Notes on collaborative annotation with the AdA Filmontology

Anton Buzal, Yvonne Pfeilschifter, João Pedro Prado, Thomas Scherer, Jasper Stratil, Rebecca Zorko, and Jan-Hendrik Bakels.

BMBF Junior research group "Audio-visual rhetorics of affect," Freie Universität Berlin.

English version (a German version can be downloaded here)
Translated by Kevin Huthmann, Iris König

Version 1.1 (August 2023)

Table of Contents

Introduction

Within the framework of the junior research group "Audio-visual rhetorics of affect" (2016-2021), a systematic method for analyzing dynamic affective patterns of audiovisual images was (further) developed, which is based on the creation and visualization of annotations. The collaboration between film scholars from the Freie Universität Berlin and computer scientists from the Hasso-Plattner Institute at the University of Potsdam was funded by the German Federal Ministry of Education and Research (BMBF). The research subject were audiovisual images of the global financial crisis (2007–).

More information on the BMBF junior research group "Audio-visual rhetorics of affect" (AdA): https://www.ada.cinepoetics.fu-berlin.de/en/Ueber-Ada/index.html

AdA Toolkit: https://www.ada.cinepoetics.fu-berlin.de/en/ada-toolkit/index.html

> ***AdA Filmontology – Levels, Types, Values***: *An overview of the AdA project's systematic, film-analytic descriptive vocabulary.*

> ***Manual: Annotating with Advene and the AdA filmontology*** *– A detailed introduction to annotating and working with Advene annotation software and the AdA filmontology:*

Annotation Guide to the AdA Filmontology

This document provides practical guidance for annotating with the AdA filmontology, i.e. with a set of its basal annotation types. The annotation guide serves as an additional orientation aid in the practice-oriented handling of the individual basic tracks, their respective characteristics and requirements. They are based on the experiences and concrete agreements gathered during the annotation processes in the AdA project, i.e. they originate from a project context in which several analysts consistently and simultaneously described an extensive film corpus: Intersubjective agreement on specific analysis decisions are central here and are the result of coordination processes between expert annotators within the project. Some comments refer directly to the corpus of financial crisis films and news segments used by the project. Nevertheless, the insights gained may be useful and helpful for other contexts as well. The following pages are slightly revised internal protocols of the project-specific annotation process. This document is meant to accompany and build upon the *AdA filmontology* as well as the *annotation manual*. The links to both documents can be found at the top of this page and are referenced several times below.

General notes on annotating

Before annotating, first consider the following:

- Trust your own intuition and viewing experience.
- Base your work sessions on scenes (normally approx. 4-5 hours for one session).
- If you are uncertain about how to annotate individual cases, it can be helpful to mark them, e.g. with a keyword (*tbd* for to be discussed) that can be searched for in the annotation package, and to compare them with the viewing impressions and annotation experiences of others.
- Overlapping annotations are to be avoided, but as a tool deliberately used in special cases they are acceptable (e.g. overlapping music pieces).

Notes on the annotation of the individual basic tracks

Segmentation (Seg)

Segmentation (Seg) includes, for example, tracks that map argumentation units (*Seg | Argumentation Unit*) or scene boundaries (*Seg | Scene*). They divide the film or video into coherent sections.

Seg | Scene

If the video sequence to be annotated exceeds a few minutes, it should first be divided into scenes. For each scene, an annotation is created on the *Seg | Scene* track. The scene boundaries are not always clear and can therefore be discussed with another person who has also viewed the entire movie. After agreeing on all scene boundaries, the beginning and end of the annotations should be checked frame-by-frame.

- The scene segmentation can also serve as a division of work packages, since scenes are usually, though not necessarily, between 1 and 5 minutes long. The annotation of all basic tracks of a scene or a work package usually takes 3 to 8 hours, so 1 to max. 2 working days.

- Tested procedure: 2-3 annotators watch the whole film independently from each other and then divide it into scenes. Then a comparison or an editorial round takes place in order to determine the final scene boundaries – ideally, a person who has viewed the entire film but not yet divided it into scenes is involved.

- Scenes should be annotated as perceptual units, i.e. they should be oriented towards narrative units or compositional breaks; there are often audiovisual markers, such as the use of music or fades to black.

