfectly complement the corresponding visual cues.

Stage 3: Assembly. Sound Integration and Action Synchronization

Having generated the necessary sound elements, the algorithm proceeds to arrange them into a cohesive and synchronized sequence that aligns seamlessly with the visual narrative. During this stage, the algorithm also employs audio transformations—like pitch shifting, equalization, and reverb—to enhance the integration of sound and visuals. Advanced techniques such as clustering algorithms are utilized to group sounds, forming layers or sequences that result in a multitrack arrangement.

The Balance of Predictability and Surprise in Algorithmic Sound Design

The algorithmic approach to sound design provides a unique blend of efficiency and creativity. By leveraging AI's capabilities for the more time-consuming tasks of identification, production, and assembly of sound elements, the human designer is left with the crucial task of evaluation and refinement. However, the end goal is not just about precision. The process encourages exploration and allows for unexpected results, ensuring that the sound design remains fresh and thought-provoking. Thus, through this blend of AI and human creativity, we manage to maintain a dynamic balance between predictability and surprise, familiarity and novelty in sound design.

Chapter 3: The Development Trajectory of Sonic Alchemist

Python-based Prototype and Sound Design Methodology

The initial prototype of the software, called Sonic Alchemist, version 0.1 was presented at the Sound and Screen CAPA Conference at Nanyang Technological University Singapore in May 2022 and showed promising results. The early Python-based prototype utilized a multi-step algorithm to automate sound design tasks. The algorithm encompassed several variations, allowing for the placement of sound effects in a random or action-synchronized manner. It also facilitated the selection of appropriate sounds based on predefined categories, such as vehicles, whooshes, and metal rattles. Additionally, the algorithm could randomize pitch and speed adjustments to further enhance the sound design.

The Python-based prototype of Sonic Alchemist v0.1 was trained on a general sound effects library, containing approximately 30,000 entries. Each entry was catalogued by various parameters, such as duration, sync points (moments of high audio energy), and associated metadata that included keywords and descriptions. The action synchronisation component of Sonic Alchemist v0.1 was in its early stages, relying predominantly on human guidance and marker extraction for optimal results.

The workflow of the early prototype involved placing markers in a Digital Audio Workstation (DAW), such as Steinberg Nuendo, to specify synchronization moments. These markers were then exported as a CSV file and loaded into the prototype software. Sonic Alchemist v0.1 employed several algorithms, with the primary one capable of matching video events by randomly selecting sounds from the database and synchronizing them with video markers.

Upon executing the algorithm, Sonic Alchemist v0.1 generated a Steinberg XML file that could be directly loaded into Nuendo alongside its corresponding audio files. The resulting synchronized audio was then automatically placed on the video timeline and allowed the artist to explore the basic principles of accidental synchronicity and seek inspiration in unanticipated events. Sonic Alchemist demonstrated its effectiveness in synchronizing sound effects for various animated sequences involving racing vehicles, mechanical contraptions, and fast-paced elements.

Despite Sonic Alchemist's initial prototype proving effective in automating sound design tasks, it did exhibit several limitations. The system's reliance on manual input and control rendered it less streamlined than desired. Its sound selection algorithm was quite rudimentary, and heavily reliant on the presence of accurate metadata. On the technical front, the system was CPU-intensive, causing basic tasks to require extended processing time. However, despite these shortcomings, the prototype successfully established the groundwork for subsequent development and enhancements.

Sonic Alchemist's ability to generate unexpected details and sound combinations highlighted its potential for innovative sound design. By directing the algorithm to specific sections of the sound effects library, it could produce the required layers, such as whooshes, foley, and vehicle sounds, while also incorporating novel elements that may not have been readily apparent to an experienced sound designer. Future development would focus on enhancing the system's video detection capabilities, audio analysis, weight-based sound categorization and reducing the reliance on manual input.

Performance Enhancements and Transition to C++

The initial Python prototype, while effective, faced several limitations in terms of performance and speed. Specifically, the indexing computation of the 30,000 sound effects took an extensive 24 hours to complete on the development Windows system. To address these issues and improve portability, the algorithm was revised and rewritten in C++. This transition aimed to optimize performance, ensure cross-platform compatibility (specifically targeting Windows and macOS), and facilitate the integration of custom audio processing algorithms.

