

How sound moves the body: Felt qualities of experience in gestural enactments of film sound

Linn-Marlen Rekittke¹, Dhana Wolf^{1,2}, Irene Mittelberg¹

¹ Human Technology Centre, RWTH Aachen University, Aachen, Germany

² Department of Psychiatry, Psychotherapy and Psychosomatics, Medical School,
RWTH Aachen University

rekittke@humtec.rwth-aachen.de, wolf@humtec.rwth-aachen.de, mittelberg@humtec.rwth-aachen.de

Abstract

The approach presented in this paper aims to contribute to an embodied account of sound experience and multimodal expression. Here, gestures are assumed to be dynamically constructed and adaptive to the sounds perceived and described by their producers. Based on film retellings by 11 German native speakers, gestural re/presentations of sound experience were analyzed in terms of spatial and temporal features with the help of a motion-capture system. When describing sounds stemming from a cinematic context, study participants conveyed sound qualities and re-experienced sensations through co-speech gestures. Our findings put into relief ways in which felt qualities of this experience are expressed through gestural enactment.

Index Terms: gesture, sound, embodiment, film

1. Introduction

This study investigates how impressions of film sound and the sensations they evoke in the spectators translate into communicative gestural movements during multimodal film retellings. We are particularly interested in how speaker-gesturers relate and enact their subjective sensorial experience of being exposed to a movie in which sounds take center stage. The effects sound may have on the perceiver's emotional state have been described by previous research [1]. Smalley [2], for instance, reminds us of the link between music and emotional responses. When listening to music we sense a psychological quality reflecting qualities of the sound; the effect of the sound is psychophysical information.

In a first approach, it is necessary to distinguish basic emotions such as fear and happiness [3] from less clearly defined sensations, which are the main interest of the present paper. According to various psychological approaches [4], emotions generally comprise the following components, which may, but do not have to occur simultaneously: 1) subjective, 2) physiological, 3) behavioral (gesture and facial expressions), and 4) cognitive (evaluation). The subjective component is defined as a private experience that can only be described through words. Contradicting the assumption of emotions as being 'in-coming' sensations, Zlatev [5] argues that such an account 1) lacks vividness and does not connect emotion and self-motion and 2) is insufficiently intersubjective. Affectivity motivates the body to move, but by the same token, it may be expressed through bodily movement [5]. As gesture research from several disciplines suggests [6], [7] and as we hope to show in this paper, gestures are not only part of the behavioral component, but tend to also unite cognitive, subjective and other dimensions of emotions.

Besides spoken language, body movements and facial expressions, spontaneous manual co-speech gestures may indeed express emotional and mental states [8]. The interrelation of the cause, experiencer and effect of feelings

and emotions is very complex [9]; gestures seem particularly apt at encompassing and expressing multiple dimensions due to their visio-spatial mediality and low codification. Besides the quality of the voice and intonation, spoken language alone can hardly convey subjective experiences holistically. Bodily semiotics, such as co-speech gestures, typically contribute additional dimensions. Furthermore, the subjective and physiological components may intermingle and as such are expressed – and also evoked – by gestures.

In addition, the cognitive component of emotions may be induced by manual gestures. Horst and colleagues [10] theoretically and empirically elaborate on how interlocutors share emotional experiences through their moving hands. This analysis of interaction leads to an embodied account of sense making through expressive movements. It challenges the assumption of pure conceptual and interpretative processes of understanding, suggesting dynamic sense-making processes through the perceiving, expressing and moving body.

Müller [8] has identified the three speech-functions of Bühler's organon model [11] interacting in co-speech gestures: representation, appeal, and expression. A gesture may represent an entity or event, impact the behavior of the interlocutor, and express feelings or affective stance. In this study the expressive function of gesture is focused upon, in the sense that the *how* rather than the *what* is foregrounded in the gesture [10]. In the following, special attention will be paid to the qualities of the moving hands and the manner of the movement (e.g. velocity, tension and torsion). In the sequences discussed below, the hands' movement qualities express sensations rather than emotions [12].

In line with embodied approaches [13], [14], this paper explores the perception of film sound and its interpretation in retellings. Scherer and colleagues [15] argue that films literally move the spectator. In addition to camera movement, temporal gestalt and montage rhythm, sound composition plays a crucial role in order to develop film images as movement patterns. Here, we are particularly interested in the connection between the experience of filmic sound events and their multimodal description: How are qualities of sound – and the affective responses provoked by them – expressed through the dynamic visual-spatial modality of co-speech gestures? Drawing on Mittelberg's [16] notion of the 'exbodyed mind', we argue that gestures have the potential and tendency to express what Johnson [14, p. 31] calls "felt qualities of our experience, understanding and thought" in an immediate fashion [see also 17, 18].

