EPiC Series in Technology

Volume 1, 2019, Pages 17–30

KLG 2017. klingt gut! 2017 – international Symposium on Sound

The Emotional Impact of Sound: A Short Theory of Film Sound Design

Thomas Görne

Hamburg University of Applied Sciences, Sound Lab, ZfD
thomas.goerne@haw-hamburg.de

Abstract

Following Zillmann's Mood Management Theory, a main reason why people are watching films is the drive to modify and regulate one's mood by means of media entertainment [1]. And film is in this sense an effective medium: Narration, acting, visual design and sound design altogether contribute to its emotional impact. Accordingly, a main objective of film sound design is the communication and triggering of emotion or mood.

The paper investigates film sound design from the viewpoints of human perception, psychology and communication science. A special focus is set on the semantics of sound, communicated by means of crossmodal metaphors and symbols, on attention guiding and inattentional deafness, on the diegesis and on image / sound relationships.

1 The Auditory Object

The smallest entity of auditory perception is the *auditory object*, a simplified and categorized interpretation of the complex data collected by the ear. In this process, the sensation of sound is very likely translated into a hypothesis of an object in space as the origin of this sound. As Heidegger stated: *"Much closer to us than all sensations are the things themselves. We hear the door shut in the house and never hear acoustical sensations or even mere sounds"*[1] [2, 3]. This is what Schaeffer and Chion called "causal listening" [4, 5]: Our perception identifies the *sound* with a hypothesis of its *source*. Consequently, the sound designer's task is not creating and shaping sound, but auditory objects, the entities perceived by the audience.

1.1 Crossmodal Metaphors, Linguistic Metaphors

The idea of sound as an object or "thing" is the key to the perception of sound in terms of its *thingness*. From the early days of Gestalt psychology, these metaphoric qualities of the perceived sound have been investigated. In the late 19th century, Carl Stumpf described the metaphoric *volume* of the auditory object [6]. Wolfgang Köhler performed the famous "Maluma / Takete" experiment in the 1920's, which connects a fake word with an abstract figure (see Fig. 1) – it turns out that "Maluma" is likely identified as the round figure, "Takete" rather as

[1] *"Viel näher als alle Empfindungen sind uns die Dinge selbst. Wir hören im Haus die Tür schlagen und hören niemals akustische Empfindungen oder auch nur bloße Geräusche."*

the edgy one [7]. As the words are meaningless, the connection must result from the virtually perceived *sound* of the words.

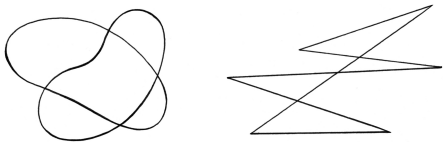

Figure 1: Maluma / Takete, after Wolfgang Köhler.

It is striking that we barely have a generic terminology to describe the sensation of sound in everyday language. Instead we are using metaphors referring to *other* senses, like high, low, deep, warm, cold, bright, dark, rough, smooth, big, small, soft, edgy, round, flat, sharp, dull, transparent, translucent, shimmering, sweet, colorful, etc. –: our auditory perception works mainly in terms of metaphoric visual and haptic properties of auditory objects.

Sources of these linguistic metaphors are the *crossmodal correspondences* of perception[2], the connections between the perception of stimuli in different sensual modalities. Thus the metaphoric descriptions of auditory perception may be called *crossmodal metaphors*.

The crossmodal correspondences have mainly been investigated since the late 1980's [8, 9, 10, 11, 12, 13, 14]; for an overview see [15]. For instance, it has been shown that in the presence of a high-pitched tone, a bright visual object presented on a video screen is detected faster [8]. Some experimentally proven crossmodal correspondences are listed in Table 1.

According to Spence [15] a crossmodal correspondence between two senses may be either caused by

- a "hard-wired", innate neural connection
  = structural correspondence,
- a neural connection by means of infant development, representing the most likely behavior of the physical environment
  = statistical correspondence, or
- a learnt connection determined by language
  = semantic correspondence.

From the sound designer's viewpoint, the first two types of correspondences (such as size, spatial height, brightness or shape) are most important, as they lead to similar crossmodal metaphors in different languages, and thus form an universal semantic code for auditory objects.

The crossmodal metaphors render auditory objects meaningful, e.g.: *a low pitched sound is a large, dark, round object with a position low in space.* This is the first step to understanding how one can communicate with the objects of a film sound design. For example, in film sound design weapons like swords, knifes or daggers are regularly complemented with semantically matching high pitched sharp sounds, even if these sounds are physically incorrect.

Furthermore, I propose that the *emotional impact* of the auditory object is created by matching linguistic metaphors associated with the crossmodal metaphors, as the whole world of *poetic* metaphors is evoked by association. For example, a low pitched and therewith large, dark, round and deep sound might trigger the connotations and associations of darkness, of something big, and of something below the surface, furthermore fueled by the psychological concept of the

[2]Terminology according to Spence; by different authors also referred to as *synaesthetic correspondences / associations* or *crossmodal equivalences / similarities / mappings* [15].

