



DATA NOTE

# Portrayed emotions in the movie "Forrest Gump" [v1; ref status: indexed, <http://f1000r.es/55o>]

Annika Labs<sup>1\*</sup>, Theresa Reich<sup>1\*</sup>, Helene Schulenburg<sup>1\*</sup>, Manuel Boennen<sup>1</sup>, Gehrke Mareike<sup>1</sup>, Madleen Golz<sup>1</sup>, Benita Hartigs<sup>1</sup>, Nico Hoffmann<sup>1</sup>, Sebastian Keil<sup>1</sup>, Malú Perlow<sup>1</sup>, Anne Katrin Peukmann<sup>1</sup>, Lea Noell Rabe<sup>1</sup>, Franca-Rosa von Sobbe<sup>1</sup>, Michael Hanke<sup>1,2</sup>

<sup>1</sup>Psychoinformatics lab, Department of Psychology II, University of Magdeburg, Magdeburg, 39106, Germany

<sup>2</sup>Centre for Behavioral Brain Sciences, Magdeburg, Germany

\* Equal contributors

**v1** First published: 13 Apr 2015, 4:92 (doi: [10.12688/f1000research.6230.1](https://doi.org/10.12688/f1000research.6230.1))  
 Latest published: 13 Apr 2015, 4:92 (doi: [10.12688/f1000research.6230.1](https://doi.org/10.12688/f1000research.6230.1))

**Abstract**

Here we present a dataset with a description of portrayed emotions in the movie "Forrest Gump". A total of 12 observers independently annotated emotional episodes regarding their temporal location and duration. The nature of an emotion was characterized with basic attributes, such as arousal and valence, as well as explicit emotion category labels. In addition, annotations include a record of the perceptual evidence for the presence of an emotion. Two variants of the movie were annotated separately: 1) an audio-movie version of Forrest Gump that has been used as a stimulus for the acquisition of a large public functional brain imaging dataset, and 2) the original audio-visual movie. We present reliability and consistency estimates that suggest that both stimuli can be used to study visual and auditory emotion cue processing in real-life like situations. Raw annotations from all observers are publicly released in full in order to maximize their utility for a wide range of applications and possible future extensions. In addition, aggregate time series of inter-observer agreement with respect to particular attributes of portrayed emotions are provided to facilitate adoption of these data.

**Open Peer Review**

Referee Status:

	Invited Referees	
	1	2
<b>version 1</b> published 13 Apr 2015	 report	 report
<b>1</b>	Dylan Wagner, Ohio State University USA	
<b>2</b>	Ariel Rokem, Stanford University USA	

**Discuss this article**

Comments (0)

**Corresponding author:** Michael Hanke ([michael.hanke@gmail.com](mailto:michael.hanke@gmail.com))

**How to cite this article:** Labs A, Reich T, Schulenburg H *et al.* **Portrayed emotions in the movie "Forrest Gump" [v1; ref status: indexed, <http://f1000r.es/55o>]** *F1000Research* 2015, 4:92 (doi: [10.12688/f1000research.6230.1](https://doi.org/10.12688/f1000research.6230.1))

**Copyright:** © 2015 Labs A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

**Grant information:** This research was, in part, supported by the German Federal Ministry of Education and Research (BMBF) as part of a US-German collaboration in computational neuroscience (CRCNS; awarded to James Haxby, Peter Ramadge, and Michael Hanke), co-funded by the BMBF and the US National Science Foundation (BMBF 01GQ1112; NSF 1129855). Michael Hanke was supported by funds from the German federal state of Saxony-Anhalt, Project: Center for Behavioral Brain Sciences.