2. Embodied sound perception and strategies of listening to film sound

In addition to speech and other vocal expressions such as screaming, we experience a variety of sounds in our everyday life. Film producers work with sounds to evoke certain expectations and sensations, e.g. fear in a horror movie. For

instance, when sound becomes louder and shriller, we anticipate a terrifying event (see also [19]).

Huvenne [20] proposes four strategies of listening to sound in film, focusing on the role perception and imagination may play in these processes. The first strategy involves the emergence of a visual image when the sound causes the spectator to connect it to the memory of a formerly experienced scene. As a second strategy, a motor image is evoked when the spectator's body resonates with the sound's source movement. Here the emphasis on a pre-reflective way of listening is crucial: "Instead of a visual image being evoked, we listen first and foremost to the dynamic qualities of the movement and the physical efforts in the steps" [20, p. 140]. It's also important to note that sound experience is not an isolated event but always interwoven with a continuous flow of multimodal action in the film as a whole. Thus, Huvennes' [20] third listening strategy implies that the temporal composition of sound is related to the temporal composition of the film fragments. Here, the sound becomes meaningful as it connects with the possible reaction to it, e.g. the sound may be interpreted as pleasant or unpleasant. Based on a certain sound we expect a certain event to happen and thus an emotional affect may arise. Scherer and colleagues [15] highlight the temporal dimension of emerging affect in film perception. Especially the change of a sound (e.g. calm turns to frightening sound) creates tension and thus a gestalt like temporal unit. When sound is recorded from the protagonist's point of view, the spectator co-experiences it. This double experience of sound defines the fourth strategy, i.e., when the spectator resonates not only with the sound, but also identifies with the protagonist of the film. Here, the recorded sound positions the spectator in relation to the sound as well as to the action or object causing it.

The present study does not investigate instances of bodily action as the source of sound. However, these listening strategies are particularly relevant when exploring the re-experiencing of film sound during multimodal narrations, as they recruit the viewer's active perception and involvement. We suggest that the description of film sound is rather a relived experience than a description of sound qualities. It is the qualities of sound that merge with the qualities of feelings.

3. Multimodal descriptions of sound impressions

When describing sound, one may generally refer to 1) the **source** (e.g. environmental/naturalistic sound produced by nature; instruments; human audible action; machines or technical devices); 2) the **type** of sound (music, noise; see Smalley [3]); 3) the **volume** (loud, quiet); as well as 4) the **physical** (e.g. emotional) **reaction** to it (e.g. frightening).

Previous research has shown that when describing music and sound events, people tend to utilize gestures [21], [22]. A study by Caramiaux and colleagues [22] presents a quantitative, multimodal analysis of movement and sound. The authors explored the relation between audio descriptors and movement features by focusing on spontaneous body movements of their study participants, while they were listening to recorded sound. Results suggest a high correlation between loudness of sound and position as well as sharpness of the sound and both the velocity and acceleration of movement. They conclude that a gestural expression and the sound it tries to capture can be seen as being intertwined instead of separate entities. In another study, Caramiaux and colleagues [21] asked people to listen to sounds and to perform gestures describing the sounds. She investigated the difference of causal (action-derived) and non-causal sounds,

whereby the causal sounds mainly induced gestures mimicking the sound-causing action. Non-causal sounds were rather represented by gestures tracing contours evoking specific acoustic features of a given sound. As we hope to show in this paper, in multimodal narrations qualities of sounds merge with felt qualities of meaning on the side of the spectator, thus motivating expressive, communicative movements of the hands and torso.

4. Research questions

Taking gesture as a starting point, this study explores the relationship between sound, emotion, and motion. It focuses on gestures produced in natural, uninstructed story retellings. Importantly, the sounds in question are contextually perceived within a story-environment and afterwards described from memory (see the third and fourth listening strategies according to Huvenne [20] discussed above).

Our main research question asks how sounds are described and enacted through co-speech gestures, and how sensations experienced when viewing the film are conveyed multimodally. For the purpose of the present paper, we focus on sound produced by non-bodily action: Small technical devices used in everyday life (e.g. coffee machine), heavy-duty machines (e.g. an electric saw), and natural environmental sound (e.g. the rustling of leaves).

In particular, we were interested in the following aspects: 1) How do gestures express sound when the source is not bodily movement but environmental sound produced, e.g. by plants or technical devices? 2) Do the observed gestures depict sound qualities as such, or rather express the impact the sound had on their bodies and emotional states? 3) To what extent may embodiment theory account for the observed bodily techniques of enacting/describing sounds?