Stimulus	Corresponding Stimulus
pitch*	vertical position (*pitch / spatial height*)
pitch*	brightness (*higher = brighter*)
loudness	brightness (*louder = brighter*)
pitch*	shape (*higher = edgier / sharper*)
pitch*	size (*higher = smaller*)
pitch*	spatial frequency (*higher = finer structure*)
pitch*	movement (*rising pitch = upwards*)
pitch*	taste (*higher = sweeter*)
consonance	taste (*more consonant = sweeter*)

Table 1: Some crossmodal correspondences [15, 13]. – Note that there exist no everyday language crossmodal metaphors for the correspondences of loudness and brightness and of pitch and spatial frequency, as well as for the correspondences of aural and gustatory perception (even though we intuitively understand that a violin sounds sweeter than a viola).
* *"pitch" refers to the signal frequency as well as to the spectral weight of broadband or noise-like signals.*

dark and frightening world of the unconscious or "subconscious" *below* us (cf. Freud's structural and topographical model of personality *id / ego / superego*[3]), and we intuitively understand that some dark demonic power from below is awakened.

Naturally, the literature is full of descriptions of sound in poetic linguistic metaphors. An early example is given by Mersenne in his 1636 published *Harmonie Universelle*: He states that the *cornet à bouquin* (a nowadays almost forgotten instrument) *"sounds like a ray of sunlight, piercing the shadows or the darkness"*[4].

Metaphoric connections as extensions of the crossmodal metaphors, and probably even beyond of the auditory object's thingness – e.g. *a low pitched sound is power* – are not just intuitively evident, they can be proved in the experiment (for this example, see [17]). In the early 1980's, Marks investigated the applicability of poetic metaphors as a means to matching light and sound stimuli; the experiment showed surprisingly consistent results [18].

As the image of the ray of sunlight piercing the darkness or the image of a silver needle has the power to evoke a sound, a sound can conversely evoke the image of a ray of sunlight or of a silver needle. Crossmodal and linguistic metaphors render auditory objects meaningful: That's the reason why the deep, deep sound effect is impeccably effective, even though it's among the corniest sound design clichés.

1.2 Sound Symbols

"Fiery the angels rose, and as they rose deep thunder roll'd around their shores: / indignant burning with the fires of Orc." (William Blake: *America A Prophecy*, 1794). – Burning, feverish angels, thundering shores, the fires of Orc: as enigmatic Blake's poem is, as unmistakable is its emotional content, imparted by powerful *symbols* – images and sounds capable of communicating even beyond crossmodality and linguistic metaphors.

[3] The threat through one's own unconscious primitive, sexual, aggressive, instinctual drives "from below" is quite effective in dreams and myths, and is a very common motif in films. Examples are the basement scenes in *Silence of the Lambs* or *Fight Club*, of course accompanied by deep sounds. Žižek's explanations on the topic [16] are illuminative.

[4] *"Quant à la propriété du son qul rend, il est semblable à lclat dn rayon de soleil, qui paroist dans lmbre ou dans les ténèbres [...]"*

Following Jungian psychology, symbols are objects that are *"a priori meaningful"* (C.G. Jung). Populating myths and dreams, they are understood as externalized visual or acoustic manifestations of the powers concealed in the unconscious [19, 20, 21]. Thus they provide semantic and emotional loading of auditory objects in addition to the above discussed metaphors.

Sounds with symbolic powers are mainly nature sounds: *Wind* is the breath of spirit, *"an invisible presence, a Numen, brought to life by neither human expectation nor by arbitrary scheme"* [19]. The invisible ghostly presence is scary; consequently the wind is a cliché in horror movies.

Thunder is the expression of supreme, creative power and divine anger, it is the voice of the gods and the destroyer of spiritual enemies. In most cultures, the sound of the thunder has been mimicked with the *drum* for ritual purposes. The symbolism of the thunder adds to the above discussed example of the cliché-like yet effective deep sound effect.

Water, in form of the abysmal dark lake or the ocean, stands for the unconscious itself. According to Jung it is the *"living symbol of the dark psyche"* [19]. Water is the principal life-giver and can be a deathly threat, and it is – in film often in form of rain or breaking waves, to which the characters are exposed – symbol of surrendering to the forces of nature, i.e. being emotionally overwhelmed. Schafer states: *"Of all sounds, water, the original life element, has the most splendid symbolism"* [22].

Silence, with its ambiguous meaning referring either to denial or prohibition of communication (as the opposite of speaking), or to low sound level and low complexity (as the opposite of noise), can symbolize non-communication, isolation, purity, peaceful stillness or the calm before the storm, or it might signify an otherworldly place.

The sound of the *bell* is specific among the sound symbols, as it is a cultural artifact, and in this sense not an universal symbol. Nevertheless, through very different cultures, bells (and their cousins, the gongs) have been used as ritual devices, their sound marking important events in the society (see Fig. 2). Mirroring its ritual and social function, the sound of the bell is a powerful symbol of fate.

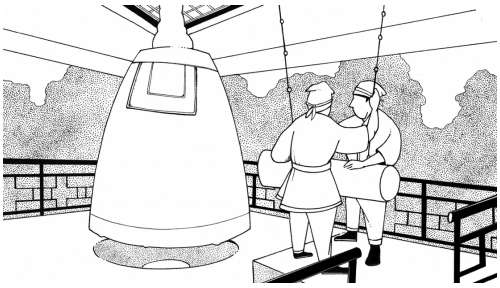

Figure 2: The "Divine Bell of King Seongdok", Korea, cast in 771 A.D., still in use for more than a millennium.