**Competing interests:** No competing interests were disclosed.

**First published:** 13 Apr 2015, 4:92 (doi: [10.12688/f1000research.6230.1](https://doi.org/10.12688/f1000research.6230.1))

**First indexed:** 27 Apr 2015, 4:92 (doi: [10.12688/f1000research.6230.1](https://doi.org/10.12688/f1000research.6230.1))

## Background

Recently, we have published a dataset with high-resolution functional magnetic resonance (fMRI) data of 20 participants listening to a two-hour audio-movie version of the Hollywood feature film “Forrest Gump”<sup>1</sup>. Using prolonged complex naturalistic stimuli, such as this one, is one approach to study cognitive processing in situations that resemble real-life contexts more closely than controlled laboratory settings typically employed to investigate individual cognitive functions in isolation<sup>2</sup>. However, multidimensional stimuli and the resulting lack of experimental control can make it harder to isolate the intervening parameters<sup>2</sup>. The goal of this study was to extend the information about this movie stimulus in order to enable further analysis of the already published brain imaging data as well as to investigate the utility of this particular stimulus for future studies and additional data acquisitions. To this end, we focused on a highly relevant aspect of social cognition that is, at the same time, difficult to infer from an audio-visual stimulus by means of computer algorithms: portrayed emotion (the expression of the emotional state of a movie character by an actor).

The presentation of movies (clips) to elicit emotional responses is an established component of emotion research, in particular with respect to more differentiated emotional states, such as remorse or pride, that go beyond primary emotions, like fear, in complexity and time-scale<sup>3</sup>. Emotion cues in movies can be manifold: facial expressions (e.g. smiling)<sup>4</sup>, verbal cues (e.g. swearing), voice characteristics (e.g. trembling voice)<sup>5</sup>, or even context cues that can be used to reason about emotional aspects of a situation based on abstract causal principles rather than direct perceptual cues<sup>6</sup>. Moreover, the emotional response to a stimulus in a real-life setting is further dependent on additional factors. For example, observing a facial expression of sadness may yield an emotional response of pity or satisfaction, depending on whether the person is a friend or a punished offender of social norms. A smiling face is not an expression of happiness when the larger context identifies it as a strategy to avoid unpleasant social interactions (a fake smile). Consequently, labeling portrayed emotions in feature films is a task that requires human observers to perform complex judgments.

Two groups of frameworks for systematic description of emotions are distinguished in the literature: a) schemes with discrete emotions (labels) and b) dimensional models<sup>7</sup>. Models using discrete emotion labels vary considerably in the number of differentiated emotion states. Many theories assume few basic, innate emotions<sup>8</sup>; others discriminate up to 36 affective categories<sup>9</sup>. Dimensional models, for example the circumplex model, locate different emotional states in a two-dimensional space, commonly using the axes arousal and valence<sup>10</sup>.

Current emotion assessment tools, like the Facial Action Coding System (FACS) or the Specific Affect Coding System (SPAFF)<sup>11</sup>, attempt to combine both approaches. However, they typically involve following complex instructions regarding the interpretation of facial expressions or other physical and cultural indicators of emotion (11, p. 281) and thus require an intensive training of

observers. While these tools provide reliable procedures for rating emotions, they are not very intuitive and are consequently only accessible to experts.

In this study, the primary goal was not to generate an objective labeling of portrayed emotions in the Forrest Gump movie. Congruent with our goal to enable further studies of already published brain imaging data, we rather aimed at producing a description of the emotional content of the Forrest Gump movie stimulus as it was likely perceived by the participants in our past and future brain imaging studies (potentially biased by age, native language, or education). Therefore, our approach was to collect data from multiple observers that stem from the same student population, using a simplified procedure that does not require extensive training.

Portrayed emotions were independently annotated for the audio-only version of “Forrest Gump” (used in Hanke M *et al.*<sup>1</sup>) and for the original audio-visual movie in order to obtain information about which aspects of portrayed emotions are congruent between both stimulus variants.

The resulting dataset of annotations of portrayed emotions combines a dimensional rating with a categorical labeling of emotions and a description of their associated perceptual evidence. In the following, we provide evidence that our procedure yielded a reliable description that can be used to segment the movie into episodes of portrayed emotions. Moreover, we show that the time course of inter-observer agreement (IOA) with respect to many aspects of portrayed emotion is a reliable measure of their perceptual ambiguity throughout the movie.