As the retellings were recorded in a Motion-Capture Lab, precise capturing, visualization and comparison of the gestural motion features was possible. Similarly to Caramiaux and colleagues [22], we determined whether verbal descriptions of the volume of a given sound (e.g. quiet, loud) relate to measurable gestural features (e.g., torsion and velocity).

5. Study

The study was conducted in the Natural Media Lab¹ at RWTH Aachen University. 11 native-speakers of German (6 female) participated in the study. They were asked to watch a short movie and then to retell the story line and reflect on it while being recorded by video and infrared cameras using a motion-capture system.

5.1. Stimulus Movie

All participants watched the German short film 'The Archivist' by Michael Cherdchupan². The film is 15.44 minutes long and shows the daily routine of an archivist who listens to and archives tape recordings of various kinds of sounds in a dark archive. Day by day he meticulously analyzes the recordings and documents their properties in a catalogue. Being alone and separated from social life the sounds ripple his imagination. All sounds (e.g. leaves rustling or noise coming from a sawmill) were edited to create a unique sound experience while watching the film. In correspondence with how the protagonist seems to perceive the various sounds,

¹ <http://www.humtec.rwth-aachen.de/nmlab>

² <https://www.youtube.com/watch?v=IIIBRvLvHu1A>



Figure 1: Stills from the stimulus film “The Archivist”

they are conveyed with exaggerated loudness and distinctiveness. Close-ups of the protagonist as well as of the items and events causing loud and striking sounds create an intimate relation between the viewer and the protagonist (as described by Huevenne’s [20] fourth listening strategy). Some sounds are perceived before their source appears in the visual scene, e.g. the vibrating sound of a cellphone, thus giving the viewer time to anticipate the source and also possible effects.

5.2. Data recording

The narrations were captured with an HD video camera and an optical Motion-Capture system. The latter is an infrared camera system capturing – with the help of reflective markers – the position and movement of the hands in three dimensional space. It further allows the visualization of movement traces of gesturing hands (see Fig. 2 A and figures further below). Participants were asked to wear 42 reflective markers. For each hand we used 14 markers including one on the wrist. Additionally, markers were placed on the head, neck, shoulders, arms, and knees. During the narration task, the interviewer was sitting directly behind the video camera placed opposite the participants (see Fig. 2 B & C).

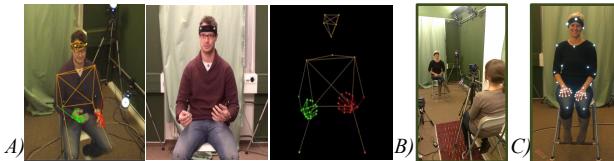


Figure 2: A) HD Video camera and Motion Capture Technology; B) Natural Media Lab; C) Marker Positions

5.3. Coding of the video data

The video data from the narration sessions were annotated and transcribed by a single coder using the software ELAN. All recorded gestures were segmented into gesture units, phrases and phases [23], which were coded independently for each hand. Only expressive gestures phases (i.e. strokes and holds; [24]) were considered, since they are taken to contribute most significantly to multimodal meaning construction. For each expressive phase, the four basic kinematic parameters – hand form, movement, position and orientation of the palm – were annotated [25], [26].

Speech was transcribed according to DuBois [27]. The transcription was added into the ELAN annotation. Transcripts were then searched for verbal terms referring to sound events. A description such as *you hear a cup clattering* (Germ. orig.: ‘Tasse hört man klappern’) was counted as one verbal reference. Then the ELAN annotations were searched for expressive gestural phases co-occurring with these verbal expressions. Here, we only analyzed those gestures that express loud or quiet sound.

5.4. Processing and visualization of the motion-capture data

The Motion-Capture system records the participants’ kinesphere and reduces the data into a three-dimensional matrix [28]. To describe the gestures’ geometry and dynamics we lean on Caramiaux and colleagues [22] by deriving multi-dimensional coordinates of torsion (maximal distance of

measured movement trace) and velocity (coordinates v_x , v_y , v_z , measured from the index finger) as two movement types from motion-capture data. The motion-capture data-based gesture visualization allows for a more objective account when comparing the gestures in terms of spatial and temporal features. The stick figures shown below are a way to map the body of the participants and to depict the trajectory of the moving hands through the blue lines. To reduce complexity, only the trajectory of the upper index finger marker was selected to trace the movement of the entire hand. The starting point of a trajectory is marked by the cones. The hands’ locations are documented using McNeill’s gesture space model [25] as well as information regarding the distance between the speaker’s hands and body. Hand configuration is described as clenched, bend or stretched¹. A statistic analysis of the data is not included in this paper; we will pursue a quantitative approach using these data in a follow-up study.