The first public C++ prototype, Sonic Alchemist v0.5, was presented at the Griffith Film School in Brisbane, Australia, in October 2022. Notable improvements included drastically reduced sound library scanning times and the implementation of multithreading. The indexing process of the same sound effects library ran over 20 times faster. The prototype also functioned as a standalone desktop application, eliminating the need for dependency on complex system-dependent scripts and third-party software.

The C++ prototype integrated an AI-powered video analysis feature that significantly reduced the need for human guidance. The system could automatically generate video events (sync points) by analysing the input video, resulting in a more streamlined and efficient sound design workflow.

Enhanced DAW and NLE Compatibility

To ensure compatibility with a wider range of Digital Audio Workstations (DAWs) and Non-Linear Editing systems (NLEs), the C++ prototype shifted from a Steinberg Nuendo XML-based workflow to a more universally compatible approach. The Advanced Authoring Format (AAF) was initially considered but ultimately abandoned due to its limitations in handling mixed sample rates and channel counts, both of which were crucial requirements for Sonic Alchemist.

Instead, the decision was made to utilize the Broadcast WAV timestamping feature, which allowed the generated audio files to be automatically synchronized when imported into DAWs or NLEs. This approach enabled Sonic Alchemist to accommodate sound files of varying sample rates and channel counts, ensuring compatibility with a wide range of audio production systems.

Integrating Computer Vision for Audio Synchronization

For the effective synchronization of sounds with visual cues in Sonic Alchemist, the generation of a comprehensive list of sync points, relevant parameters, and associated keywords was necessary. This detailed collection of data would be used to guide the software in making critical decisions about the placement and selection of sounds. Various technical methods were thoroughly explored to meet this requirement. The following sections delve into this investigative process, examining each considered method and its potential application in sound design.

Section 1: Sound Examples and Potential Techniques for Video Analysis

Optical Flow Analysis (Barron et al., 1994) can be used as a technique for tracking and matching sound effects to the intensity and speed of movements within a video. Depending on the intensity and pace, these can range from sharp whooshes for fast or abrupt motions to subtler sweetener sounds like soft rustling or lapping water for slower, more gentle movements. Certain actions or events in a video, like impacts and object collisions, may necessitate the addition of sound effects with a fast attack. Motion Detection and Deep Learning-Based Object Detection can be employed to identify these actions and generate more accurate synchronization points (Moeslund et al., 2006; LeCun et al., 2015). Scene Understanding and Semantic Segmentation techniques can be utilized to discern the context and mood of a scene, aiding in the appropriate assignment of background sounds, which can range from environmental sounds for natural scenes to room tones for indoor sequences (Garcia-Garcia et al., 2017). There can be instances where sound effects can be anticipated based on the unfolding sequence of events in a video. For example, the sound of a splash can be anticipated if a person in the video is shown running towards a pool or the sound of a car crash can be anticipated if two vehicles are shown heading towards each other at high speeds. In such scenarios, Generative Adversarial Networks (GANs) can prove to be particularly effective in predicting and aligning these sound effects accurately (Goodfellow et al., 2014).

Section 2: Sonic Alchemist's Hybrid Audio-Visual Synchronization System

While these video analysis techniques show great potential, their integration into a sound design tool presents a host of challenges. Elements like lighting conditions, camera motion, and occlusion can significantly influence the effectiveness of these techniques. The creation of robust algorithms that can withstand these variations is essential for the precise synchronization of sound effects. For sound designers to leverage video analysis without requiring in-depth technical knowledge, these advanced techniques must be enveloped within user-friendly interfaces.

A hybrid approach was therefore adopted for Sonic Alchemist. The resulting system did not aim for perfect synchronization but instead embraced a certain level of unpredictability. This system upheld Sonic Alchemist's mission of inspiring users to discover unexpected audio-visual connections and fostering creative exploration in sound design. The choice was influenced by aims to secure swift processing times, preserve user privacy and security through an entirely offline process, and maintain the software's philosophy of provoking exploration and embracing unexpected combinations.

Advanced Audio Classification Techniques for Sound Design

Section 1: Audio Analysis Techniques

The identification of suitable sound effects, designed to align with specific on-screen actions, emerges as a complex undertaking due to the broad range of available sound effects and unique scene requirements. Sophisticated audio analysis technologies are deployed at this juncture, playing an instrumental role in automating portions of the process and fostering disruptive, creative ideas.