## Possible applications

In combination with the already publicly available brain imaging data, these annotations form a two-hour high resolution fMRI measurement for auditory emotion cue processing from 20 participants. With the addition of a future publication of brain imaging data recorded from a stimulation with the audio-visual movie, the full dataset will enable comparative studies investigating the processing of rich emotional stimuli via different sensory pathways. Moreover, these new annotations of portrayed emotions are another step towards a comprehensive description of this reproducible movie stimulus<sup>1</sup> and improve its general utility for independent studies on social cognition with a focus on the perception of emotions in real-life situations.

## Materials and methods

### Stimulus

The annotated stimulus was a slightly shortened version of the movie “Forrest Gump” (R. Zemeckis, Paramount Pictures, 1994). Two different variants of this movie were annotated separately. The first was a German audio-description as broadcast as an additional audio track for visually impaired listeners on Swiss public television (Koop, Michalski, Beckmann, Meinhardt & Benecke produced by Bayerischer Rundfunk, 2009). This audio-only stimulus was the same as the one used in Hanke M *et al.*<sup>1</sup>, and is largely identical to

the dubbed German soundtrack of the movie except for interspersed narrations by a male speaker who describes the visual content of a scene. These descriptions take place when there is no dialog, off-screen speech, or other relevant audio-content in the movie. The second variant is the audio-visual movie with the original dubbed German soundtrack (without additional narrations) and the exact same timing (and cut scenes) as the audio-only version (1, contains instructions on how to reproduce the stimulus from the DVD release).

### Observers

Annotations were created by the authors themselves as part of a practical course on scientific observation. A total of 12 students at the Otto-von-Guericke-University in Magdeburg, Germany participated in this effort and received course credit. Based on personal preference, students assigned themselves to one of two groups tasked with the annotation of portrayed emotion in either the audio-visual movie, or the audio-only version. The audiovisual group consisted of 9 students (all female), the audio-only group comprised 3 students (all male). This gender bias could have had an impact on the perceived emotions. None of the observers participated in the previous brain imaging study<sup>1</sup>.

### Procedure

Annotations were performed with the help of Advene<sup>12</sup>, a dedicated open-source video annotation tool that offers convenient navigation of structural units of the stimulus and provided uniform position information with subsecond precision. Each observer annotated all of 205 previously identified movie scenes in an individual random order to minimize “carry-over” effects and help observers focus on current indicators of portrayed emotions in contrast to more slowly changing characteristics, such as mood, or other biases introduced by the movie plot. Scenes were used as units for annotation, because transitions between them typically involve a change of location or fast progression of time that make persistence of the nature of portrayed emotions across scenes less likely. Observers used a scene annotation of the movie imported into Advene to aid navigation within the movie. They only had to click on the visual representation of a scene (a numbered box) in a time line display to start the playback at the start of a scene. Advene could also be configured to play a selected scene in a continuous loop. For composing the annotation content, observers used a separate spreadsheet application and not the data input features of Advene.

Observers were instructed to work alone in a setting free of distractions and to use headphones for optimal saliency of the stimulus. There was no strict guideline regarding the length of an annotation session, but observers were informed to stop working for any number of breaks whenever they could no longer guarantee a high level of attention for the task. Sessions were typically spread over multiple days over the course of three weeks. On average, observers reported that it required around 30 hours each to complete the annotation.

### Annotation content

Each observer collected annotations in a spreadsheet with columns for start time and end time of an emotion; a label of the movie

character portraying the emotion; variables for valence, arousal, and direction of the emotion as well as an optional emotion category label; and a list of identified indicators for the start and end of an emotion. Each of these variables are described in more detail in the remainder of this section. In addition, observers also recorded the index of the scene containing a respective emotion. This variable was only used for error detection (mismatch of emotion duration with the start and end time of a scene) and is not included in the published dataset.