6. Results

In total, 11 narrations with 76 verbal references of sound experiences were analyzed. In their verbal descriptions, participants referred to either the source, the quality, the type, or the effect of the sound. They further express their individual involvement in speech and gesture. As previously observed by Caramiaux and colleagues [21], our study participants tend to imitate the action, thus the source of the sound, when verbally referring to sound produced by bodily action. For instance, the speaker moves her hand as if she was wiping a table. Only very few gestures represent gestural beats or point to the imagined location of the sound via pointing gesture. As mentioned earlier, our analysis focuses on those gestures that present the quality and effect of perceived film sound: either undefined surrounding sound or specific sound caused by, e.g., leaves or a sawmill. Special attention is paid to the description of loud versus low sound and its impact.

Gestures relating loud vs. low sound differed with respect to the hands’ positions and torsions. A weaker difference was observed in the hand configurations and measurements of velocity of the hand movements (mean of the velocity of the stroke). In terms of **position**, gestures expressing loud sound were either produced close to the speaker’s body or the body part perceiving the auditory information, e.g., the ears (see [18] on metonymic processes recruiting such body-part indices). Gestures that express a quiet and calming sound were either movements produced along a horizontal axis right above knee level or starting from close to the body, in the ‘center center’ of gesture space (‘center center’ is defined as the position of the gesture right in front of the upper body, between chest and navel of the gesturer; see [25]), going outward and downward. In terms of **torsion**, gestures expressing loud sounds tend to show several comparatively large twists, whereas gestures concerning low sound move only in one direction: either outwards producing a horizontal line, or away from the speaker downwards in a small slope. Comparing the **configuration** of the hands revealed only a slight difference in the sense that gestures expressing loud sound tend to be crooked, whereas gestures evoking low sound tend to be executed with either bent or straight hands. The **velocity** of the gestures was not strongly influenced by the sound quality. Here, the observations based on the video sequence suggested a difference in the senses that loud sounds

¹ We did not measure the muscle tension (for example using electrodes) and refrained from measuring the curvature (i.e. distance between the index, pinky and wrist) using the motion-capture data, as this would only make sense in an intra-subject comparison.

were portrayed by speedy gestures, whereas quiet sounds were expressed through gestures exhibiting less speed. The motion-capture data revealed that this perception was misleading, as the velocity of gestures expressing loud sound and those expressing low sound was not strongly different in the analyzed examples. In the following sub-sections, we will discuss a set of gestural examples in detail.

6.1. Gestures expressing loud sound

In the first example (see Fig. 3), the participant refers verbally to ‘*very loud surrounding sounds*’ (Germ. orig.: ‘*sehr laute Umgebungsgeräusche*’). On *very loud* he clenches both hands while lifting them from the ‘lower center’ (‘lower center’ is defined as the position in gesture space right above the legs of the seated gesturer) to the ‘center center’ of gesture space relatively close to his chest. His hands go quickly up and down (left hand: 808 mm/s; right hand: 832 mm/s) with a comparatively larger torsion (170 mm). His clenched hands express the strong effect the loud noise had on him.

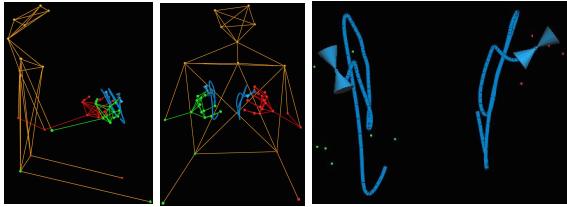


Figure 3: Gesture expressing loud surrounding sound

In the next example, the participant describes the impact of the *very loud* sound experience mentioned earlier on her own emotional state. She says the following: *there one notices how much sound may influence oneself, because* for example that* with the saw room, that immediately trou* or I myself was immediately troubled by it* (Germ. orig.: ‘*Da hat man mal gemerkt wie sehr Geräusche einen doch beinflussen können, weil* zum Beispiel das* mit dem Sägeraum da, das hat einen direkt beun* oder mich hat es direkt beunruhigt*’). Verbally she first mentions the general affect of the sound and then specifies her own reaction triggered by the sound. The first gesture co-occurs with the verbal phrase *sound may influence oneself* and is produced laterally on the right side with a small distance to her body (see Fig. 4). The dynamics of the gesture