Animal voices are the last example of symbolic sounds: The dog barking in the distance, the voice of the animal of prey – mainly the big cat –, the happy songbird, symbol of life itself, or the crow as harbinger of death are potentially powerful symbols and, like other meaningful sounds, can be a cliché if used too obviously.

Besides specific sounds like this, virtually everything can become a symbol, dependent of one's personal experiences, as Jaffé pointed out [23].

Of course, sounds like wind, water, thunder, animal voices, the toll of the bell or even silence can just be what they are – the weather, an animal, a silent place. But in a sound design they

might unfold their symbolic power, particularly when they appear in unusual context – from the extradiegetic bells in *The Matrix* when a defenseless Neo seemingly gets murdered, to the roaring of the lion accompanying a violent boxing scene in *Raging Bull*.

1.3 Spatio-Temporal Congruency, Causal Thinking and Semantic Overload

The mechanism creating the crossmodal metaphors is similar to the magnetism that combines visual and auditory objects to an audiovisual object in case of spatio-temporal congruency (this is what Chion [5] calls "synchresis"). Both are following the likely behavior of the physical world. Evolutionarily they can be understood as mechanisms helping the individual to orientate quickly in an unknown complex environment.

From the constructivst viewpoint, the magnetism of synchronous visual and auditory objects is an expression of the inherited *causal thinking*: whenever two events occur simultaneously and roughly close to each other in space (hence *spatio-temporal congruency*), our perception compellingly creates a causal connection [24].

In film sound design, this mechanism not only allows creative Foley work (...like the sound of biting in a juicy apple, created by ripping apart duct tape in the hollow hands), but particularly makes *semantic overloading* of the auditory or audiovisual object possible by synchronizing / layering the image or sound with dissimilar meaningful sounds (e.g. symbolic sounds like animal voices, water, wind, thunder). A notorious example is the combination of the sportscar with the sound of a big cat or elephant to communicate the lively power of the wild animal (and of course poetic associations like freedom, or majestic ferocity) within the sound of the car motor. Due to the mechanism of causal thinking, the voice of the big cat or elephant will very likely not be perceived as an independent auditory object, but as belonging to or created by the car.

And the magnetism of spatio-temporal congruency allows comical effects, best known from the classic Warner or Disney cartoons: the tension between a visual cue and a seemingly causal connected "wrong" sound might be relieved in humor, when Donald or Goofy hitting the wall sounds like a crash cymbal or kettledrum, when Daffy tumbling from a tower is combined with the downwards glissando of a slide whistle.

1.4 The Ambiguous Object

A specific means of emotional communication is through the *ambiguity* of an auditory or audiovisual object. Ambiguity can lead to a feeling of *strangeness* or *alienness* and therewith cause strong emotional reactions, as it stirs the archetypal symbol of the *shadow* [19], the primal fear of one's own dark side, externalized as a projection (which is a psychological source of xenophobia). Piegler points out that with this projection of the evil and terrifying aspects of one's own personality into the outside world *"the own becomes the good and the alien becomes the evil"* [25].

A similar mechanism can be supposed as creating the emotional response to ambiguous or unknowable auditory objects (Flückiger calls them "unidentifiable sound objects" or "UKOs" [26]).

Ambiguity and alienness might be achieved by

- combining dissimilar objects (two dissimilar sounds, dissimilar sound and image), layering / overloading objects (see above),

- disfiguring or distorting the auditory object (e.g. by filtering, modulating, granular processing, ...),
- alienizing the auditory object through the context.

Examples for the latter are the rainforest sound in the shower in *Paranoid Park*, or the children's voice in *The Blair Witch Project*: its terrifying impact derives from the fact that it is heard in the middle of the night and in the middle of a dark forest (which is another Jungian symbol of the unconscious, similar to the dark deep lake).

It should be pointed out though that the ambiguous object doesn't necessarily lead to the experience of unease or fear. But at least it challenges the perception and tends to catch one's attention, as it defies the categorization as an object of the physical world.

1.5 Musical Structures

Further semantic and emotional communication might be achieved by means of *musical structures*, namely *rhythm* and *harmony*. Rhythmic structures are set up e.g. by clocks, footsteps, machinery and the like, but can also be created from other diegetic auditory objects. Melodic or harmonic structures ca be created from any tonal diegetic sound (see e.g. the windmill sound in the long initial scene of *Once Upon a Time in the West*, with its falling major third anticipating the iconic harmonica melody, with its slow and uneven rhythm emphasizing the passing time and helping to build tension). A sound design with dominant *non-diegetic* rhythmic or melodic / harmonic elements crosses the line between soundscapes and music: A filmic soundscape can even be *Musique Concrète*-like, as well as music can become sound design (see e.g. Bernard Herrmann's sharp, dissonant violins in the "shower scene" of *Psycho*).