**Start and end time** The start time identifies the onset of the first observed indicator of an emotion. Likewise, the end time corresponds to the offset of an emotion — defined as the time where no evidence for an emotion is present anymore. Observers were instructed to aggregate short segments of uniform emotional episodes into a single annotation. For example, if a character was happy and smiling throughout a long period in a scene, the episode would span all the time from the beginning to the end of the smile even if the character’s face was not visible the entire period, as long as no evidence of a change of emotion was found. All times are reported in seconds. Although Advene offers time stamps with the precision of the movie frame rate (25 frames per second), pre-tests revealed that the temporal accuracy of human observers for complex emotion indicators, like a developing facial expression, is much coarser. In order to reduce the chance of unrecoverable typos in the time stamps, observers were instructed to record timing information with second-precision only.

**Onset and offset cues** Onset indicators describe what kind of evidence for an emotion was detected in the movie stimulus. We distinguished facial expressions, gestures or body language, context information, verbal, and non-verbal audio cues. Observers could record multiple onset indicators. Despite the term “onset”, these indicators did not all have to be present at the very beginning of an emotional episode. For example, an extended period of sadness could start with a facial expression and later on add a congruent body language cue. In that case, the respective labels were aggregated into a list. In the audio-only stimulus, the narrator was frequently the only source of information regarding the emotional state of a character; thus, we included a dedicated category for this scenario.

In addition to onset indicators, observers also recorded the type of evidence for the end of an emotional episode. Four conditions were distinguished: 1) changing from one emotion to another 2) entering a neutral emotional state 3) a character leaving an ongoing scene with no further evidence for its emotional state 4) the end of a scene. [Table 1](#) describes all distinguished cue categories.

**Character label** This label identifies which movie character is portraying an emotion. Observers were provided with a list of character labels that was derived from an annotation of the movie’s dialog in order to achieve a consistent labeling. However, some characters portray emotions but have no lines of dialog, and, in other instances, observers chose inconsistent labels. During data curation, character labels were therefore manually consolidated into a set of 35 categories that uniformize labeling across observers. The full list of character labels is shown in [table 2](#) and was derived

**Table 1. Categories for cues indicating the onset or offset of an emotion.**

Label	Description
<i>Onset</i>	
AUDIO	Non-verbal audio cues, feelings expressed without words. For example: laughing, crying, wheezing, sighing
CONTEXT	Emotion portrayed through context information and not directly by a character. For example, a social rejection of an individual by a group without visible emotional response
GESTURE	Body language cues
FACE	Facial expressions
NARRATOR	Description of an emotion given by the narrator (only for audio stimulus)
VERBAL	Emotion expressed through words either as a direct description (e.g., a self-report) or by means of a particular choice of words (e.g., harsh language)
<i>Offset</i>	
EMOCHANGE	Present emotion is replaced by another, different emotion
EMOFADED	Emotion fades with a change to an emotion neutral state
CHARLEFT	Character leaves a scene (no further evidence for an emotion)
SCENEEND	Scene transition with a change of time or location indicates the end of an emotion

**Table 2. Consolidated character labels.** Any annotation of a portrayed emotion is associated with exactly one of these 36 labels. While main characters, their relatives, and famous characters have individual labels, all remaining characters in the movie are aggregated into generic categories that preserve information on gender, age group, and number.