Language (Lg)

> Language (Lg) includes tracks in which spoken language is transcribed (*Lg | Dialogue Text*). Written language visible in the film can also be annotated (*Lg | Text Diegetic* and *Lg | Text Nondiegetic*).

Lg | Dialogue Text

If a subtitle file is available, the annotation of this track can be accelerated. Otherwise the track is filled with transcripts of the spoken dialogue. The annotation boundaries can be set freely and should correspond to speech units or phrases. That means, in case of a (clear) dialogue pause, the track remains free. Also, changes of the person speaking are marked by annotation boundaries.

The annotation lengths should roughly correspond to the audible duration of the dialogue in order to gain an impression of the speech rhythm.

- It is usually useful to first create the annotations based on the audible speech units and then, in a second pass, fill them with the transcription.

- <u>Note on annotation with Advene (see *annotation manual*)</u>: For longer dialogue units the *'Edit View'* is suitable, for proofreading the *'Transcript View'*.

- Annotate a new unit with every speaker change and at longer, distinctive pauses as well as at recognizable breaks (units of meaning; these can also be supported by the montage, for example).

- Large dialogue units make temporal localization of individual keywords more difficult during later searches of the annotations.

- Annotate clearly understandable song lyrics in Dialogue Text.

- Do not annotate vocals that are overlaid by dialogue or difficult to understand, and do not annotate "wohos" and "yeahs".

- Annotate sounds such as clapping, laughing, and whistling in Sound Gesture Dynamics, not in this track; if they are part of a linguistic utterance (e.g., laughing in a sentence), they can be markers of (dialogue) emotion.

- Mark slips of the tongue (e.g. in speeches of the Bundestag) with (sic) and possibly note what is meant in brackets; annotate only in very noticeable cases.

Image Composition (ImCo)

Image Composition (ImCo) includes tracks that concern aspects of arrangement and composition of the image, such as the field size (*ImCo | Field Size*) or the relative brightness of shots (*ImCo | Image Brightness*). It is not about completeness, but about capturing the most important visual impressions, always with respect to the standard of the sequence or video being annotated.

ImCo | Field Size

For this track, too, comparability within the video to be annotated is more important than absolute values, because there is an ambiguous area in the transition from one field size to the next. Dynamics of narrowing and widening in the course of the shots should be traceable through the annotation values. However, field size values should be used consistently within a video.

- Note on annotation with Advene (see *annotation manual*): This track is annotated throughout, so first copy all annotations from the *Seg | Shot* track to the *ImCo | Field Size* track and delete their contents using the *'Search/replace'* function. Then the *'Quickedit'* in combination with the *'Quickfill'* is recommended to quickly annotate the field sizes for each shot.

- Trace the dominant development if possible, since only one [TO] per annotation for annotating the continuous transition of two values is possible.

- Critically examine extreme values (extreme long shot or extreme closeup).
  → Use only for accents and thus relatively rarely.

- If two frames of reference for field size values are suitable, decide for one; annotate *neither* **only in exceptional cases.**

- Annotate *neither* for intertitles, animated graphics, etc.

- Check identical field sizes again at the end.
  → "Is annotation the same for the same reference object?"

- The same shot can be annotated differently if the focus (and with it the frame of reference) shifts within the image.

- When dealing with objects without reference size:
  → Determine field size via context (impression of proximity/distance in relation to shots before and after).

- When dealing with objects without reference size and context (e.g. newspaper articles):
  → Annotate as *neither* (e.g., document without reference to a human space).

- Annotating photos: Whether the field size refers to the content of the photo or the photo as an object, depends on the specific object character of the photo.
  → If the photo is recognizable as an object and located in a space (e.g. in a picture frame on a table), the field size refers to the photo as an object. If what is depicted in the photo dominates and can hardly be located as a photo per se, the field size refers to what is depicted in the photo.

- The annotation of the field size in situations with a newscaster and one or more additionally inserted pictures is based on the field size related to the newscaster.

When annotating the colours of a shot, the most prominent colour is named first. This is often also the colour that covers the largest area – but not necessarily. Thus, the most important colours of the whole shot should be named one after the other, in descending order.

- This track is annotated throughout, but not necessarily at shot level.
  → Choose a larger reference unit.

- Do not focus on individual images but characterize coherent colour spaces.

