



# CoHEARence of audible shapes—a qualitative user study for coherent visual audio design with resynthesized shapes

Lars Engeln<sup>1</sup> · Rainer Groh<sup>1</sup>

Received: 22 November 2019 / Accepted: 5 March 2020 / Published online: 17 March 2020  
© The Author(s) 2020

## Abstract

One way to achieve an intuitive sound design are visual approaches for synthesis and sound collages. Therefore, during spectral synthesis and editing, the sound is designed in a visualization of the frequency domain. In order to create a coherent workflow between visuals and the resulting audio, the stimuli should be matched to each other. In this work, a qualitative user study is presented, which is supposed to show the intuitive understanding from the shape to the sound. The shape is hereby the spectral envelope. The general aim is to find out whether there is a connection between the visual shape and the subsequent auditory impression.

**Keywords** Design study · Coherence · Visual audio design · Spectral editing · Audio-visual

## 1 Introduction

Synthesizer and DSP algorithms are often adjusted via value inputs, sliders or knobs, similar to their hardware counterparts. However, sound design and music production becomes more creative through a visual manipulation of audio.

Furthermore, it was shown that multi sensory convergence exists in the low-sensory processing [34] and for visual-auditive stimuli (see [14, 27]). This low-level processing of coherent sensor inputs allows the improvement of visual and acoustic perception by simultaneous matching stimuli (compare also [4, 34]). The early convergence suggests that acoustic and visual stimuli have a positive effect on each other [26]. The correlation between color and sound was examined with an empirical approach [17] and showed that there is a strong correlation between loudness and saturation as well as tonality and brightness.

Besides feature mapping [17], timbre design [29], wave terrain synthesis [31], scanned synthesis [33], and geometric oscillator [24], spectral editing is a common way to have a direct relationship between the visual (spectrogram) and the (resynthesized) sound. In addition, the research field of *soundtextures* (compare [28]) describes a graphical synthesis that is sometimes done in the frequency domain. As for spectral editing in connection with *SpecDraw* [7], *AudioSculpt* [3] provides tools for damping, enforcing, and duplicating spectrals, as well as for time-stretching [2]. *TAPESTREA* [23] allows the creation of sonic spaces from different sounds. With *Meta-Synth*, images can be used as input within the frequency domain. *SPEAR* [20] got a sparse representation of the frequency domain and is thereby abstracting the spectrogram with partial lines. Furthermore, a touch interface for partial line mapping corresponding to morphing was shown [12]. Moreover, filters and transpositions with metaphorical manipulations of the frequency domain with fluids [8] were proposed, and besides that, affine transformation and image processing effects are introduced with *VisualAudio-Design* [10].

In order to achieve a visual sound design, a user should be able to understand why the visual arrangement sounds the way it does after synthesis. To better understand how sound is shaped, subjects were asked to perform hand gestures for given melodic phrases [19]. To *sketch* sounds, not only graphical shapes are discussed but also the voice

---

✉ Lars Engeln  
lars.engeln@tu-dresden.de

Rainer Groh  
rainer.groh@tu-dresden.de

<sup>1</sup> Chair of MediaDesign, Technische Universität Dresden, Dresden, Germany

as input for complex ideas of sounds is used [18]. On the contrary, with *Track-draw* [1], a speech synthesizer was controlled by drawing. Also, it is proven that movements induced by sounds of drawing scratches are predictable [32]. So, whether a circle, ellipse, stroke, and so on was drawn. In addition, drawing graphic scores were utilized for public and collaborative creation [25]. Likewise, text as symbolic input was sonified as spectral shapes (sine sweeps and envelopes of white noise) [30]. Besides that, with *Sound Mosaics* [15, 16], a system for sound synthesis via influencing graphical variables was created. With the terms of *gestural similarity* [22], it can be assumed that a visual and a sound are coherent to each other, if they are created by gestures and the *gesture generator* remains the same.

To examine the coherence between the visual design and the (re-)synthesized audio, a small qualitative study was conducted with audio- and design-related students. However, the subjects did not order the sounds to the shapes (spectral envelope), or evaluate an existing assignment. Their main task was to visually describe the sound with sketches, whereby the sounds were made by visual shapes in the frequency domain. In this way, the association to a shape while hearing the sound is investigated, and therefore the connection from the visual to the resulting audio could be shown. In addition, the compositional problem according to Klingbeil [21] is been addressed, to compose music in the frequency domain.

The results are classified by function and by visual semiotic meaning. The subject of semiotics is described with its *pragmatic*, *semantic*, and *syntactic* function (compare [6]). Pragmatic describes the effect of a subject, i.e., how a user thinks and feels about a subject. The semantic describes the meaning of the subject and the syntax the general structure of the subject. Thereby, the syntactic structure influences the semantic meaning and the semantic meaning the pragmatic effect. The semantic is differentiated into the *symbolic*, *indexical*, and *iconic* meaning. *Symbols* are characterized by social conventions, like a vocabulary an user has to learn. In the semiotics, the metaphoric meaning is more like a subclass of symbolic, but it is leading to indexical. The indexical meaning references to hints, which allude a contextual meaning. For instance, the pictogram of a floppy disk has the indexical hint (being a storage medium) to save the current documents. For younger users, who do not grew up with floppy disks, the meaning becomes a symbol, although the visual is still pictographic. Thereby, the iconic meaning is the direct reflection of the thing itself (*the storage medium*). The semiotic classification of the semantic was compared to a classification of earcons [11]. Doing this, abstract earcons up to metaphoric earcons were differentiated and organized.