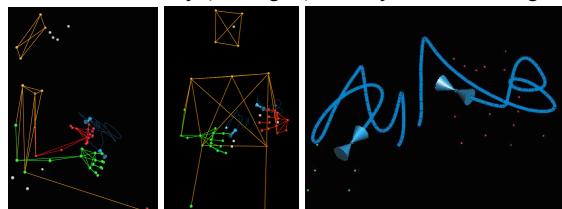


Figure 4: Gesture expressing the impact of the sound

is comparable to the examples described above, also showing a quick pace (left hand: 680 mm/s; right hand: 921 mm/s) and larger torsion (151mm). The second gesture (Fig. 5) occurs with the words *immediately trou* or I myself* and is

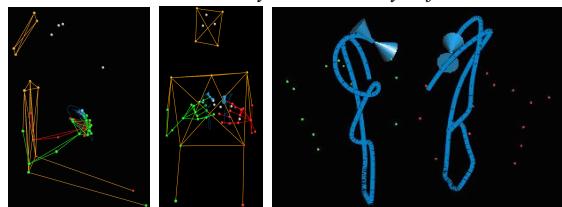


Figure 5: Gesture is performed close to speaker's body

made very close to her own body with her finger tips point at her chest, thus directing the attention to her own experience. The velocity of this gesture is much less (left hand: 49 mm/s; right hand 43 mm/s) with a larger torsion (161 mm). In the next example (Fig. 6), the participant verbally refers to noise becoming louder and louder. On the mention of *then one hears more and more white noise* (Germ. orig.: ‘*dann hört man immer mehr Rauschen*’) she lifts both hands towards her ears. With a very quick motion of the right hand (left hand: 271 mm/s; right hand: 1038 mm/s) she draws a cycle-like trace. The hands are bent as in the two examples above; here with the fingers pointing at the ear. Thus an indexical relation between the gesturing hand and the body part perceiving the sound is being created. Again the torsion is rather large (206 mm).

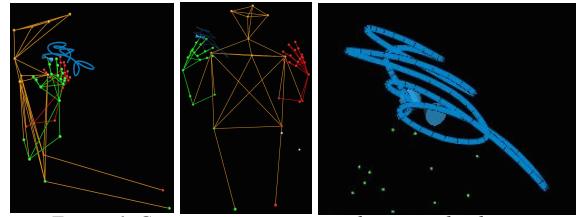


Figure 6: Gesture expressing noise becoming louder

6.2. Gestures expressing low/quiet sound

The participant compares the described loud sound with another scene in the forest, which was quiet. She connects the sound with its influence on her emotional state, as the sound calmed her down: *in the forest, for example, everything was very quiet, which immediately calmed one down* (Germ. orig.: ‘*Im Wald, die* zum Beispiel war alles sehr ruhig und das hat einen dann direkt beruhigt*’). This utterance is accompanied by two gestures. The first gesture in Figure 7 precedes the verbal reference to the sound in question – *very quiet* – and co-occurs with the verbal reference *everything*. The fact that the gesture precedes the verbal sound description may reflect that the experience was foremost physical and only secondarily verbal. From the captured motion trace, the arched trajectory of the gesture is clearly visible. The left hand starts from the ‘center

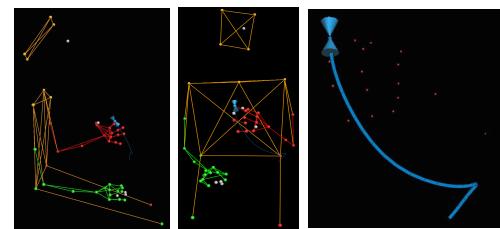


Figure 7: Gesture expressing quite surrounding

center’ of gesture space close to the gesturer’s body, moving out and downward in a well-formed arch. In comparison to the examples of the loud sound descriptions the torsion of this gesture is very small (11mm). The velocity (left hand: 527 mm/s; right hand: 105 mm/s) is slower than in the examples with participants describing loud sound; however, this difference is not strong. While moving the left hand is relaxed showing a bend configuration. The second gesture (Fig. 8) co-occurs with the verbal reference *immediately calmed down* with a low position of the hands and the wrist still leaning on the knees. The fingers of both hands are stretched forward and then draw a horizontal curve laterally outwards to both sides. Here, the velocity is rather high (left hand: 241 mm/s; right hand 907: mm/s), and the torsion small (58mm).

