Immersive VisualAudioDesign: Spectral Editing in VR

Lars Engeln
Technische Universität Dresden
Dresden, Germany
lars.engeln@tu-dresden.de

Natalie Hube
Technische Universität Dresden
Dresden, Germany
natalie.hube@tu-dresden.de

Rainer Groh
Technische Universität Dresden
Dresden, Germany
rainer.groh@tu-dresden.de

ABSTRACT

VisualAudioDesign (VAD) is an attempt to design audio in a visual way. The frequency-domain visualized as a spectrogram construed as pixel data can be manipulated with image filters. Thereby, an approach is described to get away from direct DSP parameter manipulation to a more comprehensible sound design. Virtual Reality (VR) offers immersive insights into data and embodied interaction in the virtual environment. VAD and VR combined enrich spectral editing with a natural work-flow. Therefore, a design paper prototype for interaction with audio data in an virtual environment was used and examined.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; *Sound-based input / output*; Gestural input; Interface design prototyping;

KEYWORDS

VisualAudioDesign, Virtual Reality, Immersive Analytics, Immersive Audio, Spectral Editing, Paper Prototype

ACM Reference Format:
Lars Engeln, Natalie Hube, and Rainer Groh. 2018. Immersive VisualAudioDesign: Spectral Editing in VR. In *Audio Mostly 2018: Sound in Immersion and Emotion (AM'18), September 12–14, 2018, Wrexham, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3243274.3243279

1 INTRODUCTION

In the beginning of the past century, electronic devices like the playback-machine PBG 2 (cf. [1], inventory number: 30613) indicates, that a spectrogram can be transformed back to the time-domain. It was demonstrated in electric-acoustical lectures to help students understand the frequency-domain and the corresponding transformations.

Nowadays, spectrograms are a common method to analyze audio signals. For instance, after applying and processing audio effects, the spectrogram can demonstrate the impact of the modifications. However, process and analysis tools are usually separated. Although interaction within the visual analysis provides a more comprehensible audio design [7] (*VisualAudioDesign*, VAD).

Common spectogram manipulations are transposition, restoration, and compression. Thus, the visual manipulation can also be utilized for more advanced spectral editing [16]. Particularly because little spectral modifications are not experienced as unnatural or synthetic (compare [11]).

A spectrogram, construed as a pixel-based representation, can be treated with image processing to achieve a VAD. This is an opportunity in contrast to direct value input of the DSP algorithm's parameter, for a more comprehensible sound design. In recent years, Virtual Reality (VR) became popular in the game industry, but also adds great potential for visualization and immersive analytics by means of natural interaction. VAD manipulation in the VR is a promising way to work with audio data in a 3D environment in a creative and comprehensible manner.

2 RELATED WORK

The related work is divided into the research fields VisualAudioDesign (image-based spectral editing) and Virtual Reality, which are then combined in a single concept in Section 3.

2.1 VisualAudioDesign

Spectral editing is often achieved as some kind of a drawing tool. Frequencies are filtered (rejected) with an eraser, as well as rectangles are cut out of the frequency-domain, which are then transformed and transposed [7]. In addition, polygonal and freehand selection techniques [3] allow even more freedom for damping, enforcing, and duplicating magnitudes, as well as for time-stretching [2]. Moreover, filters and transpositions with metaphorical manipulations of the frequency-domain with fluid simulation are indicated [8].

Also, a sparse audio representation is used to manipulate audio in a vector graphic based manner [17], instead of actual pixel manipulation.

Analyzed partials are replicated for creating new or hybrid sounds (spectral mapping) [21] and are used for audio morphs. Furthermore, partials are edited via multi-touch as well [9].

2.2 Virtual Reality

To understand the concept of VisualAudioDesign (VAD) in conjunction with Virtual Reality (VR), it is important to first understand the terms *immersion* and *emersion* in context [12]. Immersion is defined as diving into the virtual environment and refers to the complete entry into an artificial world, the virtual reality [4].

This concept of immersion refers to a condition, in which the user loses consciousness of being in an artificial world. The user engages with all his senses in the experience and can interact within this environment. In contrast, emersion provides an overview, rather like the bird's eye perspective as a top-down view.

For instance, the metaphor of the *window* in desktop applications, in which the user perceives and acts *from the outside*, refers to the term emersion (c.f. [12]).