## 2 User study on resynthesized shapes

This user study extends the pre-test [9] by asking additional 16 persons to perform the same tasks as the pre-test. The goal was to get more insights in what people are doing when drawing the sounds they are hearing. That is why the pre-test is briefly considered first; after that, the setup and the tasks for the study are described. Then, the findings are reported, which are discussed afterwards.

### 2.1 User (pre-)test

A pre-test was conducted (see [9]), in which 6 subjects (2 female, 4 male) aged 20–25 took part. Because the pre-test was conducted within a seminar for audio processing, all subjects were at least interested in audio or working with audio. This pre-test should show whether it is generally reasonable to investigate the connection between sound and visual with drawings in depth and which direction should be taken with this very tasks as study method. In Fig. 6, all results for task 1 (see Section 2.4) are shown. The results from the pre-test are A to E in Fig. 6. Only the drawings from five subjects are displayed, because one subject only described the imagination with keywords.

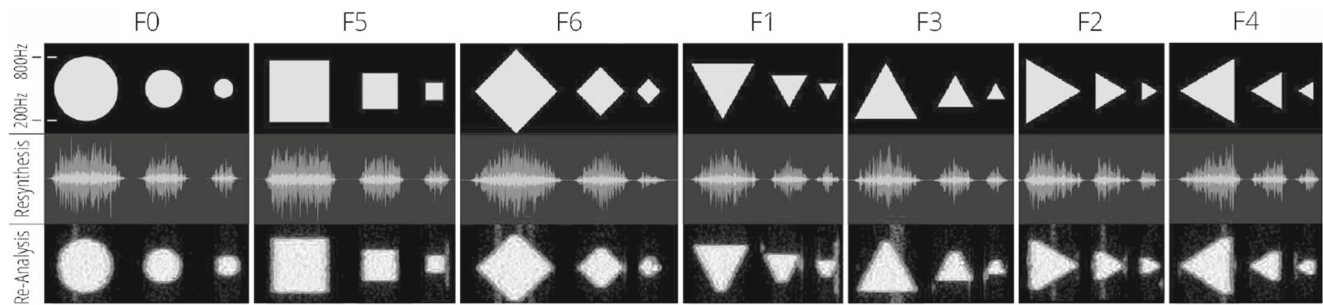
Thereby, A and C drew forms, where A is the only one who has drawn solid forms. D and E used just one stroke and characterized only one single sound from the repetition. A's drawings seem to be based more on spectrograms, B's on waveforms and sinusoids, C's on waveforms, and D's on envelopes (loudness-curve). It seems as if the drawings are based on the iconic meaning, i.e., the representation of the physical signal (with taking the perception into account).

### 2.2 Setup

Seven elementary shapes in different rotations were examined (see Fig. 1). Each shape was repeated three times and with each repetition the shape became smaller, to decouple the shapes size for the examination.

These shapes are interpreted as magnitudes and were resynthesized with a ISTFT using VisualAudio-Design [10]. Although VisualAudio-Design is able to estimate phases iteratively, the phases were taken from white noise, because the shapes can be assumed as band-limited envelopes of the white noise. The window size was 4096 and the lowest frequency of a shape was 200 Hz and the highest 800 Hz (each for the biggest), except for *F6*. For *F6*, it was approx. 10–990 Hz, because *F6* was a affine transformation of the square (*F5*).

The range 200–800 Hz (roughly two octaves from G3 to G5) was chosen to guarantee an almost flat slant in the isophones. These sounds files were made public for



**Fig. 1** Resynthesized shapes for the user test—the shape as input magnitude for the STFT (*upper row*), the resynthesized waveform of the shape (*middle*), and the re-analysed spectrogram of the waveform (*lower row*)

reproduction with the publication of the pre-test [9] on researchgate.<sup>1</sup>

The subjects did not know that the sounds are created by elementary shapes and they were allowed to hear the audio file multiple times. But, they knew that there were three repetitions of the same sound for each  $F_n$ , which become *smaller* and more quiet. The order of the sounds to test were static according to their naming ( $F0$ ,  $F1$ ,  $F2$ , ...,  $F6$ ). So no randomization of the files for each subject was done, although it might be that some people have first listened to all the sounds, or have worked on the sounds in an irregular order.

### 2.3 Participants

In addition to the 5 participants of the pre-test who turned in drawings, a further 16 (3 female, 13 male) took part at the extended study. The mean of the age is 26.7 and has a range of 21 to 43 (compare Fig. 2). All participants are related to (media) computer science and/or media design.

The test group has a equally distributed center at 2.94 (5 is best, 1 is worst) for the experience in graphic design (see Fig. 3) and a lower mean of 2.50 for the experience in sound design. There is also a wide range of how often they make music with a mean of 2.81. Therefore, the test group can be regarded as largely unbiased, although there is a shift towards graphic design and an overhang towards male subjects. More subjects with knowledge in DSP would have been appreciated.