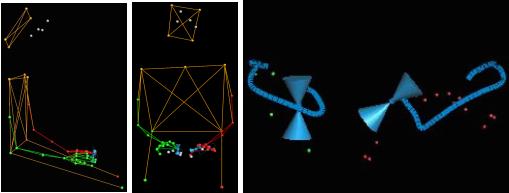


Figure 8: Gesture expressing the process of calming down

Interestingly, another participant (Fig. 9) is also leaning back with her shoulder while making a similar gesture evoking calmness. The trace of the shoulder is also depicted in Figure 9. Verbally she refers to her reaction to the sound *I noticed how I myself became very quiet to observe it* (Germ. orig.: ‘Ich merkte wie ich selber ganz leise wurde um das aufzunehmen’). Thus, the process of relaxing in response to the sound is described here. The concurrent gesture is similar to the gesture in Figure 8, which also draws a horizontal line. Here, the velocity is rather low (left hand: 497 mm/s; right hand 522 mm/s), with a rather small torsion (72mm).

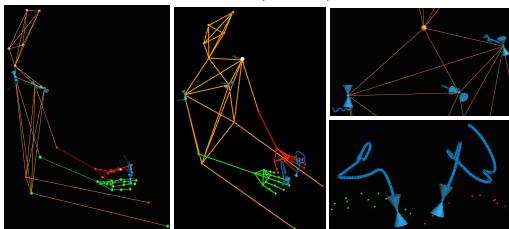


Figure 9: Participant leans back; gesture expressing her becoming quiet

7. Discussion

Using film sound as stimulus for analyzing expressive gestural movements in communicative acts has proven to be insightful, as the film retellings evoke sensorial qualities of experience as part of embodied listening strategies. While watching the movie, the participants experienced the sound. During the narrations they re-experience the sound and the effect it evoked in them. When reflecting on the filmic events they activate the physical dimensions of the listening and viewing experience. Our analysis has revealed that the volume of the sound and its physical reaction to it merge in accordance with Merleau-Ponty [29] who argued against an arbitrary relation between bodily experience and expression, suggesting a rather gestalt-like structure. The fact that the tension and relaxation of the speakers’ hands reflect the loudness or calmness of the respective sounds supports the understanding of sound perception as a comprehensive sensorimotor and psychological experience [2].

As suggested in the beginning of the paper, gestures are far more than a behavioral component of emotions and particularly apt at expressing subjective stance and feelings. In Figure 7 the gesture precedes the verbal sound description underlining the assumption that the re-experiencing of sound qualities is foremost physical and only secondarily verbal. Whereas speakers tend to describe the quality of a given sound verbally, their gestures tend to express the effect of the sound. According to our view [16], such ‘exbodyed’ instances of experiential essence indeed invoke “felt qualities of our experience” [14, p. 31]. With these types of co-speech gestures, the gesturer does not refer to, but rather senses the sound through the way her or his own body resonates with the sound experience.

The strongest finding is the positioning of the gestures as well as their torsion in relation to the sound experience. In accordance with Huevenne’s fourth listening strategy two

factors play a central role here: the participants 1) listened in an intersubjective way and 2) were situated listeners as the recorded sound invited them to react. Listening to a sound situates the listening body in space and time in relation to the sound, thus creating a certain nearness or distance to the sound. When perceiving a loud sound, one feels its closeness, vibration, and force. While retelling, the sound experience led to quick movements with bigger torsions located close to the body or to the ear as the body part perceiving it. Listening thus is an embodied practice of experiencing the sensorial environment including films as part of the semiotic world we live in.

The change in volume of sound gives the spectator the impression as if s/he was moving closer towards or further away from the source, thus leading to the embodied sensation the film evokes. The spectator experiences how his/her own body resonates with and is involved in the film. Thus, on the basis of bodily and affective sensing the participants grasp the story line and make sense of it cognitively and emotionally [30]. The participants found themselves attuned with the film’s sounds (see also [30] on this effect of images instead of sounds). When verbally referring to the process of ‘calming down’ the hands build up a tension, which gets released outward or downward. In a very engaged way, the participants empathize with the protagonist, e.g. through anticipating what a certain sound may forecast (see third listening strategy above). This kind of connection between the present moment and immediately following moments leads to a gestural (re-) presentation of the effect the sound evoked.

The sounds have an effect on their perceivers, and this effect is an affect in the sense of a physiological reaction that in turn motivates actual physical, communicative motion. In most of the examples, the sound talked about, movements and sensations merge in the multimodal description of the film scenes. The participants do not only refer to the sound quality itself but also express the influence the sound had on them and their bodily experience and memory of it. In summary, our analyses of multimodal film descriptions support (in line with Koch and colleagues [31]) an embodied account of perceived sound qualities, sound-induced sensations, as well as of their multimodal enactments.