AM'18, September 12–14, 2018, Wrexham, United Kingdom

ACM ISBN 978-1-4503-6609-0/18/09.
https://doi.org/10.1145/3243274.3243279

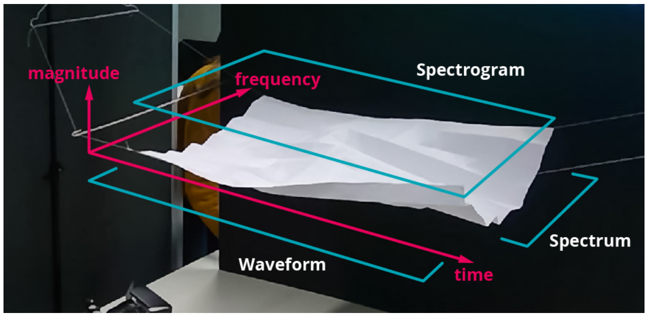

Figure 1: Paper prototype setup.
A consistently folded piece of paper impersonates a 3D spectrogram for real-world examination. Different views at the time-frequency-domain yield waveform, spectrum (see Fig. 2), and spectrogram as typical audio visualizations.

With a glance at the benefits of VR, [23] recommend to use those in order to allow the user to accomplish or display things, that are usually not possible within a real environment and use it for a data-driven experience. Additionally, interactive systems can significantly influence and improve the user's experience [13].

In terms of VAD it is relevant to mention, that the structure of a sound is difficult to recognize without proper training, and any visual assistance enhances that recognition [15]. Therefore, the aspects of enhancing the experience via interaction and visualization are critical, if developing user-centered VR applications.

Few examples exist in the field of VR for audio applications. [19] allow user's observation and navigation control with focus on human perception of sound and color, while [14] visualizes sound fields with see-through head-mounted displays to facilitate their understanding. [20] emphasize the importance of understanding audiovisual perception to develop interactive virtual reality systems in the context of audio visualizations.

In addition to these works, depth cues are substantial to create presence for the user and intensify the immersive experience [10]. VR is also used to visualize room acoustics [6] and planning sounds in urban design [20].

3 VAD IN VR

The challenge for VisualAudioDesign is to bring the user closer to an understanding in the virtual environment. Since common visualization tools primarily rely on 2D visualizations, full-immersive 3D environments are rather new to the user. Thereby, the core idea is to combine strengths of 2D and 3D representations, as well as their inherited benefits: 2D can be utilized to gain a quick overview of the whole problem space, whereas 3D is more suitable for detailed investigations on specific points. These aspects refer to the idea of emersive and immersive representations [12].

A paper prototype was build to examine the usage of a 3D spectrogram for VR and how gestures for natural interaction are perceived by the user.

Figure 2: Spectrum (colored edge) with indication of the temporal progression (visible paper).

3.1 Paper Prototype

The design prototype (see Fig. 1) is made of structured paper. This way, mountains can be folded. These peaks and valleys represents a 3D spectrogram.

Immersion grants the potential to move freely in the data set. Different views at the spectrogram offer variable meanings (see Fig. 1). Looking at the frequency-magnitude plane (orthogonal projection, so that the time-axis collapses) a spectrum analyzer

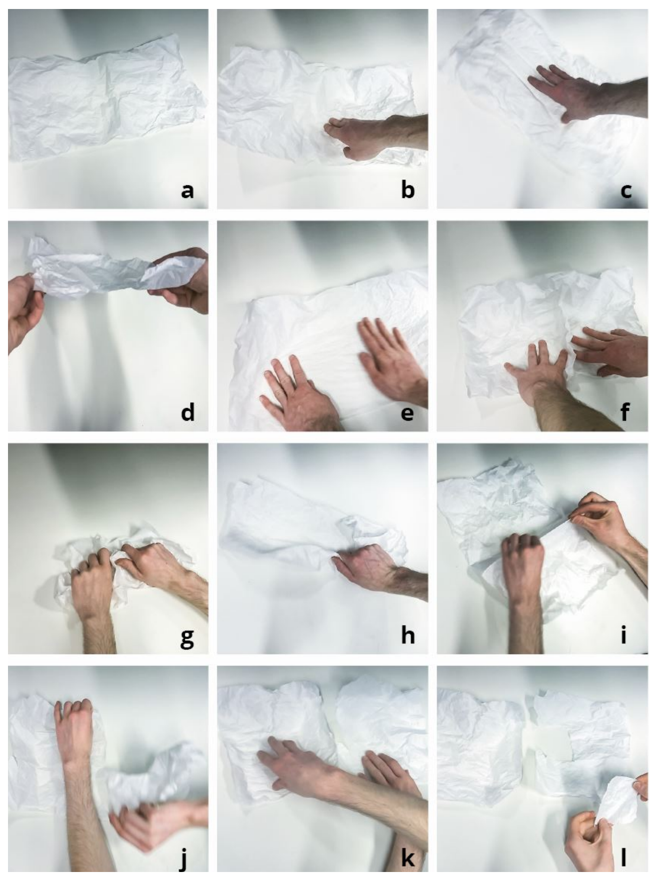

Figure 3: Natural interaction with a 3D-spectrogram. (a) paper sheet - 3D-spectrogram, (b) push, (c) rotate/flip, (d) free view, (e) flatten, (f) squeeze, (g) crunch whole, (h) crunch partly, (i) nip, (j) rip, (k) interchange, (l) overlap/arrange.