Convolutional Neural Networks (CNNs) and 3D Convolutional Neural Networks (3D CNNs) are employed in the analysis of spectrograms of sound effects, thereby enabling sounds to be classified based on spectral features learned (Krizhevsky, Sutskever, & Hinton, 2012). For instance, in scenes involving a busy city street, a CNN is capable of distinguishing between sounds of car horns, chatter, and background music (Krizhevsky et al., 2012).

Long Short-Term Memory (LSTM) networks offer valuable assistance when dealing with sequential data. They can help understand sequences or changes in sound over time, making them useful for classifying chains of sound effects representing specific events or actions (Hochreiter & Schmidhuber, 1997).

Deep Neural Networks (DNNs) play a critical role in recognizing different types of sounds. For example, a DNN can be trained to discern various environmental sounds, extending beyond mere sound identification to comprehend sound characteristics as well (LeCun, Bengio, & Hinton, 2015). Advanced applications of DNNs in sound design involve classifying sounds based on the emotions or moods they evoke (LeCun et al., 2015).

Autoencoders are used to create compact representations of audio data. For instance, if a specific type of dog bark is needed as a sound effect, an autoencoder could assist in finding the most similar sounding barks within the sound library (Hinton & Salakhutdinov, 2006).

The ability to classify sounds automatically can be significantly enhanced by leveraging pre-trained models from TensorFlow, such as VGGish and YAMNet (Hershey et al., 2017; Gemmeke et al., 2017). For instance, VGGish can assist in classifying diverse animal sounds, while YAMNet simplifies the process of identifying suitable sound effects by categorizing the sound library automatically (Hershey et al., 2017; Gemmeke et al., 2017).

The Fourier Transform technique is used for translating an audio signal from the time domain to the frequency domain, which reveals the diverse frequencies that make up the audio signal. This technique helps to categorize sounds based on their unique frequency content, thereby creating sound 'signatures' that assist in identifying and categorizing unfamiliar sounds (Bracewell, 2000).

Section 2: Sonic Alchemist's Audio Engine

A decision is made to design a hybrid audio classification system for Sonic Alchemist's audio engine, echoing the approach employed for the video engine. A variety of techniques are used to extract audio information, and a weight-based distinction mechanism is integrated. The audio engine functions independently of the existing BWAV metadata, allowing users to effectively navigate sound libraries.

Individual events from audio files containing multiple instances of similar sounds can be isolated by the audio engine in Sonic Alchemist. Moreover, an algorithm for automatic sound trimming and fading is employed. Sonic Alchemist's audio engine uses an auto-mixing algorithm, which adjusts to the audio content and corresponding visual content, avoiding distortion. An adaptive threshold system is used to handle files of varying volume.

A variety of features, such as tape-style pitch shifting, reverse effects, frequency shifting, and convolution reverb are included in Sonic Alchemist's effects engine. The audio engine supports files of varied channel counts and sample rates, accommodating up to 8-channel interleaved files at a sample rate of 384kHz.

Chapter 4: Understanding Sonic Alchemist

This chapter provides a comprehensive exploration of the finalized software, delving into its core concept and unique functionalities.

At the heart of the Sonic Alchemist interface are two fundamental concepts - Events and Materials. These two elements form the backbone of the sound design process within the software. In this context, 'Events' correspond to sync points, the moments in time that are identified as requiring an audio accompaniment to enhance the visual content. They may be actions in a scene, transitions, or any other significant moments in the timeline that would benefit from a sound effect. 'Materials', on the other hand, refer to the sound effects themselves. They encompass a broad spectrum of auditory elements that can be matched to the events.

The essence of Sonic Alchemist's functionality is the matching of Materials to Events. When a sketch is generated, the software intelligently pairs sound effects from the Materials with the defined Events. This process is driven by a hybrid AI system.