As a next step, a quantitative analysis of the narrations is planned. A detailed observation and measurement of all gestures expressing sound experiences may give further insights into the active perception of film sound. We are interested in whether the tendencies identified in this study also holds for the entire data set. Furthermore, we plan to look at the multi-modal descriptions of additional sound events and sources.

Another question that arose from this study concerns the relation of metonymy and metaphor when describing and enacting (the influence of) sound qualities multimodally. The linguistic expression “the sound moved me” reflects, according to Conceptual Metaphor Theory (CMT), a conceptual metaphor, whereby the emotional response evoked by the perception of a sound is described in terms of ‘physical movement’ [32]. Foolen [9] argues that the figurative expression of emotion has itself an expressive motivation and suggests that the predominant form of figurative language about emotions is metonymic as an ‘effect for cause’ relation. In a similar vein, Mittelberg and Waugh [33], [18] have described how metaphoric gestural portrayals are rooted in fundamental metonymic processes. We thus will investigate whether multimodal descriptions of sound experience provide further evidence for metonymic/metaphoric relations between sound, sensation, motion, and meaning.

8. Conclusions

In this paper, we used auditory stimuli for multimodal descriptions to see how speakers describe with words and gestures sound impressions. Sound as such is invisible, but the sounds' sources and their effects on our physical and mental state can be described and perhaps imitated. Compared to traditional video analysis – relying on visual observation – Motion-Capture system allowed us to visualize and measure movement qualities more objectively. As these preliminary results show, there is a distribution of semantic and emotive dimensions across modalities. When describing from memory sounds stemming from a film context, participants conveyed sound qualities and the sensations they evoked through both speech and gestures. Our observations suggest that sensations caused by sound impressions link sound with motion and meaning physically. That is, it is not the sound that moves the person watching the movie, but the sensations and feelings evoked by the sound. Multimodal descriptions of experience and understanding as discussed in this paper may help the interlocutor not only to understand different aspects of the sound experience put into words, but also the speaker's sensorial and emotional experience while watching the film scenes. While the work cited in this paper no doubt represents a considerable step towards a fuller understanding of gesture's capacity to convey felt dimensions of experience, expression, and meaning more empirical research is needed to see how these various dimensions interact when people interact face-to-face or try to make sense of semiotic universes such as films and other multimodal forms of communication.

9. Acknowledgements

This study was funded by the Excellence Initiative of the German Federal and State Governments. We thank the members of the Natural Media Group for valuable input at various stages of this project, especially Bela Brenger and Ugur Töman for technical assistance, and the study participants for supporting our research.