The perception of rhythmic structures gains its impact from *temporal attention guiding*: The attention focuses on the point in time when the next occurrence of a sound is expected. Three similar events with even time spacing are enough to form such a structure [27, 28]. Furthermore, tension can be created with an event appearing out of its expected time. The *speed* of a rhythmic structure is perceived in relation to one's "spontaneous tempo", an inherent tempo, correlating with the average speed of one's own body movements e.g. when walking [28]. For young adults this is approximately a distance of 600 ms in between events [29], which equals some 100 bpm.

The perception of harmonic structures depends on the idea of a matching, fitting, pleasing harmonic sound as opposed to the tension and unease created by a dissonant sound. Interestingly, musical harmony is not a universal quality, but a cultural standard. Besides the "consonant" intervals of western / European music, there exist quite different cultural codes of pleasing, fitting intervals and therewith different definitions of "consonance" and "harmony" . Recent findings in Ethnomusicology suggest that even the idea of "pleasing consonance" and "annoying dissonance" is a cultural code [30]. Nevertheless, consonance and dissonance are powerful sound design tools. But even though in applied film sound design the western / European system of harmony is mostly regarded as a standard, taking into account the harmonic systems of other cultures might open the sound designer's horizon of communicating with sound; an overview on the topic is given e.g. in [31].

For a deeper look into the perception of musical structures see [28] and [32].

2 The Auditory Scene

As the perception identifies and categorizes parts of the information provided by the ears as discrete auditory objects, everything that's *not* categorized as an object is either perceived as part of an unspecific background, or not perceived at all.

2.1 Complexity

In the 1950's, Miller showed in a famous meta-study that the maximum number of objects that can be perceived consciously and simultaneously is typically 7 ±2 [33]. An implication for applied sound design is that a rather complex virtual auditory scene has a sufficient complexity with few discrete objects before a diffuse background ("atmo", "ambience" or "environmental sound"). The more complex a soundscape is built, the more likely the discrete auditory objects will melt into a diffuse conglomerate.

As of now there exists little experimental evidence of the maximum perceivable complexity of an auditory scene dependent of similarity and spatial distribution of auditory objects, but it seems like Miller's "magical number" tends to be even smaller when the objects are similar and located close to each other in space. For example, to record Foley footsteps for a mass scene, one needs barely more than three or four tracks, two or three of them to create sync sound for a few unique, peculiar characters, the last track to fill up with (then asynchronous) footsteps for the complex rest.

Although this is a well-known fact in practical sound design, even professionals get trapped in the "complexity pitfall" from time to time. As Walter Murch reports from his experiences of the sound design for *Apocalypse Now*: When he started to combine six premixes of the "Kilgore / helicopter attack scene", each composed of some 30 tracks, *"by some devilish alchemy they all melted into an unimpressive racket when they were played together"* [34]. It's not exactly devilish alchemy, it's Miller's number and the limited capacity of our conscious perception.

nota bene: The rise of object-based audio production, specifically for VR / 3D-Audio applications, shines a new light on the issue, as one cannot deal as easily with "object / background" categorization (i.e. few specific, macroscopic sound events before a background of stereophonic, quadraphonic or surround ambiences comprising a large number of microscopic sound events). A typical "hybrid approach" is the distribution of virtual loudspeakers for the playback of diffuse multi-channel ambiences within an object-based scene construction.

2.2 Physical vs. Perceptual Realism

Regarding a physically accurate construction of an acoustic scene as a prerequisite to auditory realism is a misapprehension of the term "realism". The perceived realism is not solely dependent on the physical accuracy. As we consciously perceive only a small part of a complex scene, the key to realism is not the sheer amount of detail. Instead, it is most important that the *essential* elements of the scene are convincing and persuasive.

Placing a few strong and convincing auditory objects before a rather blurry background anticipates and guides the auditory focus of attention, similar to how the depth of sharpness of the image anticipates and guides the visual focus of the eye. Thus we can differentiate between two approaches of auditory scene design:

- Guiding the audience's focus of attention through a complex scene by highlighting the important auditory objects. The goal of this approach is the *perceptual realism.*

- Constructing a detailed complex scene, where the audience can freely move the focus of attention. The goal of this approach is the *physical realism*.

If both are performed properly, it's likely that the naïve listener will not notice any difference, even though the soundscapes might be – physically, technically – quite different (cf. Miller's number).

The pitfall of the latter approach is that – although Robert Altman is known for trying this in his films – it hardly works for highly complex scenes in traditional technical formats like stereo or surround. The audience, instead of being able to focus the attention on parts of the scene, might then experience just a noisy chaos. But with the most advanced Audio VR technologies, physical realism might become a creative option and might be, following the idea of virtual realities, the adequate approach.

On the other hand, the approach of perceptual realism with highlighted auditory objects in reduced soundscapes is quite common among sound designers.

2.3 Gorillas at the Cocktail Party

The *attention* is a mechanism to control the conscious perception. By far the most information recorded by the sensory organs is processed automatically in the unconscious: That's the reason why we can drive a car and simultaneously have a discussion. While the attention focuses on the discussion, the complex process of conducting the machine is mainly performed automatically and unconsciously. In fact, most of our behavior is controlled by means of such unconscious automatic processes, and emotional reactions can be evoked automatically by unconsciously perceived stimuli likewise [35].