- Annotation is to be exclusively carried out with the following reduced colour selection:

red
royal blue
dark cyan
green
saddle brown
yellow
dark orange
deep pink
purple
antique white
white
grey
black

- The goal is the annotation of all image-defining colours, i.e. also possible accent colours, under the premise that a maximum of 4, in exceptional cases 5 colours per unit are annotated.

- The tendency to annotate rather small units – especially in documentaries – is no problem, as long as this does not cause additional work (smaller units can e.g. also represent a result in contrast to feature films).

- Additionally, annotate (*bw*) in brackets for black-and-white images.
  If a colour image within a shot becomes black-and-white, annotate the entire shot as (*bw*); images that are only black-and-white (e.g. filmed documents) are not automatically annotated as (*bw*), unless they are also decidedly shot in black-and-white.

- Usually skin colour is not annotated. If relevant, annotate with existing colour values.

- Recurring/continuous small graphic units that are superimposed on the image in post-production (lower third, TV logos, etc.) should be taken into account for colours if particularly noticeable.

All tracks, but especially the annotation types of the image composition category, are annotated with reference to a standard or average value of the video. Thus, a shot in a generally dark film can be annotated as "bright" even at an average brightness compared to other footage.

- This track is annotated throughout, but not necessarily at shot level.
  → Choose larger reference unit.

- Do not focus on individual frames, but characterize coherent brightness spaces.

- In order to maintain a relative uniformity of the brightness scale within a video, sample screenshots can be created for each brightness value for the respective film. These can be consulted in borderline cases and in case of uncertainties. They can also be used as a regular adjustment for one's own perception in order to compensate for different ways of perception due to different screens or a different brightness of the annotating person's surroundings.

ImCo | Image Intrinsic Movement

This track is intended to capture the perceived overall degree of movement of all things within the frame. Both the intensity of the movement and how much space it takes up in the frame are taken into account. Ultimately, it is the impression of movement that counts, with camera focus and direction of attention also playing a role – movement in the foreground is more noticeable than in a blurred background. *ImCo | Image Intrinsic Movement* is annotated throughout, the value 0 is not limited to freeze frames, but can also be used for very minimal movement of a shot, if it is only noticed on second viewing.

- It is annotated throughout (also 0 for static frames).

- The majority should be classified in 2 or 1. 3 being reserved only for extreme values (motion is often overestimated).

- For static images, which are interrupted by single movements (e.g. a car in the image background), annotate [*VS*], e.g. 0,[*VS*],1 (in this order).
  → Precondition: Two forces/motion dimensions are opposite to each other.

- Do not annotate images with constant minimal motion in the image (e.g. rustling of leaves) with [*VS*].

- Annotate with [*VS*] for clear background-foreground dynamics.

- If the motion changes in the course of the shot, work with [*TO*].

- Use [*TO*] for back and forth modulations (1→ 2→ 1→ 2 corresponds to 1,[*TO*],2). If the order is clear, annotate in that order.

- If both [*TO*] and [*VS*] can be annotated, opt for the stronger impression.

- Do not consider any fading in and out of the whole shot.

- Incorporate movement in the image as well as movement of the image in the motion impression, e.g. an incoming image.

- Brightness dynamics, e.g. fading in and out or timelapses, fall more into the Image *Brightness track* (mark with [*TO*]) but can also trigger or intensify an impression of movement: Take into account only if noticeable.

- Movements of text take place on a different level of motion than the movement 'in' the image (principle of layering influences the visual impression).
  → Images with text insertions are always annotated with [*VS*] (motion impression of the image,[*VS*],motion impression of the text; in this order).

- Split screen: In case of a difference in motion, annotate the values separated by commas (same as for *Field Size*, see section *How to handle split screens* at the end of the document).

- Difficulties in distinguishing image intrinsic movement from camera motion in digitally generated motion may arise.
  → Separate according to visual impression/intuition, in case of doubt annotate as image intrinsic movement.
  → In any case, ensure that all decisive movements are captured with *Dominant Movement Direction* and *Camera Movement Direction*.

  - Potentially set default values for the film similar to *Image Brightness.*

  - News broadcasting: *1,[VS],0* for the studio situation in front of the green screen if it is static.

This track is selectively annotated, preferably directly after the annotation of the *ImCo | Image Intrinsic Movement* track. The aim should be to note a clear direction of the movement (or in case of the value 'undirected': strong movements in different directions'').