### 2.4 Tasks

For each shape,  $F0...F6$ , the subjects performed task 1 and task 2. In addition, they performed task 3 with the variations  $A1...A3$  (blurred shapes, see Fig. 4) and  $B0...B6$  (outlined

shapes, see Fig. 5). Thus, task 3 was about the influence of variations (blur and outline) of the shapes.

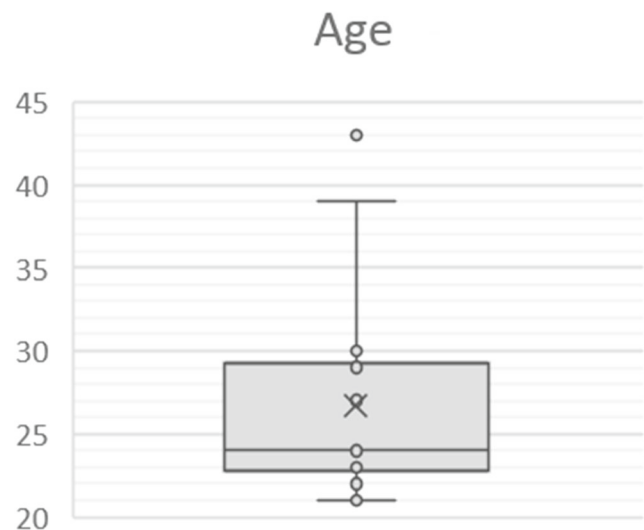
**Task 1:** “How does the audio sound to you? Illustrate your imagination (paint a picture or sketch a shape). Record your sound impression with few key points.”

**Task 2:** “Sort the sounds from ‘I like best’ to ‘I like least’. Give a point from 0-100 (100 is the best, 0 is the worst) and give a short explanation.”

**Task 3:** ( $3a$  will stand for task 3 with  $A_i$  and  $3b$  for  $B_k$ )

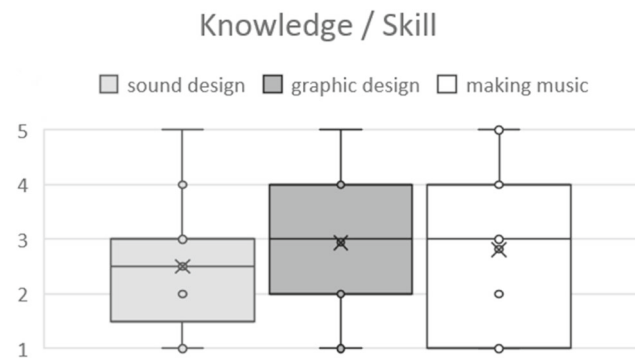
“How do the  $A_i$  and  $B_k$  sounds differ each from each other and from the  $F_n$  sounds?”

Initially, all tasks were performed by the subjects. Afterwards, the results from the pre-test were discussed individually with each participant. Subjects in the extended study only considered their impression with key phrases briefly. To deal with the visual impression of the sound as key points

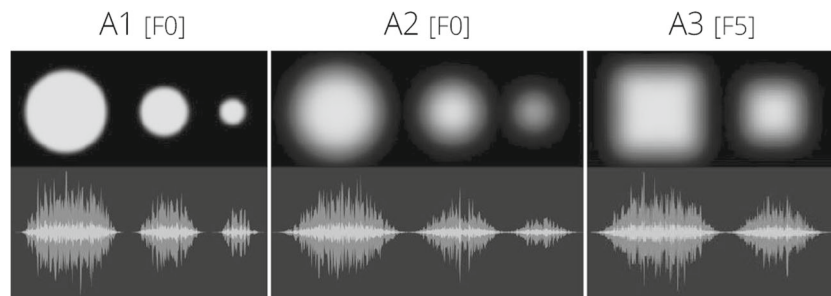


**Fig. 2** Age distribution of 16 subjects from the extended study

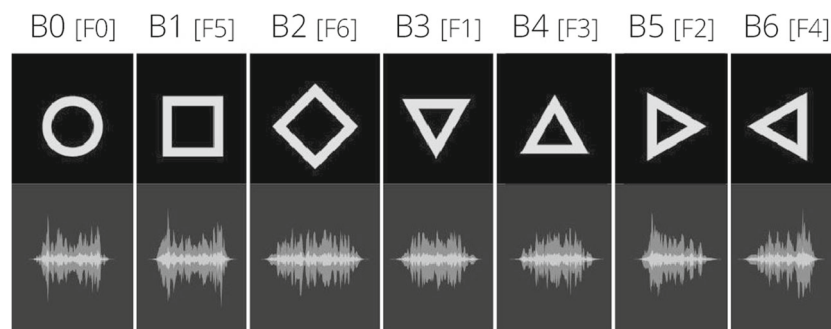
<sup>1</sup>Resynthesized shapes as sound files for reproduction: [https://www.researchgate.net/publication/339178523\\_CoHEARence\\_resynthesized-shapeszip](https://www.researchgate.net/publication/339178523_CoHEARence_resynthesized-shapeszip).