The user interface of Sonic Alchemist is designed to be intuitive and comprehensive, enabling users to navigate through various controls from a single window:

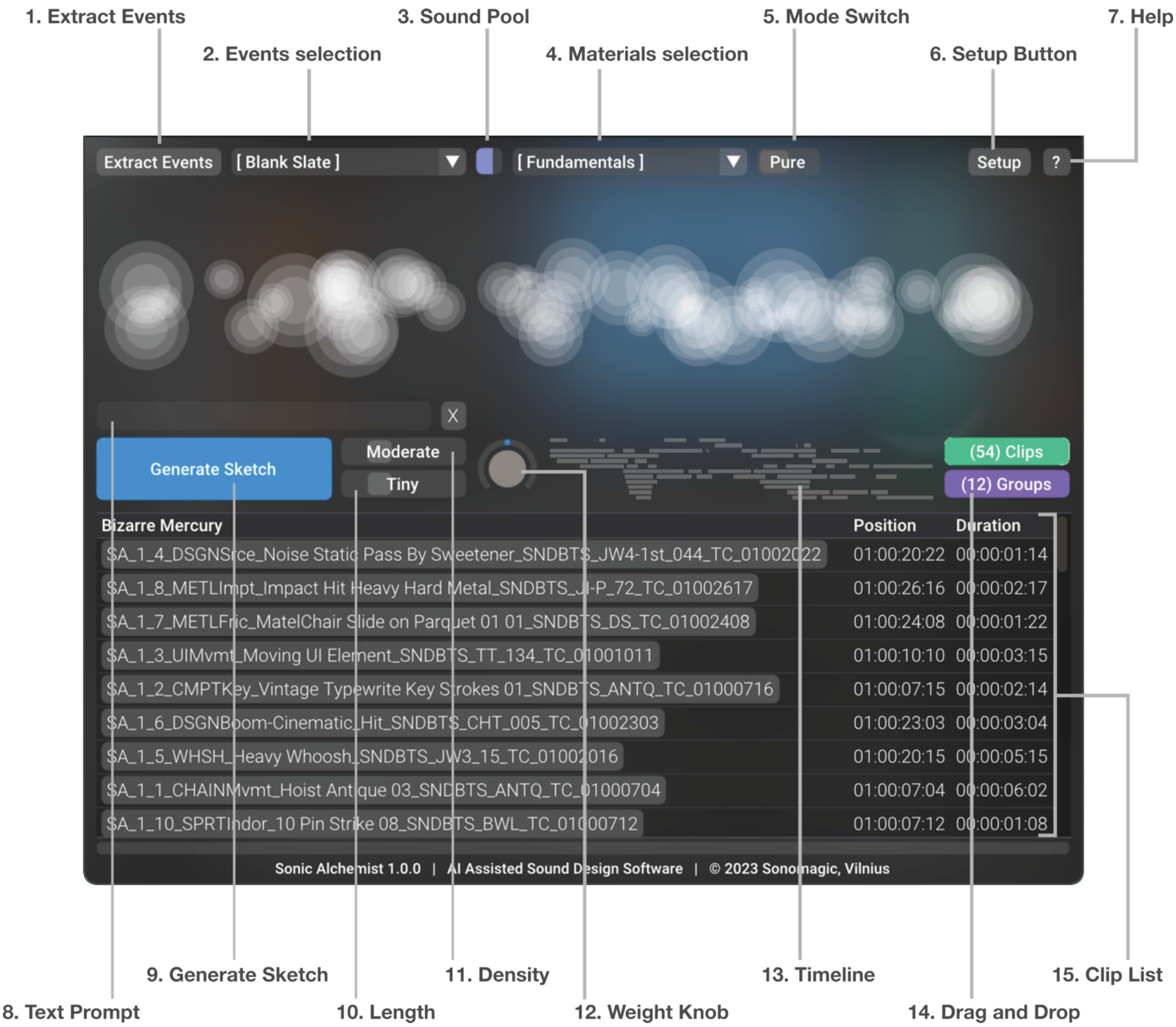

Fig. 2 Sonic Alchemist: User Interface

1. **Event Extraction**: This allows users to import events from video files or load markers from Pro Tools Session Info and Steinberg Nuendo Marker CSV files.
2. **Event Selection**: Users can select, rename, or delete the extracted event files, and choose the least predictable 'Blank Slate' mode.
3. **Sound Pool**: Provides a visual indication of how much a sound effects pool aligns with the defined criteria based on weight and text prompt input.
4. **Material Selection**: Allows users to select a sound effects library that will be used for selecting sounds.
5. **Mode Switch**: Offers two sound generation modes - 'Pure' and 'Alloy', which are explained in more detail further in this text.