- If *Image Intrinsic Movement* is attributed with value 0, no movement direction is annotated here.

- Do not annotate all movement directions within a shot, but only the dominant/especially emphasized movement direction.
  → If there is no dominant movement, nothing is annotated.

- Use *undirected* as a value for talking, gesturing people.
  However, only if there is a dominant movement!
  → In case of doubt, do not annotate (meaning: annotate a wildly gesticulating person, but do not annotate a calmly speaking person).

- The values *outward* and *inward* do *not* refer to On and Off, but denote movements of concentric expansion (e.g. water drop on water surface) or contraction (e.g. drain or black hole).

Camera (Cam)

The Camera category gathers tracks that describe camera-specific qualities of the footage, such as the camera angle (*ImCo | Camera Angle*) or the ratio of recording speed to playback speed (*ImCo | Recording/Playback Speed*). If a shot has several camera-specific properties, such as a pan to the left followed by a pan to the right, the values are named one after the other in chronological order. It is annotated according to importance and clarity: Very slight low angles or minimal camera movements are only annotated if the annotator perceives them as distinct and relevant.

Cam | Camera Movement Direction

In some cases, the annotator has to decide whether to annotate movements here or under *ImCo | Image Intrinsic Movement* – for example, in case of vehicular footage. If it is clearly the camera moving past the landscape, the movement is annotated as camera movement. However, if the impression is created that the image itself is moving, as in the case of a full frame of flowing water or drifting clouds, image-intrinsic movement is annotated, although steady camera movement can also create or enhance the impression. If the ambiguity of movement is deliberately played with, how that movement is annotated is up to the annotator.

- This track is to be annotated selectively.

- Minimal movements are not annotated. Capture dominant movements to reflect visual impression.

- Employ *undirected* as a value to annotate *shaky* camera movements.
  It is not always necessary to annotate *shaky* camera movements, only if they seem dominant in terms of movement impression.

- Split screen (see also section *How to handle split screens* at the end of the document): Annotate the values separated by commas in case of different movement (directions) (same as for *Field Size*).

- Complex camera movements may be described with many values, possibly *undirected* is also suitable.

- Annotate *circle*, if a circular movement is clearly discernable.

- Difficulties in distinguishing between image intrinsic movement and camera movement in digitally generated movements (see *Image Intrinsic Movement*) may arise.
  → Separate according to visual impression/intuition, in case of doubt annotate as image intrinsic movement.
  → In any case, ensure that all crucial movements are captured with *Dominant Movement Direction* and *Camera Movement Direction*.

Cam | Camera Angle

This track is annotated throughout and is closely related to the *Cam | Camera Position* track, which describes the relative height of the camera position in relation to the standard of the entire movie. This standard height is in many cases the eye level of the characters. The impression of a high camera angle can also be achieved by a strikingly high camera position, without a strong camera tilt. If this impression is clear, it should also be captured in *Cam | Camera Angle*, since the track *Cam | Camera Angle Vertical Positioning* is not one of the basic tracks.

- Intertitles, animations, graphics, writing, etc. are also annotated (perspective often corresponds to *straight-angle*).

- Use *straight-angle* also as container for slight high- or low-angles, so that clear *high-angles* and *low-angles* are not lost (in part changes are also captured via the annotation of the camera movement).

Cam | Recording/Playback Speed

A still (photo), a timelapse or a slow motion sequence often clearly stand out from the rest of the film, which establishes a standard. This standard is fundamentally based on the correspondence between recording and playback speed and thus roughly corresponds to an everyday visual impression. But there are also borderline cases where one has to take a closer look. Here, too, the visual impression is the deciding factor and not meticulous research into the production process.

- Form larger units up to entire scenes.

- Annotate *still* for still images (no signs of internal movement, e.g. photos), as well as static images where it is difficult to tell whether they are photos or camera recordings. Annotate (digitally animated) movement across these still images as camera movement or image intrinsic movement.

- Annotate intertitles, graphics, etc. according to visual impression.

- Split screen: In case of different recording/playback speeds, annotate the values separated by commas (same as for *Field Size*).

- News broadcasting: Annotate the studio situation as normal despite the still in the background/ next to the news anchor.