**Fig. 3** Experience distribution of 16 subjects from the extended study. They were asked for the knowledge and/or the skill in the specific domains (5 is best, 1 is worst)



**Fig. 4** Variations of  $F0$  and  $F5$  for task 3a (see Section 2.4)—the shape is slightly blurred (A1) and even more blurred (A2, A3)



**Fig. 5** Variations of all shapes for task 3b (see Section 2.4)—the shapes are outlined (only the second biggest shape was used here)

**Table 1** A brief overview of the results showing stylistics in contrast to the experience (+ is 4 and 5, – is 1). Thereby, *lines* and *shapes* are iconic, *metaphoric* is indexical or symbolic, and *abstract* is symbolic in terms of the representation of the signal

| Subject        | A | B | C | D | E | F | G | H | J | K | L | M | N | O | P | Q | R | S | T | U | W |
|----------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lines          |   | x |   | x | x |   | x | x | x | x | x |   |   |   | x | x |   | x | x | x | x |
| Shapes         | x | x | x |   |   | x |   |   |   | x |   | x | x | x | x |   | x |   | x |   |   |
| Metaphoric     |   |   |   |   |   | x | x |   |   | x |   | x |   |   |   |   |   |   |   |   | x |
| Abstract       |   |   |   |   |   |   |   |   |   |   |   |   |   |   | x | x |   |   |   |   |   |
| consistent     | x |   | x | x | x |   |   | x | x |   | x |   | x | x |   |   | x | x |   | x |   |
| Single sound   |   |   |   | x |   |   | x |   |   |   |   | x | x |   | x |   | x | x | x |   | x |
| Sound design   |   |   |   |   |   |   |   | – |   |   | + |   |   | – |   | – | + | – |   | + |   |
| Graphic design |   |   |   |   |   | – |   | + |   | – | – | + | + | + | + | + |   |   |   |   | + |
| Making music   |   |   |   |   |   | – | + | – |   | + | – |   |   | – |   | – | + |   | + | + |   |

**Fig. 6** Results from the drawing task (task 1, see Section 2.4). Drawings A to E are the results from the pre-test [9]. Participants with high knowledge (4 and 5, compare Fig. 3) in sound design are L, R, U and in graphic design are H, M, N, O, P, Q, V. The participants G, K, R, T, U are making a lot of (4 and 5) music. H, O, Q, S have no knowledge (1) in sound design and F, K, L in graphic design. F, H, L, O, Q do not make any (1) music





**Table 2** Combined characterizations for each  $F_n$  with same order like Figs. 1 and 5 with corresponding comparison to the outlined variations  $B_k$ 

|      |   |      |                                |
|------|---|------|--------------------------------|
| $F0$ | Bulging, blunt, rounded, noisy                  | $B0$ | Soft, round, balanced          |
| $F5$ | Hard, sudden, garish, scratchy, thick, powerful | $B1$ | Scratchy, novel, nice          |
| $F6$ | Pumping, futuristic, roller coaster, thrilling  | $B2$ | Slidy, spongy, vague           |
| $F1$ | Pulsating, vivid, rounded, dull, unpleasant     | $B3$ | Bright, cute, lightly, minimal |
| $F3$ | Swinging, hollow, balanced, symmetric           | $B4$ | Centric, slight breeze         |
| $F2$ | Gurgling, sloshing, pointed, firm, stable       | $B5$ | Simple, minimalistic, blunt    |
| $F4$ | Hollow, windy, yawning, swelling                | $B6$ | Rounded, moving                |

of task 1, the possible sound of the geometric shapes circle, square, and triangle was discussed with the group of the pre-test.

### 3 Findings

The pre-test (see Section 2.1) suggests that the subjects uses the iconic meaning of the perceived signal, so that they would draw lines to visualize envelopes/loudness and waveform-like curves or shapes to visualize bandwidths/spectrograms (compare Table 1 A to E). But in addition, the extended study shows that some subject uses metaphoric and abstract meanings of the sound, for instance, *whirling a rope* (M F6 and F3, see Fig. 6) or *leafing through paper* (V F3). These drawn mental images are actually iconic in themselves, because they reproduce a natural action, but viewed from the perspective of the sound, they are considered as symbolic.

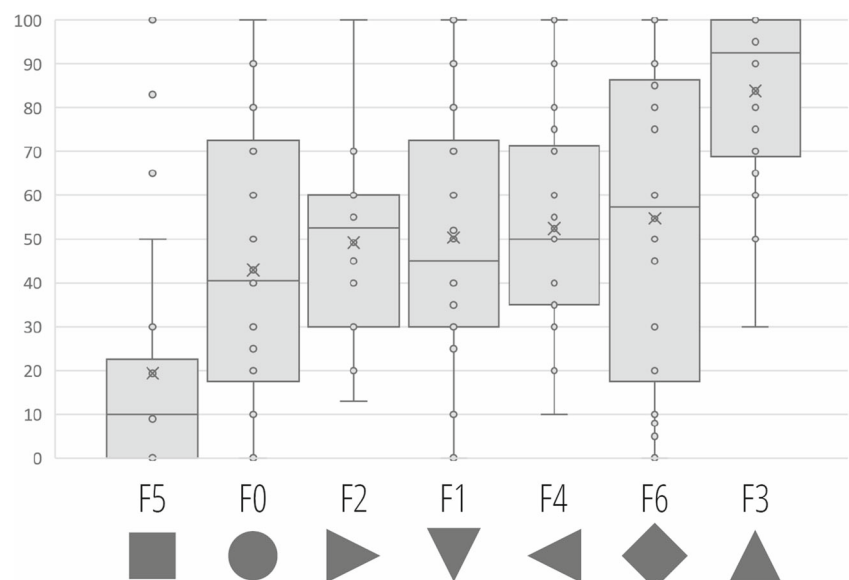
All subjects who used iconic meanings used the  $x$ -axis for the time (from left to right), except for subject D. D used the  $y$ -axis (top-down). For better comparison, D's results were rotated, so that the direction of time matches to the other results. Some people drew only one sound (see

Table 1), although each file had repetitions (see Section 2.2), others drew each repetition solo, or in conjunction with the others.