6. **Setup Button**: Allows for importing sounds, applying effects, adjusting project parameters, and managing options.
7. **Help**: Provides software information and a quick start guide.
8. **Text Prompt**: Allows users to input a short scene description or keywords to refine their selection.
9. **Generate Sketch**: Facilitates the creation of a foundational sketch by matching, trimming, fading, synchronizing, and timestamping audio to the picture.
10. **Length Control**: Lets users set the sound duration, with the algorithm tending to favor shorter or longer sounds.
11. **Density Control**: Defines timeline density, offering a range from Sparse to Crowded and Jammed.
12. **Weight Knob**: Allows users to focus on specific sounds based on their weight - heavy to light.
13. **Timeline**: Offers a multitrack view of the sketch as arranged over time.
14. **Drag and Drop**: Allows users to import the generated sounds directly into their digital audio workstation (DAW) or non-linear editing system (NLE) by dragging either individual timestamped audio clips, or a multitrack arrangement.
15. **Clip List**: Displays the name, position, and duration of each clip. Users can preview sounds individually or try out combinations in real time polyphonically.

Fig. 3 Sonic Alchemist. Video analysis window.

Sonic Alchemist: Workflow

This chapter outlines the step-by-step process of creating a sound design sketch using the Sonic Alchemist software.

Import and Analyse Video

The initial step in the Sonic Alchemist workflow is importing a video file. This could be a scene from a film, an animated sequence, or any visual content that requires sound design. After import, Sonic Alchemist's video engine scans the visual content, identifying potential synchronization points (Events) that could benefit from accompanying sound effects.

Selection of Events and Materials

Once the Events file is generated, it can be selected in the software interface. The user then chooses a Materials library that contains the sound effects they wish to incorporate into their project. These Materials, or sounds, will be matched to the Events, or sync points, identified in the previous step.

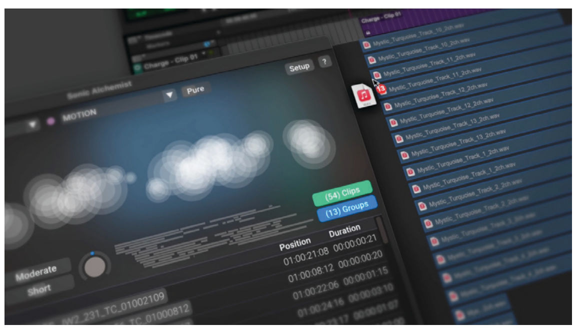

Fig. 4 Sonic Alchemist. The process of importing a multitrack arrangement into Avid Pro Tools.

Generate Sketch

The "Generate Sketch" function initiates the main algorithm that is then used to match, collect, trim, fade, apply effects to, and timestamp the chosen audio to each Event. If selected, Sonic Alchemist can also assemble a multitrack sequence, using a clustering algorithm to optimize track usage.

Preview and Edit

Upon completion of the sketch, it can be dragged directly into the user's Digital Audio Workstation (DAW) or Non-Linear Editing (NLE) system for previewing in sync with the picture. The software supports BWAV timestamping, allowing automatic placement of the clips at the appropriate timeline positions. Following preview, users can retain, remove, or alter sections of the audio as required.

Iterative Layering

The next stage involves iterative layering of additional sketches. This involves generating further sketches to overlay on the existing sequence, enhancing and diversifying the sound design. The Sonic Alchemist workflow encourages users to experiment with different weight and length settings to construct diverse audio layers. Additionally, users can use the Text Prompt feature to refine sound selection by entering keywords or descriptions, which are processed by Sonic Alchemist's natural language engine.

Event Marker Import

In the event of undetected visual actions requiring specific sound effects, Sonic Alchemist allows for marker track import from Pro Tools or Nuendo. The software interprets these markers as sync points and generates a new Events file. Users can assign different keywords to individual markers, allowing tailored sound selection for each sync point.

Mode Selection

Sonic Alchemist offers two modes for users - Pure and Alloy. In Pure mode, the software uses original audio copies, maintaining the character of the source material. This mode is preferable when source quality preservation is paramount. Conversely, the Alloy mode uses a custom algorithm to breed new sounds by merging features extracted from different sounds in the Materials library, offering extensive possibilities for sound variation.

Blank Slate Mode

For a more explorative approach, Sonic Alchemist provides the Blank Slate mode, where the algorithm picks and places sounds at unexpected positions, stimulating creative possibilities through the principle of accidental synchronicity.