Cherry's famous investigation of the "Cocktail Party Effect" [36] showed not only the ability of the perception to focus on one source (in Cherry's experiment: dialogue) in a complex scene, but shine a light on the finding that one knows very little of what's happening *outside* one's focus of attention. Furthermore, it turns out that the focus of attention has a spatial dimension, i. e. the attention can focus on a specific location in space (e.g. on the soloist in front of the orchestra, but then one might miss what's going on in a different part of the hall), and it has a semantic dimension, i. e. the attention can focus on a specific part of a complex message (e.g. on the melody played by the soloist, but then one might miss what the celli are playing).

The equally famous "Invisible Gorilla" experiment by Simons and Chabris [37] investigated further the effect of *inattentional blindness*, the blindness to a stimulus in the middle of the visual field but outside the focus of attention, which has been reproduced in a similar way for auditory perception, proving the existence of *inattentional deafness* [38, 39].

It could be shown recently that not only the inattentional deafness is in effect with typical film sound designs, but also that a specific meaningful sound outside the focus of attention – so to say an "inaudible gorilla" in the soundscape – can emotionally intensify the audience's experience of a film scene [40].

The *orienting reflex* or *orienting response* forces the focus of attention to a specific object and direction in space, mainly through a stark contrast in the soundscape – an sudden loud sound, abrupt silence, an abrupt change in complexity – or by contradictory stimuli [41, 42]. It can be utilized in sound design to force the audiences attention and conscious perception to a certain part of the scene.

It can be concluded that potent auditory objects can be hidden in a soundscape if the audience's attention is fixated on a different part of the audiovisual message (e.g. the dialogue or the action). The sound symbols in *Matrix* and *Raging Bull* mentioned above in Section 1.2 are examples of such "inaudible gorillas".

3 Sound in Film

In film theory, the term *diegesis* refers to the relation of image or sound to the virtual world of the film [43]. A *diegetic* sound belongs to the world of the film, it exists as an acoustic signal in the virtual world. A non-diegetic sound can either be *metadiegetic*, i.e. subjective (e.g. an inner voice in the character's head, or an "semantically overloaded" or by other means alienized sound mimicking a character's perception), or it can be *extradiegetic*, i.e. completely outside the world of the film, audible only for the audience, not for the characters (e.g. typical film music or typical fancy "sound effects").

Possibly due to the history of film, starting with silent movies accompanied by live music, we are accustomed to feature film soundtracks with lots of non-diegetic elements, whereas the *picture* is typically completely diegetic (Lars von Trier's drama *Dogville* is one of the rare exceptions). In a typical production from the early days of "talkies", the only non-diegetic element was the music.

A rather modern approach of film sound design should not distinguish between soundscapes and music, but rather between diegetic and non-diegetic sounds, and it should take the relation of image and sound into account.

3.1 Sound Design with Respect to Diegesis

Though it seems that the diegesis of an auditory object is a binary criterion – a sound is diegetic or not –, it actually can be regarded as a continuously variable feature. As an auditory object can have an ambiguous identity and an ambiguous meaning, its diegesis can be ambiguous as well, tending to be either rather diegetic, or rather meta- or extradiegetic. And if of an audience of 100, 50 listeners judge a sound to be diegetic and 50 listeners judge it to be non-diegetic, then its diegesis obviously is "fluid".

A natural auditory object like a wind sound, the roar of a lion, the toll of the bell, a voice, or a footstep can be either diegetic or non-diegetic or in between, depending on the context. A synthetic or electronic sound is non-diegetic by definition, unless it is the diegetic sound of machinery, or unless it mimics a natural sound (but then it likely will sound "a bit strange", or not perfectly diegetic).

A sound design solely constructed from diegetic sounds may be called *documentaristic* or *naturalistic*, it is rather easy to perceive as it presents nothing but the sound of the world that is shown in the image. Its informational content – the information rate – is low due to redundancy of image and sound, and the soundscape is likely to be perceived as "objective". An impressive example is the sound design of the dogme95 drama *Festen* by Thomas Vinterberg, which remains completely on the documentaristic level.

A sound design from only meta- or extradiegetic sounds may be called *surreal* or *mystical*. As it presents mainly information that is not communicated within the image, the redundancy is low and the information rate is high, thus it is challenging and probably stressful to perceive, and it is likely to be perceived as "subjective". Compelling examples for sound design on the surreal, mystic level can be found e.g. in Darren Aronofsky's *Pi*.

Between the documentaristic and the surreal sound design I suggest to identify two more levels. One is the *quasi-documentaristic* or *attention-guiding* sound design, constructed from mainly diegetic sounds, with reduced soundscapes to guide the audience's attention to specific parts of the scene, and with more or less meaningful (metaphoric / symbolic) sounds; examples can be found as "hyperrealistic" soundscapes in virtually every modern major film production. As this attention-guiding sound design follows the idea of perceptual realism and shows

no obvious "sound effects", the audience might judge it as realistic as the documentaristic / naturalistic sound.

The other level in between is the *supernatural* sound design, constructed from lots of meaningful sounds with dominant meta- or extradiegetic elements – obvious "sound effects", so to say (examples can be found in David Lynch's *Mulholland Drive*, e.g. in the "diner scene"). Alternatively, the soundscape might be filled with ambiguous or unknowable sounds, auditory objects that are neither clearly diegetic nor clearly non-diegetic (see Ridley Scott's *Alien*, e.g. the "facehugger attack scene").