A is the only one who has drawn solid forms, whereas O and R visualized a solid form with additional lines in the shape. At a first glimpse, the  $F0$  of B, H, J, R, T, U and the  $F5$  of D, H, J, L, N, O, P, R, S, T, U and the  $F4$  of D, G, J, S match perfectly with the corresponding shape, as well as C's  $F5$  and  $F6$ , due to the spikes on it outlining the corners of  $F6$ . B reports that  $F6$  sounds like two tones, so apparently with G's and T's  $F6$ . The combined impression from all subjects for each  $F_n$  is shown in Table 2 (the impressions are stabilized with the feedback from the extended study). Thereby,  $F5$  was mostly negatively characterized, and  $F3$  and  $F6$  mostly positively.

This is also reflected in the rating (see Fig. 7) from task 2. The box plot indicates that  $F5$  is significantly less appealing than  $F3$  and might be less appealing than  $F2$ ,  $F1$ , and  $F4$ , but no variance analysis has been done yet. The outlined versions  $B_k$  shifted the appearance drastically (see Table 2). A sudden garish sound ( $F5$ ) became novel and nice ( $B1$ ), and a *thrilling* sound ( $F6$ ) became spongy ( $B2$ ) for instance. With blurring ( $A_i$ ), the sounds became more pleasant and less sharp, but more noisy. A correspondence between a

**Fig. 7** Results from the rating task (task 2, see Section 2.4) as box plot with exclusive median, sorted by rating—the higher (and therefore more right) the more appealing the sound is





**Fig. 8** One possible syntactic order of all drawings in relationship to their meaning and function. Thereby, equivalence classes for their resembling function like *sinusoidal*, *waveform*, *envelope*, *shape*,

and *spectrogram* (all being indexical) and resembling meaning like *metaphoric* and *abstract* emerges

noisy sound and a blurred shape appears as consistent, due to most of the subject reported this.

## 4 Discussion

First, the results and findings are interpreted. After that, the scope and further fields of interest for coherent spectral design are indicated. Finally, the interface for resynthesis used here is briefly considered.

### 4.1 Interpretation of the results

Various findings should be elaborated and demonstrated individually in specific studies and a clear formalization of coherent result should be done.

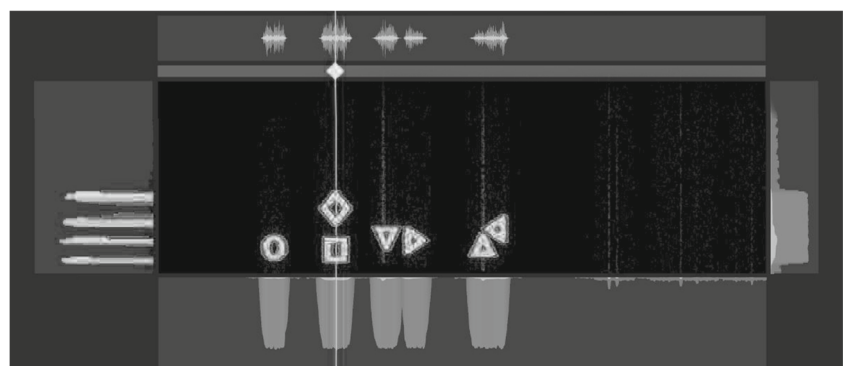
The two triangles pointing up ( $F3$ ) and down ( $F1$ ) were mostly illustrated round and in the case of  $F1$  also verbally

described as a round sound. This impression is reinforced by resynthesis with the ISTFT, since the large analysis window smooths finer resolutions in time like the peak of the triangle (compare the round edges in the re-analysis in Fig. 1). Also,  $F2$  had a rounder impression than  $F4$ , although they were smoothed in time the same way, because of just being flipped according to the time. Here, temporal (post-)masking might benefits the impression of a harder edge of  $F4$ .

The subjects said that they disliked the hard *edges* (no attack and no release) of a sound like  $F2$ ,  $F4$ , or  $F5$ . That could be why the square  $F5$  had the worst ratings (having two *edges*), followed by  $F2$  (no attack) and  $F4$  (no release). The blurring ( $A_i$ ), which softens the hard edges, also made the sounds more pleasant.