10. References

- [1] Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A. and Turnbull, D., "Music emotion recognition: A state of the art review", in Proc. ISMIR, Utrecht, The Netherlands, 2010.
- [2] Smalley, D., "Spectromorphology: explaining sound-shapes", in Organised Sound 2, 107-126, 1997.
- [3] Ekman, P., "Methods for measuring facial action", in K. R. Scherer and P. Ekman [Eds], Handbook of Methods in Nonverbal Behaviour Research (pp. 45-90), Cambridge: Cambridge University Press, 1982.
- [4] Brandstätter, V., Schüler, J., Puca, R. M. and Lozo, L., "Motivation und Emotion. Allgemeine Psychologie für Bachelor", Berlin: Springer, 2013.
- [5] Zlatev, J., "Prologue", in Foolen, A. et al. [Ed], Moving Ourselves, Moving Others, 1-25, 2012.
- [6] Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. and Teßendorf, S. [Eds], "Body – Language – Communication: An International Handbook on Multimodality in Human Interaction", Handbooks of Linguistics and Communication Science (38.1). Berlin: Mouton de Gruyter, 2013.
- [7] Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D. and Bressem, J. [Eds], "Body – Language – Communication. An International Handbook on Multimodality in Human Interaction", Handbooks of Linguistics and Communication Science (38.2), Berlin/Boston: De Gruyter Mouton, 2014.
- [8] Müller, C., "Gestures as a medium of expression: The linguistic potential of gestures", in C. Müller et al. [Eds.], 202-217, 2013.
- [9] Foolen, A., "The relevance of emotion for language and linguistics", in A. Foolen et al. [Ed], Moving Ourselves, Moving Others, 349-368, 2012.
- [10] Horst, D., Boll, F., Schmitt, C. and Müller, C. "Gesture as interactive expressive movement: Inter-affectivity in face-to-face communication", in C. Müller et al. [Eds.], 2112-2125, 2014.
- [11] Bühler, K., "Sprachtheorie: Die Darstellungsfunktion der Sprache", Stuttgart, Fischer. First published [1934], 1982.
- [12] Damasio, A., "Looking for Spinoza: joy, sorrow and the feeling brain", New York, Harcourt Brace, 2003.
- [13] Gibbs, R. W. Jr., "Embodiment and cognitive science", Cambridge University Press, Cambridge, 2006.
- [14] Johnson, M., "The philosophical significance of image schemas", in B. Hampe [Ed.], From Perception to Meaning: Image Schemas in Cognitive Linguistics. Edited in cooperation with J. Grady. Berlin/New York, Mouton de Gruyter, 15–33, 2005.
- [15] Scherer, T., Greifenstein, S., and Kappelhoff, H., "Expressive movements in audio-visual media: Modulating affective experience", in Müller et al. [Eds.], 2080-2092, 2014.
- [16] Mittelberg, I., "The exbodyed mind: Cognitive-semiotic principles as motivating forces in gesture", in C. Müller et al. [Eds.], 750-779, 2013.
- [17] Mittelberg, I., "Balancing acts: Image schemas and force dynamics as experiential essence in pictures by Paul Klee and their gestural enactments", in M. Bokrent, B. Dancygier, J. Hinnell [Eds.], Language and the Creative Mind, 325-346, 2013.
- [18] Mittelberg, I. and Waugh, L. R., "Gestures and metonymy", in C. Müller et al. [Eds.], 1747-1766, 2014.
- [19] Boltz, M., "Musical soundtracks as a schematic influence on the cognitive processing of filmed events", Music Perception, 18: 427-455, 2001.
- [20] Huvenne, M., "Sound in film as an inner movement. Towards embodied listening strategies", in H. De Preester [Ed], Moving Imagination, John Benjamins, 133-148, 2013.
- [21] Caramiaux, B., Bevilacqua, F., Bianco, T., Schnell, N., Houix, O. and Susini, P., "The role of sound source perception in gestural sound description", ACM T. Appl. Percept., 11(1), 2014.
- [22] Caramiaux, B., Bevilacqua, F. and Schnell, N., "Towards a gesture-sound cross-modal analysis", in S. Kopp, I. Wachsmuth [Eds], Lecture Notes in Computer Science, Embodied Communication and Human-Computer Interaction, 8th International Gesture Workshop, 158-170, Heidelberg, 2010.
- [23] Kendon, A., "Gesture: Visible action as utterance", Cambridge u.a.: Cambridge University Press, 2004.
- [24] Kita, S., Gijn, I. van and Hulst, H. van der., "Movement phases in signs and co-speech gestures, and their transcription by human coders", in I. Wachsmuth and M. Fröhlich [Eds], Gesture and Sign Language in Human-Computer Interaction, 23–35, Berlin/ Heidelberg, Springer, 1998.
- [25] McNeill, D., "Gesture and Thought", Chicago, London, University of Chicago Press, 2005.
- [26] Bressem, J., "A linguistic perspective on the notation of form features in gestures", in C. Müller [Ed], 1079–1098, 2013.
- [27] DuBois, J. et al., "Outline of discourse transcription", in J. A. Edwards and M. D. Lampert [Eds], Talking Data: Transcription and Coding in Discourse Research. Mahwah, Lawrence Erlbaum Assoc., 45-87, 1993.
- [28] Priesters, M., Mittelberg, I., "Individual differences in speakers' gesture spaces: Multiangle views from a motion-capture study", Proceedings of the Tilburg Gesture Research Meeting (TiGeR), June 19-21, 2013.
- [29] Merleau-Ponty, M., "Phenomenology of Perception", London/New York: Routledge. First published [1945], 2005.
- [30] Schmitt, C. and Greifenstein, S., "Cinematic communication and embodiment", in Müller et al. [Eds.], 2061-2070, 2014.
- [31] Koch, S. C., Fuchs, T., Summa, M. and Müller, C [Eds.], "Body Memory, Metaphor and Movement", Amsterdam: John Benjamins: 2012.
- [32] Lakoff, G. and Johnson, M., "Philosophy in the flesh. The embodied mind and its challenge to western thought", Basic Books, New York, 1999.
- [33] Mittelberg, I. and Waugh, L. R., "Metonymy first, metaphor second: A cognitive-semiotic approach to multimodal figures of speech in co-speech gestures" in: C. Forceville and E. Urios-Aparisi [Eds], Multimodal Metaphor, Berlin/New York, Mouton de Gruyter, 229-356, 2009.