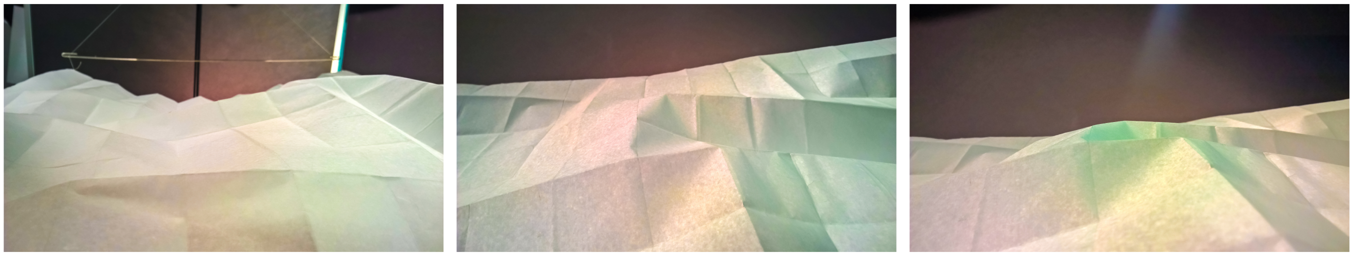

Figure 4: Immersive point of view within the data - climber metaphor: standing on a mountain top.

(cf. Fig. 2) is given and is modifiable via an equalizer. At the time-frequency plane the 3D spectrogram flattens to a 2D spectrogram. At the time-magnitude plane a waveform can be visualized and the amplitudes are manipulated. The data is displayed with no occlusion or perspective deformation.

All orthogonal projected planes (i.e. spectrum, 2D spectrogram, and waveform) are associated with operational emersive tasks. A 2D spectrogram is also useful as a map to improve wayfinding [5].

In contrast, standing on a mountain-top (see Fig. 4) provides an immersive insight into spectral data and allows a creative handling. In addition, manipulations of the 3D spectrogram's spectral data with natural and metaphoric interactions are immersive tasks.

3.2 Natural Interaction

A small design study suggests different natural interactions (manipulative gestures) with paper representing the 3D spectrogram (see Fig. 3). The meaning according to similar manipulated spectral data are explained in Tab. 1 briefly.

Furthermore, the hand is not limited to interact with the 3D spectrogram from a single perspective. Especially in case of working with bi-manual gestures, one hand can push from above and the other from below to shape the spectral data (see Fig. 5).

3.3 Metaphoric Interaction

When thinking of shaping mountains, metaphoric interactions are conceivable. If water (or sand) is poured in a valley, the valley is filled up. This raises the magnitudes in the affected area and thereby a temporal and band-limited noise is added. In case of cutting a mountain-top, it leads to a temporal and band-limited distorted overdrive effect.

A virtual lamp tool manipulates shadows (cf. Fig. 6). By doing so, reverberation, ducking, and frequency-masking like filtering can be created. The light source's angle according to the spectrogram determines the length of the shadow and the distance to the spectrogram determines the softness. For instance, in case of reverberation the shadow length corresponds to the room size (i.e. reverberation length) and the softness corresponds to the damping and diffusion.

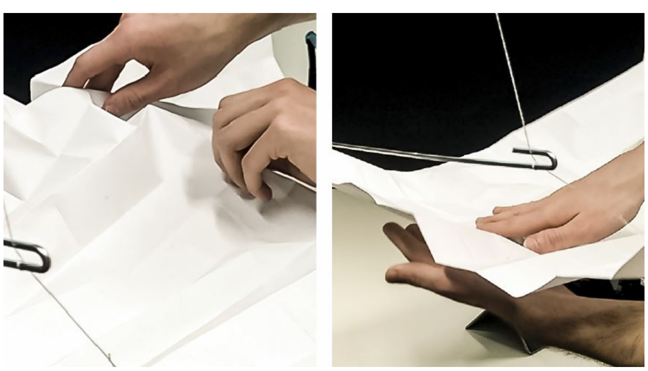

Figure 5: Bi-manual manipulation of the 3D spectrogram - from one side (left) and both sides (right) of the surface.