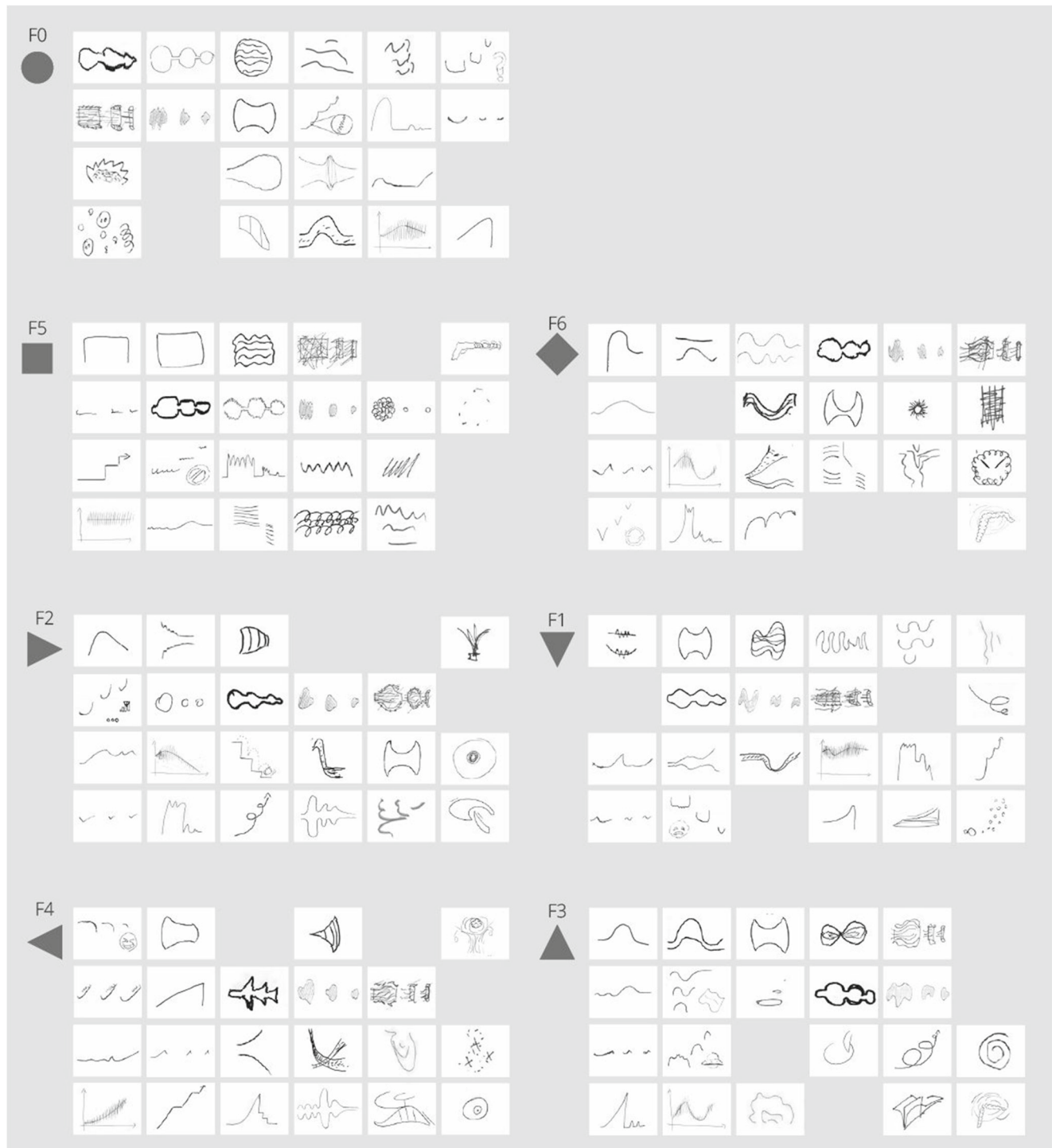
Interestingly, the attack into the higher frequencies seems to be more appealing than the attack towards the lower frequencies (see Fig. 7). This is also the key characteristic

**Fig. 9** Spectral editing with interactive widgets surrounding the spectrogram: spectrum (left), spectral power (bottom), long-time spectrum (right), waveform (top)



for  $F4$  by D, E, F, H, J, and S. The rating of the stacked combination  $F6$  of  $F1$  and  $F3$  is between them. For the subjects, the swelling and un-swelling aspect of  $F2$  and  $F4$  is more important, in contrast to the *direction of change* of  $F1$  and  $F3$  (Figs. 8 and 9).

A huge variety of different styles for each shape is given (see Fig. 10). This is emphasized in Fig. 8 by having nearly a consistent transition in the syntactic order from symbolic (the very left) and iconic (the very right). No real clusters of unique visual descriptions appeared. More drastic is that the



**Fig. 10** Overview of all drawings syntactically ordered for each shape



triangular shapes are usually perceived as rounded, which leads to confusion with the circle (compare Fig. 11).