Fig. 3 depicts this 4-stage scheme of levels of intensity of film sound design, the levels depending on the diegesis, the grade of subjectification and the information rate: The more diegetic a sound is, the smaller its information rate, and the more it is perceived as a seemingly meaningless "objective representation" of a physical world. The less diegetic a sound is, the higher its information rate, and the more it is perceived as "subjective" and meaningful.

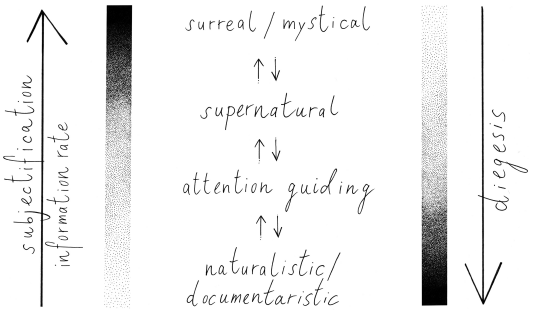

Figure 3: Four levels of sound design in dependence of diegesis (bottom: diegetic, top: non-diegetic), subjectification (bottom: objective, top: subjective) and information rate (bottom: low, top: high).

Typically, an arc of suspense can be created by means of proceeding from a lower level (i.e. the naturalistic or attention-guiding level) up to the supernatural or even surreal level, and back down again. An impressive example is the "murder in the restaurant" scene in *The Godfather*, which starts on the naturalistic level when the characters enter the restaurant, gets attention-guiding with few distinct auditory objects (wine bottle, eating) embedded in silence, and then, as the sound of passing trains – an initially diegetic, later metadiegetic, meaningful, metaphoric element – gets louder and louder and the dialogue disappears, it leads to the supernatural and finally surreal levels, culminating in the massive, shocking sound of the gunshots. With this climax at the end of the scene the metadiegetic sounds end abruptly. The soundscape drops back to the naturalistic level and the music starts.

3.2 Image / Sound Relations

Regarding image and sound as independent communication channels of the complex audiovisual channel "film", one can identify the image / sound relations as depending of the messages transmitted via the respective channels[5]. The scheme proposed here is based on Pauli's scheme of

[5]Of course, this model is simplified, as one might consider narration, mise en scène, acting, colors and lighting, montage, camera angles, etc, as well as diegetic sounds, non-diegetic sounds, dialogue, sound montage, etc, as separate channels, capable of communicating information independent of each other. But for the analysis of the role of sound with respect to the image, the simplified look is helpful. It might likewise be applied to multidimensional relationships.

music in silent movies [44], extended with the terms *incongruence*, *parallelization / complement* and *irritation*, as a sound design with diegetic and non-diegetic elements allows more complex relationships.

Assume that a filmic communication channel like image or sound is capable of communicating a message **A**, a message with a different but related content **B**, a completely independent message **X**, or no message **0**, i.e. an emotionally neutral content. ¬**A** is **A** negated. Then the relation of image and sound can be described as follows:

1. **Polarization**: image **0**, sound **A**.
2. **Paraphrase**: image **A**, sound **A**.
3. **Parallelization / Complement**: image **A**, sound **B**.
4. **Incongruence**: image **A**, sound ¬**A**.
5. **Irritation**: image **A**, sound **X**.

Polarization means that an emotionally neutral image is loaded by meaningful, emotionally effective sounds, e.g. ambiguous auditory objects, meta- or extradiegetic sounds. A compelling example is the above mentioned "diner scene" at the beginning of David Lynch's *Mulholland Drive*.

The *paraphrase* is redundant communication. The messages of image and sound are similar, which can easily be boring, trivial or obtrusive. On the other hand, the paraphrase might be a proper tool for the emotional climaxes of the narrative, like e.g. the final encounter of the lovers at sunrise in Joe Wright's *Pride and Prejudice*.

Parallelization and *complement* are common tools for condensation of the story; the most common form is the complementing diegetic offscreen sound. What can be told in the sound need not to be told in the image.

Challenging to the sound designer and provocative to the audience are *incongruence* with contradicting messages in image and sound, and *irritation* through an independent, asynchronous soundtrack, as they allow communication beyond the image and create tension.

The *incongruence* of image and sound is one of the most effective tools in film sound design. According to Schulz von Thun's model of human interaction, communicating incongruent, contradictory messages in different channels (i.e. verbal and nonverbal) is a communication disorder, causing a *cognitive dissonance* at the recipient [45, 46].

Translating this theory of human communication to media communication, one can conclude that contradictory messages in image and sound might likewise cause a cognitive dissonance, resulting in inner tension and unease in the audience (the paraphrasing sound would be *congruent* communication). A cliché in the horror and thriller genres is the incongruence of a dark and scary sound, prefiguring bad things to come, with a nice and positive image.

The last image / sound relation, the *irritation*, can be regarded as the "suspended chord" of image and sound. As a sound design completely independent of the image might even be perceived as a technical mistake, the only regular usage is the *overlapping* sound that anticipates the following scene. Like in western classical music a suspended, dissonant chord typically resolves to a major or minor chord after a few measures, the overlapping (asynchronous and extradiegetic) sound typically resolves to synchronous and diegetic as soon as the image changes. In the above mentioned "Kilgore / helicopter scene" in *Apocalypse Now*, the howling of a helicopter's turbine is used as an overlapping sound in the "Charlie don't surf" campfire scene before: an alien, irritating element, foreshadowing the early morning helicopter attack, a sound that resolves as a part of the diegetic soundscape a few seconds later.