Montage (Montg)

> The annotations of montage-specific tracks (Montg) describe how the images relate to each other: At the shot level (*Montg | Montage Figure Micro*) or taking into account the whole film (*Montg | Found Footage*). Also the basic track Shot (*Montg | Shot*) – i.e. the division into individual shots – belongs to this category.

Montg | Shot

This track numerically lists all of the film's shots, thereby establishing the shot boundaries upon which many other tracks are built. A shot consists of a coherent movement of the image material, a break in this movement is caused by a cut. See also section '(automatic) shot segmentation' in the *Advene manual*.

- ATTENTION: This track should be annotated first, because it serves as the basis for many other tracks, including the *Seg | Scene* track.

- *Montg | Shot* is annotated by one person for the entire film and <u>not</u> at the scene level (see *Seg | Scene* section).

- After the shots are detected, they are renumbered once.

- Before annotating, check (using the 'Checker') if there are any overlapping annotations or annotations with empty content.

- Inaccuracies with the *automatic shot segmentation* by **one** frame or 42 milliseconds (i.e. the deviation between the shot validation view and the timeline) do not have to be corrected. This is an inaccuracy that varies from player to player.
  → If no greater accuracy is possible, take the less labor-intensive option.

- When working with news broadcasting, the *automatic shot segmentation* should cover the entire program and not just the segment to be annotated.

- Shot and scene boundaries have to match.

- In the case of transitions, the cut is made when the prior image has completely disappeared, i.e. at the first frame in which the prior image can no longer be seen.

- Split screens in which cuts in the individual image segments occur at different times are treated as one shot.

- Insert missing cuts in all relevant tracks.
  → After merging all scene packages, the shots are renumbered (see sections '*Merge packages*' and '*Renumber annotations*' in the *annotation manual*, link at the beginning of this document).

Montg | Shot Duration

This track displays the duration of each shot. Copying all annotations from the *Seg | Shot* to the track *Seg | shot duration* will suffice, the duration is automatically calculated by Advene (in the AdA-template).

- ATTENTION: The annotations in this track are created by one person for the entire film.

- Regarding split screens: While they are not recorded in the *Shot* track, cuts that do not take place simultaneously in all image segments of the split screen (e.g. only cut in the right screen) are annotated in the *Shot Duration* track (to record this dynamic).
→ See also the *Split screens* section at the end of the document.

Montg | Figure Macro

For the annotation of a *shot-reverse-shot* sequence, all shots, which show one character look and another look back, are combined into one annotation. The same applies to a *cross-cut*. If the sequences are interrupted by individual shots that do not fit into these montage figures, it is up to the annotator to decide whether the annotation is also interrupted.

- This track is selectively annotated in larger units that capture the entire montage figure.

- Use *montage* in the narrow sense for decidedly outstanding 'montage sequences' (that are e.g. not so strongly guided by voice).
→ The choice/change of music (song) can be an indicator, as can time- and/or space-ellipses.

- With two comprehensible possibilities of annotation, the exact boundaries of the unit are less important than recognizing the overall montage figure.

- In case the annotation of one montage figure overlays the annotation of another, there are different possibilities:
→ Only annotate the dominant montage figure (and align the unit with it) if the other is negligible.
→ Select the larger unit and annotate both values there (if the chronological localization of the smaller unit is negligible).
→ Create overlapping units for smaller units that should definitely be located within a larger unit (they are just as visible in the visualization, i.e. the data can still be evaluated).
→ Split the larger unit if it is interrupted by another montage figure.

- Annotate *framing* when a unit of staging is perceived (e.g. through an interview part at the beginning and end of a coherent explication).
Strong impressions of the **beginning** and **end** of a unit are crucial.
→ When *framing* is used, only annotate the first and last shot.

- *Circular* denotes a recurring pattern in a sequence of shots which not only contrasts two types of shots (ABABA...), but also creates more complex repeating patterns (e.g. ABCDABCDA).

- Only annotate *continuity* if there are noticeable continuity sequences, e.g. in the news, and not in every feature film scene that adheres to continuity editing.

- *Split screen*: Annotate a coherent image arrangement as one annotation. Create a new annotation for a new arrangement (e.g. when changing from vertical to horizontal split screen).

- Do not look for hidden patterns; this especially applies to *framing* and *circular.*

- Annotate *shot-reverse-shot* in documentaries too. Also annotate continuously if the line is crossed within the montage.