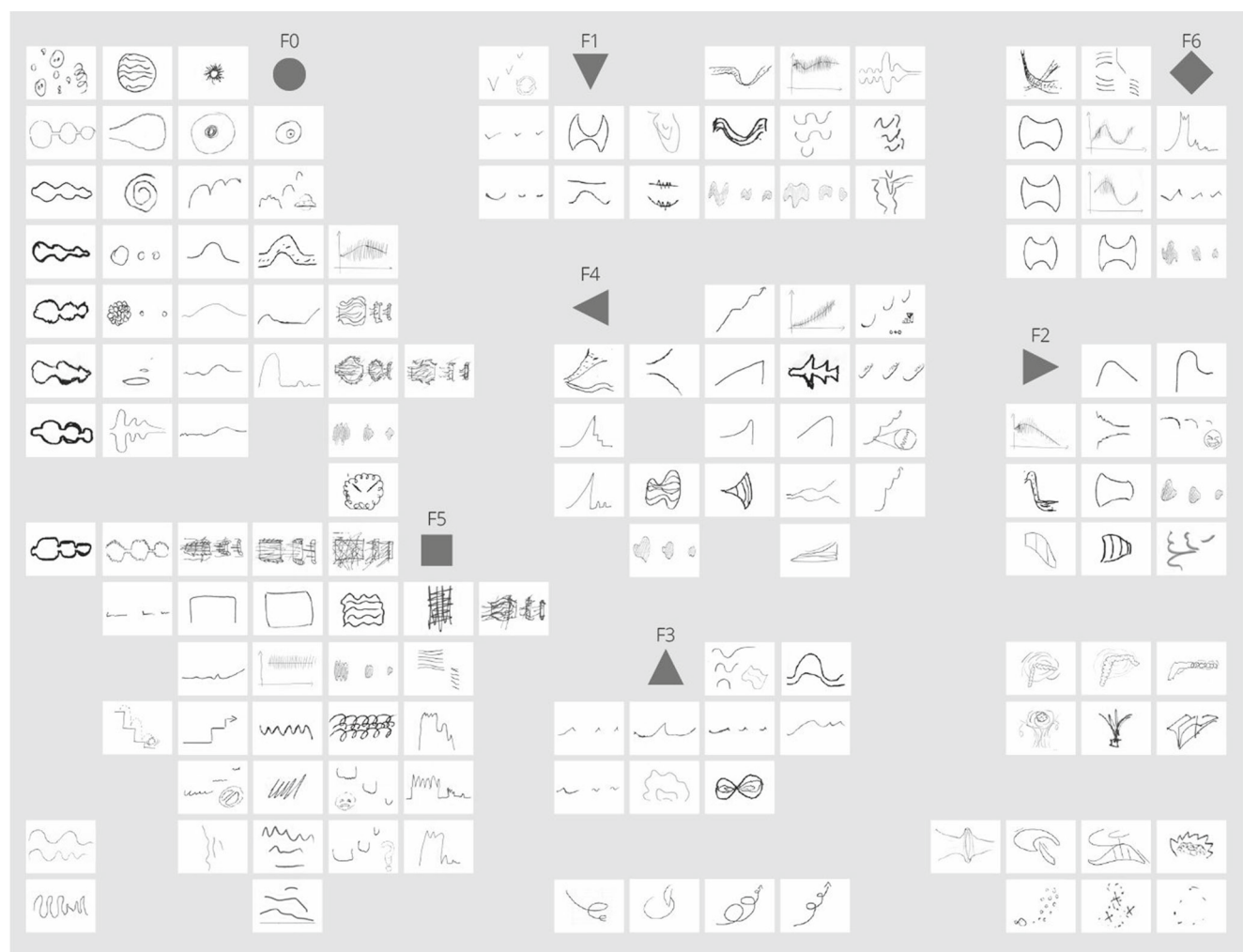
## 4.2 Topic of observation

Here, we only considered the visual shape for a resynthesized sound. But a visual design for audio means a lot more: shapes, sounds, transformations, manipulations, effects, and coloration. More in depth, what often forgotten is that additional visualizations, next to the spectrogram the spectral editing is working in, supports visually the understanding of the frequency domain (compare Fig. 9). These are topics for coherent visual audio design as well. Therefore, not only the shape or more generally the data for designing should be examined but also the entire interface and the way of interaction is important for overall coherency.

## 4.3 Interface for spectral editing

The spectral editing app used here for the resynthesis has multiple visualizations surrounding the spectrogram for more detailed insights and was tested with a brief user study [10]. The user study indicated that this spectral editing app is creative, clear, predictable, and easy to use with potential for improvement. For a joyful user experience, the interface matters and is influencing the creativity of visual audio design.

Spectral editing means not only editing existing magnitudes but also adding new magnitudes by strokes and shapes. Drawings can also be used for query-by-sketch searches for audible content (c.f. [5, 13]). Hereby, the drawing itself does not have to be the sketch to be queried (iconically) but an interpretation of the sketch can also be



**Fig. 11** A naive association of the drawings to the visually most suitable shape

used (indexically). Drawings are leading to symbols for an interpretation to the iconic spectrals. By using symbols, the interaction becomes more perceivable iconic.

## 5 Conclusion

A study method for the coherence of visual design for sound was introduced, in which the subjects only hear the resulting sound. Our approach is that when the visual impression of what is only heard correlates with the actual visual sound design, there is coherence. Nevertheless, the results have a huge bandwidth of possible drawings for the different sounds, although a visual alphabet can be derived from the results to find a translation from the visual to the sound. For this, the combination of both directions, impression from the visual to the sound and impression from the sound to the visual, should be emphasized. In future a tool for automatic clustering of the result from drawing studies should be considered, to deal with even more participants, due to this very diverse and subjective topic.