4 Conclusions

From the earliest days of sound for film, theorists like Eisenstein and Balázs demanded "contrapuntal", autonomous sound for fictional films, in contrast to the "naïve sound" (Eisenstein) that has "no right to exist if there are no dramaturgical reasons" (Balász) [47, 48]. However, in many modern productions the sound is still regarded as a mere necessity, combining location sound and Foley, maybe with additional fancy "sound effects" – much like in the times of Eisenstein and Balázs. But understanding film sound design as a means of original, autonomous communication, the filmic soundscape should be designed with regard to its semantics and emotional content, its role in the filmic reality (diegesis), and its relation to the image.

Sound is perceived as auditory objects in time and space, with visual and haptic properties according to the *crossmodal metaphors*, gaining emotional impact through the connected *linguistic metaphors*, and with an additional impact if it can be understood as an "a priori meaningful" *symbol*. An important category of meaningful auditory objects is the *ambiguous object*, either with unknowable or ambiguous identity, or semantically overloaded by means of combining dissimilar sounds or dissimilar sound and image. Emotionally effective auditory objects might be hidden in the soundscape as "inaudible gorillas" by drawing the audience's attention to other elements of the scene. Additional impact can be achieved by means of rhythmic and harmonic structures.

The function of the auditory object within the filmic soundscape is given by its *diegesis*, the object being diegetic, meta- or extradiegetic, or with ambiguous diegesis. With respect to the momentarily diegesis of the majority of the elements of a soundscape, the impact of the sound design can be *naturalistic / documentaristic*, *attention guiding*, *supernatural* or *surreal / mystical*.

With respect to the image, the soundscape can be *polarising* or *paraphrasing*, it can add offscreen information by *complementing* the image, or it can be *incongruent* or *irritating*, causing a tension between image and sound.

More detailed descriptions are given in [49].

Acknowledgments

The illustrations in this paper have been created by Hannah Brückner, Hamburg, and are taken from [49].

References

[1] D. Zillmann: "Mood Management: Using Entertainment to Full Advantage", in: L. Donohew, H.E. Sypher & T. Higgins (ed.): *Communication, Social Cognition, and Affect*, Lawrence Erlbaum Associates, 1988

[2] M. Heidegger: *Der Ursprung des Kunstwerks*, Reclam 1960

[3] D. Ihde: *Listening and Voice: Phenomenologies of Sound*, State Univ. of New York Press, 2nd ed. 2007

[4] P. Schaeffer: *Musique Concrète*, Ernst Klett 1974

[5] M. Chion: *Audio-Vision: Sound on Screen*, Columbia University Press 1990

[6] S. Volke: "Carl Stumpf und die 'Raumsymbolik der Töne' ", in: G. Rötter & M. Ebeling (ed.): *Hören und Fühlen*, Peter Lang Verlag 2012

[7] W. Köhler: *Gestalt Psychology – An Introduction to New Concepts in Modern Psychology*, Liveright 1929/1947

[8] L. E. Marks: "On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination", *Journ. Exp. Psychol.: Human Perception and Performance* Vol. 13 (3), 1987

[9] L. E. Marks: "On Cross-Modal Similarity: The Perceptual Structure of Pitch, Loudness, and Brightness", *Journ. Exp. Psychol.: Human Perception and Performance* Vol. 15 (3), 1989

[10] A. Gallace & C. Spence: "Multisensory synesthetic interactions in the speeded classification of visual size", *Perception & Psychophysics* Vol. 68, 2006

[11] K. K. Evans & A. Treisman: "Natural cross-modal mappings between visual and auditory features", *Journ. Vision* Vol. 10 (1), 2010

[12] S. Shayan, O. Ozturk & M. A. Sicoli: "The Thickness of Pitch: Crossmodal Metaphors in Farsi, Turkish, and Zapotec", *Senses & Society* Vol. 6 (1), 2011

[13] K. Knöferle & C. Spence: "Crossmodal correspondences between sounds and tastes", *Psychonomic Bulletin & Review* Vol. 19 (6), 2012

[14] K. Knöferle, A. Woods, F. Käppler & C. Spence: "That Sounds Sweet: Using Cross-Modal Correspondences to Communicate Gustatory Attributes", *Psychol. & Marketing* Vol. 32 (1), 2015

[15] C. Spence: "Crossmodal correspondences: a tutorial review", *Attention Perception & Psychophysics* Vol. 73 (4), 2011

[16] S. Žižek: *The Pervert's Guide to Cinema* (DVD), dir.: Sophie Fiennes, Amoeba Film / Channel 4 / WDR 2006, Suhrkamp 2016

[17] D. Y. Hsu, L. Huang, L. F. Nordgren, D. D. Rucker & A. D. Galinsky: "The Music of Power: Perceptual and Behavioral Consequences of Powerful Music", *Social Psychological and Personality Science* Vol. 6 (1), 2015