A Catalyst for Curiosity

A key aspect of Sonic Alchemist that warrants emphasis is its intended role as a catalyst for curiosity, not as a tool for producing a polished final product. It would be highly improbable to generate a usable, narrative-appropriate sound design exclusively through random methods. Nevertheless, Sonic Alchemist excels in swiftly populating fast-paced sequences. It can adeptly transform an empty canvas into a fully realized timeline teeming with diverse sound combinations.

A foundational sketch can be efficiently generated using Sonic Alchemist, where several passes incorporating a blend of random, semi-random, and precise sync points are layered. Once a number of promising starting points have been identified, the scene can be finalized using traditional techniques. These may include foley recording, manual editing of sound effects, dialogue addition, and music integration.

Sonic Alchemist, thus, serves as a vital stepping stone in the journey from nothing to something. It presents an unconventional entry point into the sound design process, stimulating creativity and exploration. As such, Sonic Alchemist should not be viewed as a tool for creating a final product, but rather as a mechanism to ignite the early stages of the sound design process in a refreshing manner.

Advanced Audio Features

Weight-based Sound Selection

Sonic Alchemist's weight-based audio classification algorithm, facilitated through the Weight Knob feature, provides an intuitive and efficient pathway for users to explore sounds of varying 'weights'. This allows the artist to navigate through a broad spectrum of sound effects, from the notably heavy to the subtly light, relying purely on their inherent sonic characteristics, thus eliminating dependency on text descriptions or metadata.

The Weight Knob is implemented as a continuous control. On the far left, it prioritizes heavier sounds like booms and earthquake rumbles. Gradually moving the knob towards the centre shifts the selection towards medium-heavy sounds such as roaring oceans and animal growls. Nearing the centre, it begins to concentrate on moderate sounds, including car engines and substantial door creaks. The centre position effectively disables the algorithm, permitting a mix of sounds with varying weights. Turning the knob further to the right makes the system favour lighter sounds such as smaller animal calls and running water. As we continue towards the far right, the focus transitions to even lighter sounds, like the shattering of glass and high-pitched whistles. At the extreme right setting, the algorithm zeroes in on the most delicate of sounds, such as the rustling of leaves or the faint buzzing of tiny insects.

To illustrate the utility of the Weight Knob, a session was conducted using a set of TR-808 drum samples. At the far-left position, the system picked heaviest bass drums. As the knob was turned towards the centre, it began to select low toms, along with the bass drums. Further turning of the knob focused on deep snares and toms, and close to the centre, the system selected snares. As the knob moved to the right off-centre, the system picked higher snares and claps, then claps and cowbells. Further towards the far-right, it included rides and high-hats, and at the extreme right position, it selected only the lightest sounds, like shakers and maracas.

The Weight Knob feature encourages a shift from the conventional mindset of adhering to familiar keywords, and instead, promotes an explorative approach to sound design, analogous to freestyle painting.

Alloy Mode

The Alloy mode in Sonic Alchemist operates by 'breeding' new sounds through a process called transmutation. This involves combining features extracted from different sound files in the Materials library. The software applies a set of audio processing techniques and custom algorithms that analyse and break down the input audio into various sonic elements, or 'features'. These features might include aspects such as pitch, envelope, texture, and other sonic attributes. The system then 'recombines' these features to create new, unique sounds. This blending process isn't a simple layering or mixing, but rather an intricate morphing that allows for the generation of thousands of variations from just a handful of audio files. By altering the inputs and parameters, artist can endlessly experiment with the sound output, leading to a myriad of potential sound designs. The unpredictable nature of this feature can lead to exciting surprises, yielding results that might not be traditionally conceived, but hold potential for unique and compelling sound effects.

User Engagement and Expectations in Sonic Alchemist's Testing Phase

In the post-ChatGPT era, traditional tools for fostering creativity maintain a resilient relevance. Nevertheless, the efficacy of instruments, such as card-based creative tools, hinges upon the user's willingness to both engage with the prompts and embrace the inherent uncertainty they introduce.

Throughout the testing phase of Sonic Alchemist, a recurring observation was the predominance of user expectations for immediate, conventional results that align with mainstream production norms. In this post-ChatGPT era, many users of the test software sought instant gratification in the form of solutions that echo familiar concepts from previously seen films, technically flawless output, effectively reinforcing a cycle of borrowing and recycling existing ideas.