- Annotate *cross-cut* for a phone call between two people in different locations*.*

- Do not annotate graphics as a *sequence shot*. For a sequence shot, firstly, a choreographic element must be given and secondly, the movement must take place in three-dimensional space.

Montg | Found Footage

Images that are clearly created before the shooting or production of the film/video are labeled in the track *Montg | Found footage*. If the images appear as coherent material coming from the same source due to similarity in texture, colouration, or overall impression, they are annotated as a coherent annotation. Many films use only self-produced material, some consist entirely of found footage.

The decisive factor for annotating shots as found footage is the viewer's impression that existing image material has been used again in a new context, i.e. not every "found" material is automatically *found footage.* The idea is not to find out from which sources the material actually originates. An image that does not show its foreign origin is not annotated as *found footage* (e.g. images that are used in several news programs but that do not interrupt the flow of images and that are not explicitly marked as archive material). The decisive factor is not whether the material was shot by the filmmakers themselves, but an impression, a sense of media circulation.

Deviations in quality, format, etc. can serve as a guide since material that was not filmed by the filmmakers themselves is not automatically found footage, but material that was filmed by the filmmakers may also give the impression of found footage.

- Provide more detailed descriptions in brackets, e.g. *unspecified archive (police).*

- If a picture is evoking several associations, please record all of them, e.g. *news, unspecified archive (police).*

- If known, indicate the source of the found footage in brackets, e.g. *recorded session (C-SPAN)* or *news (ARD/Brennpunkt).*

- The categories *historical* and *contemporary* do not refer to what is depicted, but to the historical or contemporary marking on an aesthetic level (news from the past year can be marked as historical by a specific design such as the 4:3 format, for example).

- The category *news* implies a journalistic view (e.g. to create a certain sense of drama with the pictures) in contrast to the purely documentary claim of *recorded sessions.*

- *Recorded session* is also used for the recording of court hearings.

- For classification: The value *witness* was originally created with the documentation of police violence in mind.

- If a case cannot be classified at all (but gives the impression of found footage), annotate it as an *unspecified archive.*

- Annotate filmed television images as *found footage* if the material differs significantly from that of the film (even if it almost fills out the screen).

- Except for *archive photography*, all values are annotated for film material (audiovisual recordings) only.
  → Photos can only be annotated as *archive photography.*

Acoustics (As)

The annotation types of the Acoustics category (As) describe the auditory characteristics of the video sequence, such as the affective register of the spoken language (*As | Dialogue Emotion*) or the music (*As | Music Mood*).

AS | Volume

This track visualizes the dynamics of volume and can be automatically generated in Advene for an entire film, see *Advene manual*, 'Soundwave', link at the beginning of this document.

- ATTENTION: The annotations in this track are to be created by one person for the entire film.

AS | Dialogue Emotion

This track is based on the annotations from the *Lg | Dialogue text* track and assigns the speech units to different emotional values, which refer to a general affective dimension of the tone of voice and the way of speaking, rather than a psychology of characters. Larger units are possible if the dynamics stay consistent.

- Reserve striking, more "extreme" emotions (e.g. sad, joyful, angry) for clear cases only.
  → Annotate *neutral* if no other value applies or if the peak towards any applicable value is too weak.

- Noises such as sniffling, laughing, etc. can appear in the annotation of the dialogue as sad, happy, etc. and are then not annotated separately as a sound gesture.
  If there is, for example, a clear laugh, sob, etc., annotate this in *Sound Gesture Dynamics.*

- Generally assume a *neutral narrator.*

- Don't let the spoken content lead you too much.

- Don't be overly influenced by the multimodal impression.
  → Note: Use headphones and close your eyes if needed.

- Be careful with units that are too large: Rather think in terms of individual dialogue units (e.g. two characters with different emotions. When can you see which character? → Capture these dynamics).

- Annotate *disgusted* for: Disgust, contempt, disapproval, arrogance, revulsion, and annoyance (i.e. emotions that create distance).

- Annotate *suffering* for pain that can be physical or emotional (despair) in contrast to *sad* as a quieter form of sadness.

- Understand *relaxed* as a positive being-in-the-world, a not-being-exposed-to-any-threats.