[18] L. E. Marks: "Synesthetic Perception and Poetic Metaphor", *Journ. Exp. Psychol.: Human Perception and Performance* Vol. 8 (1), 1982

[19] C. G. Jung: "Über die Archetypen des kollektiven Unbewußten" (1934), in: *Archetypen*, dtv 1990

[20] C. G. Jung (ed.): "Der Mensch und seine Symbole" (1964), (orig. *Man and His Symbols*), Patmos 18th ed. 2012

[21] J. Campbell: *The Hero with a Thousand Faces*, Princeton University Press 1949 / Fontana Press 1993

[22] R. M. Schafer: *The Soundscape. Our Sonic Environment and the Tuning of the World*, Destiny Books 1977, 1994

[23] A. Jaffé: "Bildende Kunst als Symbol", in: Jung (ed.): *Der Mensch und seine Symbole*, Patmos 18th ed. 2012

[24] R. Riedl: "Die Folgen des Ursachendenkens", in: Watzlawick (ed.): *Die erfundene Wirklichkeit. Wie wissen wir, was wir zu wissen glauben? Beiträge zum Konstruktivismus*, Piper 1985, 10th ed. 2016

[25] T. Piegler (ed.): *Das Fremde im Film. Psychoanalytische Filminterpretationen*, Psychosozial-Verlag 2012

[26] B. Flückiger: *Sound Design. Die virtuelle Klangwelt des Films*, Schüren 2001

[27] C. L. Krumhansl: "Rhythm and Pitch in Music Cognition", *Psychological Bulletin* Vol. 126 (1), 2000

[28] S.-L. Tan, P. Pfordresher & R. Harré: *Psychology of Music. From Sound to Significance*, Psychology Press 2010

[29] S. Grondin: "Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions", *Attention, Perception, & Psychophysics* Vol. 72 (3), 2010

[30] J. H. McDermott, A. F. Schultz, E. A. Undurraga & R. A. Godoy: "Indifference to dissonance in native Amazonians reveals cultural variation in music perception", *Nature* advance online publication http://dx.doi.org/10.1038/nature18635, 2016

[31] T. Justus & J. J. Hutsler: "Fundamental Issues in the Evolutionary Psychology of Music: Assessing

Innateness and Domain Specificity", *Music Perception: An Interdisciplinary Journal* Vol. 23 (1), 2005

[32] W. J. Dowling & D. L. Harwood: *Music Cognition*, Academic Press 1986

[33] G. A. Miller: "The Magical Number Seven, Plus or Minus Two. Some Limits on Our Capacity for Processing Information", *Psychological Review* Vol. 63, 1956

[34] W. Murch: "Womb Tone", *Transom Review* Vol. 5 (1), Atlantic Public Media / Transom.org April 2005

[35] J. A. Bargh: "Automatic Information Processing: Implications for Communication and Affect", in: Donohew, Sypher & Higgins (ed.): *Communication, Social Cognition, and Affect*, Lawrence Erlbaum Assoc., 1988

[36] E. C. Cherry: "Some Experiments on the Recognition of Speech, with One and Two Ears", *Journ. Acoust. Soc. Am.* Vol. 25, 1953

[37] D. J. Simons & C. F. Chabris: "Gorillas in our midst: sustained inattentional blindness for dynamic events", *Perception* Vol. 28, 1999

[38] P. Dalton & N. Fraenkel: "Gorillas we have missed: Sustained inattentional deafness for dynamic events", *Cognition* Vol. 124 (3), 2012

[39] P. Dalton & R. W. Hughes: "Auditory attentional capture: implicit and explicit approaches", *Psychological Research* Vol. 78 (3), 2014

[40] A. Garcia Restrepo: "Die Wahrnehmung von nichtdiegetischen und abstrakten Klängen im Sounddesign", Bachelor Thesis, Hamburg University of Applied Sciences (HAW) 2017

[41] R. I. Schaefer, M. Goos & S. Goeppert: *Online-Lehrbuch Medizinische Psychologie*, Albert-Ludwigs-Universität Freiburg, http://www.medpsych.uni-freiburg.de/OL, 2000

[42] M. A. Wirtz (ed.): *Dorsch - Lexikon der Psychologie*, 17. Aufl. 2014

[43] . Souriau: "Die Struktur des filmischen Universums und das Vokabular der Filmologie" (orig. publ. 1951), *montage/av* Vol. 6 (2), 1997

[44] H. Pauli: *Filmmusik: Stummfilm*, Klett-Cotta 1981

[45] F. Schulz von Thun: *Miteinander Reden. Störungen und Klärungen. Allgemeine Psychologie der Kommunikation*, rororo 46th ed. 2008

[46] L. Festinger: *A Theory of Cognitive Dissonance*, Stanford University Press 1962 / reprint 1985

[47] S. M. Eisenstein, W. L. Pudowkin & G. W. Alexandrow: "Manifest zum Tonfilm" (1928), after Franz-Josef Albersmeier, *Texte zur Theorie des Films*, Reclam 5th ed. 2003

[48] B. Balázs: *Der Film. Wesen und Werden einer neuen Kunst*, Globus 1949

[49] T. Görne: *Sounddesign. Klang, Wahrnehmung, Emotion*, Hanser 2017