In contrast, it was primarily seasoned film sound design professionals and experimental filmmakers who were able to fully embrace the concept of using randomness as a catalyst for creativity. These users understood and valued the notion of using random, unpredictable sketches as a starting point, accepting the unconventional direction they offered.

Yet, these observations only reaffirmed the decision to continue designing a tool that would, much like the Oblique Strategies cards, provoke unconventional thought. By offering vague and unexpected suggestions, Sonic Alchemist encourages exploration of uncharted creative territories, even if the direction seems unusual or unfitting at first. The goal is not to provide a tailor-made solution, but rather to steer the user down an unfamiliar path ripe for exploration.

Conclusions

The advent of AI has stirred apprehension in many sectors, including the film and audio industry, where concerns of AI replacing human roles are often voiced (Rainie & Anderson, 2017). However, AI's entry into these industries need not signify the end of human involvement. Instead, AI can function as an augmentation tool, enhancing human creativity rather than replacing it. Sonic Alchemist is a prime example of this approach, designed to work as a creative partner and a storytelling companion, rather than a replacement for human ingenuity.

Sonic Alchemist enhances the creative process by presenting users with a unique way of exploring the sonic landscape. It assists in creating audio sketches by intuitively matching sound effects to visual sync points, inspiring fresh insights and unexpected solutions. This aligns with the Gestalt theory of continuity, which suggests that humans innately seek patterns and continuity in sensory experiences (Wertheimer, 1923). The inclusion of the Weight Knob function provides an innovative method for filtering and selecting sounds, further promoting creative exploration of the sound library.

Sonic Alchemist integrates the principles of the Beginner's Mind concept, which encourages open-mindedness and receptivity to new ideas (Suzuki, 1970). This tool offers an unconventional approach, provoking sound designers to explore beyond established patterns and embrace innovative sound design practices. It does not aim to produce conventional results; rather, it presents a range of options that encourage exploration and stimulate creativity.

The software also enhances integration within the filmmaking process. It allows video editors and directors to sketch out ideas during editing and pre-production, thereby informing decisions before reaching audio post-production where these ideas are further refined. This integrated approach helps to create a consistent vision across the production process, from initial concept to final output.

In essence, Sonic Alchemist serves as an augmentation tool, facilitating human creativity by automating certain elements of the sound design process. This frees the user's cognitive resources, enabling them to focus more on the creative aspects of their work.

Sonic Alchemist's core philosophy is to demonstrate the irreplaceability of humans in creative contexts. It is a light-hearted tool, designed to inspire rather than intimidate, and to augment human creativity rather than replace it.

This willingness to unlearn and rediscover, even in the face of advanced tools and techniques, remains the pivotal constant in the ever-changing equation of creative innovation.

References

Apple Inc. (2023). Core ML | Apple Developer Documentation. Retrieved from https://developer.apple.com/documentation/coreml/

Bracewell, R. N. (2000). The Fourier transform and its applications. McGraw-Hill.

Eagleman, D., & Brandt, A. (2017). The Runaway Species: How human creativity remakes the world. Catapult.

Gemmeke, J. F., Ellis, D. P., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M., & Ritter, M. (2017, March). Audio set: An ontology and human-labeled dataset for audio events. In Proceedings of the IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 776-780). IEEE.

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., . . . Wilson, K. (2017). CNN architectures for large-scale audio classification. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 131-135). IEEE. doi: 10.1109/ICASSP.2017.7952132

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504-507.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Rainie, L., & Anderson, J. (2017). The future of jobs and jobs training. Pew Research Center.

Sawyer, R. K. (2006). Explaining creativity: The science of human innovation. Oxford University Press.

Suzuki, S. (1970). Zen Mind, Beginner's Mind: Informal Talks on Zen Meditation and Practice. Weatherhill.

Thom, R., & Isaza, M. (2011). Behind the Art: Randy Thom. Retrieved from https://designingsound.org/2011/01/12/behind-the-art-randy-thom/

Tzara, T. (1920). Dada Manifesto on Feeble & Bitter Love. In Motherwell, R. (Ed.), (1951). The Dada Painters and Poets: An Anthology. Wittenborn, Schultz.

Wertheimer, M. (1923). Laws of Organization in Perceptual Forms. Untersuchungen zur Lehre von der Gestalt II.