**Acknowledgments** Open Access funding provided by Projekt DEAL. The students and colleagues are appreciated for participating in the study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Assmann P, Ballard W, Bornstein L, Paschall D (1994) Track-draw: a graphical interface for controlling the parameters of a speech synthesizer. *Behavior Research Methods Instruments, & Computers* 26(4):431–436. <https://doi.org/10.3758/BF03204661>
- Bogaards N, Röbel A (2005) An interface for analysis-driven sound processing. In: *Audio engineering society convention* 119
- Bogaards N, Roebel A, Rodet X (2004) Sound analysis and processing with AudioSculpt 2. In: *International computer music conference (ICMC) (Icmc)*, 1
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the bold effect. *Neuroimage* 14(2):427–438
- Cao Y, Wang H, Wang C, Li Z, Zhang L, Zhang L (2010) Mindfinder: interactive sketch-based image search on millions of images. In: *Proceedings of the 18th ACM international conference on multimedia, MM '10*. ACM, New York, pp 1605–1608. <https://doi.org/10.1145/1873951.1874299>
- Cobley P (2014) *Introducing semiotics: a graphic guide*. Icon Books Ltd
- Eckel G (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
- Engeln L, Groh R (2017) Audioflux : a proposal for interactive visual audio manipulation. In: Burghardt M, Wimmer R, Wolff C, Womser-Hacker C (eds) *Mensch und computer 2017 - workshopband*, september, Gesellschaft für Informatik e.V., Regensburg
- Engeln L, Groh R (2019) Cohearence – a qualitative user-(pre-)test on resynthesized shapes for coherent visual sound design
- Engeln L, Groh R (2019) VisualAudio-Design – towards a graphical Sounddesign. In: *Proceedings of the 22nd international conference on digital audio effects (DAFx-19)*. ACM, Birmingham
- Engeln L, Hanke M, Auerswald T, Kallenbach F, Groh R (2019) The earconizer - a tool for constructing hierarchical earcons. In: *Mensch und computer 2019 - workshopband*. Gesellschaft für informatik e.v., bonn. <https://doi.org/10.18420/muc2019-ws-436>
- Engeln L, Kammer D, Brandt L, Groh R (2018) Multi-touch enhanced visual audio-morphing. In: *Proceedings of the international conference on new interfaces for musical expression (in press)*, Blacksburg
- Flickner M, Sawhney H, Niblack W, Ashley J, Huang Q, Dom B, Gorkani M, Hafner J, Lee D, Petkovic D, Steele D, Yanker P (1995) Query by image and video content: the QBIC system. *Computer* 28(9):23–32. <https://doi.org/10.1109/2.410146>
- Frassinetti F, Bolognini N, Lдавас E (2002) Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res* 147(3):332–343
- Giannakis K (2001) Sound mosaics a graphical user interface for sound synthesis based on auditory-visual associations
- Giannakis K (2006) A comparative evaluation of auditory-visual mappings for sound visualisation. *Organised Sound* 11(3):297–307
- Giannakis K, Smith M (2001) Imaging soundscapes: identifying cognitive associations between auditory and visual dimensions. *Musical Imagery*, pp 161–179
- Houix O, Monache SD, Lachambre H, Bevilacqua F, Rocchesso D, Lemaître G (2016) Innovative tools for sound sketching combining vocalizations and gestures. In: *Proceedings of the audio mostly 2016, AM '16*. ACM, New York, pp 12–19. <https://doi.org/10.1145/2986416.2986442>
- Kelkar T, Jensenius A (2018) Analyzing free-hand sound-tracings of melodic phrases. *Appl Sci* 8(1):135
- Klingbeil M (2005) Software for spectral Analysis, Editing, and synthesis. In: *ICMC*
- Klingbeil M (2012) SPEAR: Sinusoidal Partial Editing Analysis and Resynthesis. retrieved from on Nov 12, 3
- Mannone M (2018) Introduction to gestural similarity in music. an application of category theory to the orchestra. *J Math Music* 12(2):63–87. <https://doi.org/10.1080/17459737.2018.1450902>
- Misra A, Cook PR, Wang G (2006) Tapestry: sound scene modeling by example. In: *SIGGRAPH Sketches*, p 177
- Peschke J, Berndt A (2017) The geometric oscillator: sound synthesis with cyclic shapes. In: *Audio Mostly 2017: 12th Conf. on Interaction with Sound - Augmented and Participatory Sound/Music Experiences*. Queen Mary university of london, ACM, London, UK

25. Pon A, Ichino J, Eagle D, Sharlin E, D'Alessandro N, Carpendale MST (2012) Vuzik: a painting graphic score interface for composing and control of sound generation. In: ICMC
26. Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, unisensory processing. *Curr Opin Neurobiol* 15(4):454–458
27. Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cogn Brain Res* 14(1):187–198
28. Schwarz D (2011) State of the art in sound texture synthesis. In: *Digital audio effects (DAFx)*, pp 221–232
29. Seago A (2013) A new interaction strategy for musical timbre design. Springer, London, London, pp 153–169
30. Stolfi A, Barthet M, Goródcy F., Deusany A, Iazzetta F (2017) Open band: audience creative participation using web audio synthesis
31. Terren M (2019) Wave – an introduction to waveshaping and wave-terrain synthesis
32. Thoret E, Aramaki M, Kronland-Martinet R, Velay JL, Ystad S (2014) From sound to shape: auditory perception of drawing movements. *J Exp Psychol Hum Percept Perform* 40(3): 983
33. Verplank B, Mathews M, Shaw R (2001) Scanned synthesis. *J Acoust Soc Am* 109(5):2400–2400
34. Watkins S, Shams L, Tanaka S, Haynes JD, Rees G (2006) Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage* 31(3):1247–1256

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.