- Do not define an extra neutral value for news reporting whose basic mood may already be *angry* or *sad* (e.g. speeches in the Bundestag), but continue to assume a general understanding of *neutral.*

AS | Sound Gesture Dynamic

This track is selectively annotated. It is not about capturing every perceptible noise but about highlighting accents visible in their acoustic properties. The slamming of a car door would thus be annotated as *explosive*, the siren of a passing fire truck would first be noted as *swell*, then as *subside*.

- Do not annotate sound segments in this track, only accents.

- Atmos are usually not gestures and are therefore not annotated in this track.

- Annotate noises such as *clapping*, *laughing*, *whistling* in this track if they represent an *accent* in the sound design. If they are part of a spoken utterance (e.g. laughter in a sentence), do not annotate in this track.
  → Do not annotate every sound, only those that act as an accent (foregrounding).

- If in doubt, do not annotate individual quiet sound elements or do mark them for discussion (*tbd*), however, if a pattern results from several individual quiet sound elements, make sure to annotate them (not to be confused with atmospheric noises).

- Consider the *element of the gesture* as a criterion for the annotation.

- Sounds that cannot be clearly identified as part of the music, but also not clearly identified as an external sound...
  → … are to be annotated if the sound contrasts with the music.
  → … are rather not to be annotated if the sound integrates well with the music.

- Procedure: Should one rather perceive the overall composition and e.g. annotate accents in the music or speech in this track, too, OR should one first differentiate mentally between voice, music and noise?
  → Only annotate what really fits into the track (no music, no voice).

- Note the different perceptibility that comes with different playback devices.
  → Annotate with headphones.

AS | Music Mood

This track assigns different emotional values to related musical units based on their affective qualities. A piece of music can be divided into several units with different emotional values, since the mood can develop, for example, from a slow, sad intro (*sad*) to a dark, tense main theme (*tense*).

- Also annotate diegetic (i.e. located in the world in the film) music in this track.

- If two pieces of music can be heard at the same time (e.g. diegetic and non-diegetic music), annotate them with two emotional values and [*VS*].

- Newscast: Annotate jingles as music; this should, however, not usually be part of a scene.

- Individual, significant moments, e.g. a drumbeat, are mapped on *Music Gesture.*

Bodily Expressivity (BodExp)

Bodily Expression (BodExp) includes tracks that describe the physical expression of the actors or depicted characters, such as gestures (*BodExp | Gestures Emotion*) and facial expressions (*BodExp | Facial Expressions Emotion*).

BodExp | Body Language Emotion

This track assigns certain emotional values to the gestures and facial expressions of the visible characters. It is selectively annotated only when characters are visible, at the shot level. Larger units are possible if the dynamics stay consistent.

- Be careful with units that are too large: Rather think of individual shots
  (e.g. two characters with different emotions). When can you see which character?
  → Capture these dynamics.

- If only hands are visible, only annotate if an emotion is clearly recognizable (e.g. an extended middle finger being depicted).

- Annotate *disgusted* for: Disgust, contempt, disapproval, arrogance, revulsion, and annoyance (emotions that create distance).

- Annotate *suffering* for pain that can be physical or emotional (despair) in contrast to *sad* as a quieter form of sadness.

- Understand *relaxed* as a positive being-in-the-world, a not-being-exposed-to-any-threats.

- Annotate tension as *neutral* and describe it in terms of intensity.

BodExp | Body Language Emotion

This track refers to the track units *BodEx | Body Language Emotion* and is intended to depict a progression of the intensity of the expression of body language.

- The value *neutral* is to be provided with an intensity, too.

- The intensity value is to be understood more as a basic peak of arousal and is not directly related to the annotated emotion.

Motif (Motf)

> The Motifs category describes the content of what is depicted, for example visible objects (*Motf | Image Content*) or the place of action/setting (*Motf | Setting*).

Motf | Setting

The place of action is recorded in this track. Its annotation consists of a system of certain values and a short free text.

- Logic for annotating the *Setting*, based on screenplay logic:
  - *Interior, exterior*
  - *Day, night, twilight*
  - *Nature, rural, suburban, urban*
  - Specific designation in brackets (e.g. *Mark's living room*)
  - Not every line must be annotated with a value (e.g. if the time of day is not recognizable from inside a room, none has to be annotated either).
    → Only annotate what is clearly evident; if nothing is clear, don't annotate
    → This means annotations such as *interior* or *interior,day* are possible.
- If a place or a time of day becomes clear through the context or if a corresponding impression is achieved through the staging, then definitely annotate it
(e.g. the time of day becomes clear via the reverse-shot or a club scene suggests *night* as the time of day).