3.4 Prototypical Implementation

Developing a prototypical implementation is the next step to explore and manipulate 3D spectrograms in VR. To give users a reliable tactile and spatial feedback of the interactions VR-controller are used for precise inputs. To enhance performance of given tasks an horizon [12] and a virtual room as a reference frame [18] are implemented for better orientation.

Interaction	Description
paper sheet (a)	initial state
push (b)	shelving-/notch-filter
rotate/flip (c)	reorders bands or inverts audio signal
free view (d)	analyzing data
flatten (e)	clears spectrogram, removes magnitudes
squeeze (f)	stretches and condenses spectral data
crunch whole (g)	creates noise
crunch partly (h)	creates noise at specific time and bandwidth
nip (i)	creates a burst (vertical nip), or a sinusoidal sound (has a fixed frequency, if nip is horizontal)
rip (j)	cuts the audio signal at a specific time, or cuts out a specific spectral area
interchange (k)	arranges audio or spectral area
overlap/arrange (l)	fades/morphs and arranges/transposes spectral data

Table 1: The following alphabetic notation *(a) to (l)* refers to the equivalent sub-image in Fig. 3.

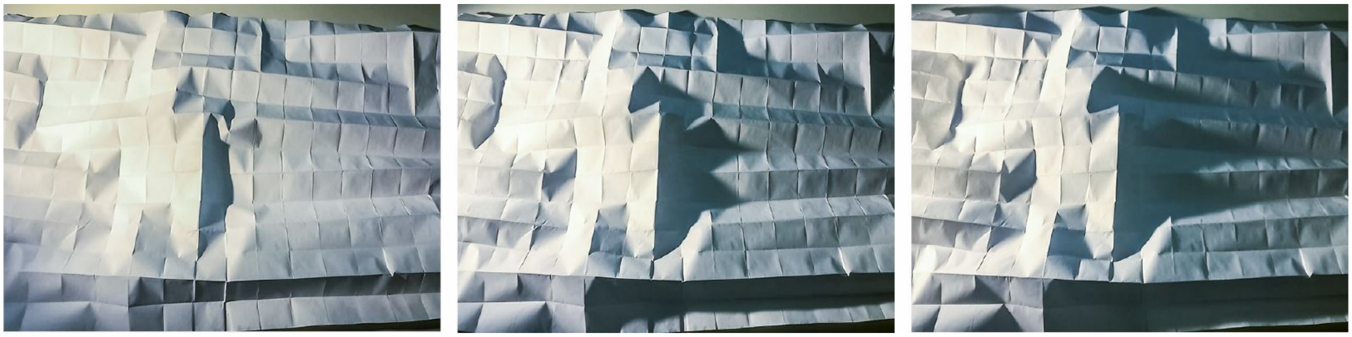

Figure 6: Shadows demonstrate the temporal masking.

The data set for the prototype is created from a 2D spectrogram, which serves as heightmap. Manipulations on the 3D spectrogram are transferred to its 2D counterpart to perform the inverse transformation.

4 CONCLUSIONS

With a paper prototype novel interaction for VisualAudioDesign in VR are demonstrated and natural gestures are proposed. VisualAudioDesign is promising in the field of sound design applications, i.e. for sound textures [22], arrangement, and manipulations in general. Also, with spectrograms in VR a playful knowledge transfer of how audio is structured can be achieved for ambitious students, as well as visitors in the context of museums dealing with electric-acoustical exhibits.

In future works the user experience of VisualAudioDesign interaction techniques and interfaces will be investigated to achieve a creative and comprehensible audio design.

ACKNOWLEDGMENTS

On behalf of Lars Engeln this research (reference no. 01UQ1601 A) has been funded by the German Ministry of Education and Research (BMBF) and on behalf of Natalie Hube, this research (project no. 100271686) has been funded by the European Social Fund and the Free State of Saxony.