- Annotate *nature, rural, suburban, urban* for *exterior* shots only.
In borderline cases (e.g. shots in the car) it is also possible to annotate *interior* and *exterior* together.

- Understand the setting in the sense of *space of action*, not in the sense of arrangement.

- Do not annotate shots that have no setting and cannot be classified by the context (e.g. intertitles, animations, newspaper articles, closeups of signs, etc.).

- If possible, form larger units.
→ If the closeup of an object cannot be clearly located, but the staging suggests that it is in the same setting as in the shot before and thereafter, the annotation of the setting should not be interrupted for the shot of this object.

- In the case of a *talking head*, if the setting cannot be assigned (e.g. due to a very close shot), mark it for discussion (*tbd*) and determine whether it is located elsewhere in relation to the film as a whole.
→ Often the person is already established in a room at an earlier point in time (e.g. through a long shot of the interview situation).

- Be sure to annotate important/known rooms (e.g. *White House, stock market*, ...).

- It can be difficult to distinguish between *court* and *hearing.*
→ Annotate *court* only if clearly recognizable, otherwise annotate *hearing room.*

- For clarity and consistency, rooms can be assigned to individual characters, e.g. *kitchen at Bobby's.*

Motf | Image Content

The annotations of this track relate to the most central "content" of the image. It is not about completeness of all visible objects, but about capturing what catches the eye.

- Annotate on shot level throughout – except for black-and-white images, do not annotate those.

- With *person, group* & *mass*, several values can be annotated, e.g. if a single person faces a group.

- If a figure is clearly in focus, definitely annotate *person* and, if necessary, add *group* or *mass.*

- A person who is ever so slightly caught in frame does not have to be annotated.

- If there are two people in the image, one of whom is clearly in focus (e.g. in the case of a shot-reverse-shot with over-the-shoulder shots), annotate *person,group.*

- If at first there is one person in the shot, then another (both people are not visible at the same time), still annotate *person,group.*

- Annotate individual people in the background as a *group*, even if they are standing alone.

- Annotate *location* for shots that e.g. establish a space (establishing shot), in which people play no role (long shot of Manhattan), or in which the space plays an important role (*location* can therefore also be annotated together with other values). If people are too small and therefore not expressive, they don't need to be annotated.

- *Writing*: Do not annotate any writing in the image (e.g. supermarket names in the image), but only when the reading mode is triggered.
  → Annotate writing that is prominent in the image (e.g. legible written lettering).

- Distinction between *writing* and *graphics*: *Graphics* do not have a strong written character, e.g. graphs, diagrams.

- Annotate this track throughout, even if it makes *object* a catch-all category.

- No annotation of black or white images (the track is therefore selective).

- A filmed screen showing a person is to be annotated as *person,object.*

- Annotate a close up of an object, even if e.g. hands can be seen, as *object* only*.*

- It is best to annotate bank and company names as *object,writing.*
  → Above all, do not annotate close ups as a *location*, since closer shots tend to emphasize the object character and do not establish a space.

- Signs can be annotated as *object,writing*.
  If there is no object character, only annotate as *writing*.
  If the writing is illegible and/or negligible, annotate as *object.*

- Newscast: Annotate lower thirds as *writing.*

- How are patterns and forms that have more of an aesthetic character annotated?
  → If no location is established, annotate as *object.*

How to handle split screens

- If there are different values depicted on individual screens, these respective values are all annotated within *one* annotation, *separated* by commas.

- The order of values to be observed is as follows:
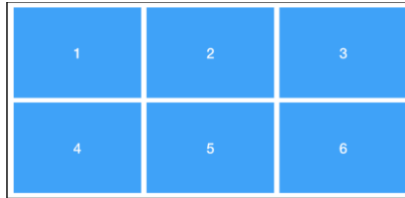
→ Generally, a logic from *left to right* and from *top to bottom* is to be followed.
→ For information on the specific handling of split screens in individual tracks, see the notes on the respective tracks in this document.
→ Generally, a shot with split screens is treated as 1 shot (separate values with commas), only if all split screens show a cut at the same time: shot boundary.