REFERENCES

[1] 2014. Historic acoustic-phonetic Collection at the Technische Universität Dresden. http://www.universitaetssammlungen.de/sammlung/687?setLocale=en
[2] Niels Bogaards and Axel Röbel. 2005. An Interface for Analysis-Driven Sound Processing. In *Audio Engineering Society Convention 119*.
[3] Niels Bogaards, Axel Roebel, and Xavier Rodet. 2004. Sound Analysis and Processing with AudioSculpt 2 To cite this version : Sound Analysis and Processing with AudioSculpt 2. *International Computer Music Conference (ICMC)* Icmc (nov 2004), 1.
[4] D. A. Bowman and R. P. McMahan. 2007. Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 7 (July 2007), 36–43.
[5] Rudolph Darken and Barry Peterson. 2001. Spatial Orientation, Wayfinding, and Representation. In *Handbook of Virtual Environment Technology*, K Stanney (Ed.). Number November 2001.
[6] Eduard Deines, Martin Hering-Bertram, Jan Mohring, Jevgenijs Jegorovs, and Hans Hagen. 2010. Audio-visual Virtual Reality System for Room Acoustics. In *Scientific Visualization: Advanced Concepts*, Hans Hagen (Ed.). Dagstuhl Follow-Ups, Vol. 1. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 303–320.
[7] G Eckel. 1992. Manipulation of Sound Signals Based on Graphical Representation-A Musical Point of View. In *Proceedings of the International Workshop on Models and Representations of Musical Signals, Capri, Italia*.
[8] Lars Engeln and Rainer Groh. 2017. AudioFlux : A Proposal for interactive Visual Audio Manipulation. In *Mensch und Computer 2017 - Workshopband*, M. Burghardt, R. Wimmer, C. Wolff, and C. Womser-Hacker (Eds.). Gesellschaft für Informatik e.V., Regensburg.
[9] Lars Engeln, Dietrich Kammer, Leon Brandt, and Rainer Groh. 2018. Multi-Touch Enhanced Visual Audio-Morphing. In *Proceedings of the International Conference on New Interfaces for Musical Expression (in press)*. Blacksburg, VA, USA.
[10] Nicolas Gerig, Johnathan Mayo, Kilian Baur, Frieder Wittmann, Robert Riener, and Peter Wolf. 2018. Missing depth cues in virtual reality limit performance and quality of three dimensional reaching movements. *PLoS one* 13, 1 (2018), e0189275.
[11] John M Grey and John W Gordon. [n. d.]. Perceptual effects of spectral modifications on musical timbres. 63, 5 ([n. d.]), 1493–1500.
[12] Rainer Groh. 2017. *An Iconography of Interaction* (neuerscheinung ed.). TUDpress. 245 pages.
[13] Marc Hassenzahl and Noam Tractinsky. 2006. User experience-a research agenda. *Behaviour & information technology* 25, 2 (2006), 91–97.
[14] Atsuto Inoue, Kohei Yatabe, Yasuhiro Oikawa, and Yusuke Ikeda. 2017. Visualization of 3D Sound Field Using See-through Head Mounted Display. In *ACM SIGGRAPH 2017 Posters (SIGGRAPH '17)*. ACM, New York, NY, USA, Article 34, 2 pages.
[15] H. G. Kaper, E. Wiebel, and S. Tipei. 1999. Data sonification and sound visualization. *Computing in Science Engineering* 1, 4 (Jul 1999), 48–58.
[16] Michael Klingbeil. 2005. Klingbeil, M. (2005). Software for spectral Analysis, Editing, and synthesis. In ICMC.Software for spectral Analysis, Editing, and synthesis.. In *ICMC*.
[17] Michael Klingbeil. 2012. SPEAR: Sinusoidal Partial Editing Analysis and Resynthesis. *retrieved from on Nov* 12 (2012), 3.
[18] Thinh Nguyen-Vo, Bernhard E. Riecke, and Wolfgang Stuerzlinger. 2018. Simulated Reference Frame: A Cost-Effective Solution to Improve Spatial Orientation in VR. *IEEE VR* (2018).
[19] Benjamin I Outram. 2016. Synesthesia audio-visual interactive-sound and music visualization in virtual reality with orbital observation and navigation. In *Mixed Reality Art (MRA), IEEE International Workshop on*. IEEE, 7–8.
[20] Gemma M. Echevarria Sanchez, Timothy Van Renterghem, Kang Sun, Bert De Coensel, and Dick Botteldooren. 2017. Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space. *Landscape and Urban Planning* 167 (2017), 98 – 107.
[21] William A Sethares, Andrew J Milne, Stefan Tiedje, Anthony Prechtl, and James Plamondon. 2009. *Spectral Tools for Dynamic Tonality and Audio Morphing*. MIT Press - Journals.
[22] Gerda Strobl, Gerhard Eckel, Davide Rocchesso, and S Le Grazie. 2006. Sound texture modeling: A survey. In *Proceedings of the 2006 Sound and Music Computing (SMC) International Conference*, Vol. 61. 1–7.
[23] Ruth West, Max J Parola, Amelia R Jaycen, and Christopher P Lueg. 2015. Embodied information behavior, mixed reality and big data. In *The Engineering Reality of Virtual Reality 2015*, Vol. 9392. International Society for Optics and Photonics, 93920E.