Character label	Description	Character label	Description
<i>Main characters</i>			
<b>BUBBA</b>	Forrest's army pal	<b>FORRESTJR</b>	young Forrest Gump
<b>DAN</b>	Lt. Dan	<b>JENNY</b>	Forrest's love
<b>FORREST</b>	adult Forrest Gump		
<i>Special</i>			
<b>FORRESTVO</b>	voice-over narrator Forrest	<b>NARRATOR</b>	audio-description narrator
<i>Famous characters</i>			
<b>ABBIEHOFFMAN</b>	activist (lookalike actor)	<b>NEILARMSTRONG</b>	first man on the moon (orig. footage)
<b>BOBHOPE</b>	entertainer (orig. footage)	<b>PRESIDENTJOHNSON</b>	US-president (orig. footage)
<b>DICKCAVETT</b>	talk master (orig. footage)	<b>PRESIDENTKENNEDY</b>	US-president (orig. footage)
<b>ELVISPRESLEY</b>	The King (orig. footage and lookalike actor)	<b>PRESIDENTNIXON</b>	US-president (orig. footage)
<b>JOHNLENNON</b>	Beatle (orig. footage)	<b>ROBERTKENNEDY</b>	Brother of J. F. Kennedy (orig. footage)
<i>Relatives of main characters</i>			
<b>BUBBASGRGRANDMA</b>	Bubba's great-grandmother	<b>JENNYSGRANDMA</b>	Jenny's grandmother
<b>BUBBASGRGRGRANDMA</b>	Bubba's great-great-grand-mother	<b>MRSBLUE</b>	Bubbas's mother
<b>BUBBASYOUNGBROTHER</b>	Bubba's young brother	<b>MRSGUMP</b>	Forrest's mother
<b>JENNYSDAD</b>	Jenny's father		
<i>Generic character categories</i>			
<b>CHILDREN</b>	group of children	<b>MEN</b>	group of men
<b>GIRL</b>	individual girl	<b>OLDMEN</b>	group of old men
<b>BOY</b>	individual boy	<b>WOMAN</b>	individual woman
<b>OLDERBOYS</b>	group of older boys	<b>OLDWOMAN</b>	old woman
<b>MAN</b>	individual man	<b>WOMEN</b>	group of women
<b>OLDMAN</b>	individual old man	<b>CROWD</b>	group of people with mixed gender

using the following rules: a) main characters and famous/familiar characters receive individual labels as do the two narrator voices b) relatives of main characters are labeled according to their relation c) all other characters are consolidated into generic categories that preserve gender, age (children, adults, or seniors), and number (individual or group).

It is important to note that only emotions portrayed by one or more movie characters were recorded. This excludes other potentially emotion-inducing stimuli (e.g. music in the soundtrack) or inanimate stimuli (e.g., the depiction of a beautiful sunset or chirping birds).

**Valence, arousal, and direction** Binary variables were used to record the valence of an emotion (positive vs. negative), the state of arousal of the character portraying it (high vs. low), and whether the emotion is directed towards the character itself or another one (e.g. feeling pity for somebody). Together, these three indicator variables offer a coarse categorization of observed emotions in the spirit of dimensional models of emotion.

Arousal was considered a global indicator of the emotional status of a character, i.e., for any given time in the movie annotations, indicate either a high or low level of arousal for a character — but never both at the same time. Unlike arousal, valence and direction

indicators were considered non-exclusive, i.e., a character could show signs of both positive and negative emotions at the same time. Likewise, a character could simultaneously exhibit self-directed emotions and emotions directed towards others. In order to annotate such situations, observers had to create records for two or more portrayed emotions with overlapping timing.

**Emotion labels** Observers were instructed to assign one of 22 labels for discrete emotion categories. These categories were derived from a schema proposed by Ortony<sup>13</sup> and concrete definitions were developed based on the Specific Affect Coding System<sup>11</sup>. The complete list of emotions is shown in table 3. Pre-tests using a larger and more detailed set of emotion categories revealed weak agreement between observers for a forced categorization of emotional episodes. Based on these results, we decided to make this part of the annotation optional, and observers were instructed to assign an emotion label only if they saw a “perfect” match with any of the categories.

#### Quality control

We used an automated procedure to check annotation records of individual observer for errors or potential problems. Observers submitted their annotations in tabular form to a script that generated a list of error and warning messages. Using this feedback, observers double-checked their annotations as often as necessary until no

**Table 3. List of 22 emotion categories the observer could (optionally) use to further qualify the nature of a portrayed emotion.**

Label	Description
ADMIRATION	Appreciation of another person or object
ANGERRAGE	Response to injured respect or autonomy of a person by another one
COMPASSION	Concern for someone's suffering
CONTEMPT	Disdain of another person or object
DISAPPOINTMENT	When expectations were not met
FEAR	Triggered by a dangerous situation
FEARCONFIRMED	The confirmation of the prospect of an undesirable event
GLOATING	To be happy because another person suffers
GRATIFICATION	Triggered by a positive event for which one credits themselves
GRATITUDE	Triggered by a positive event for which one credits another person
HAPPINESS	Pleasant and joyful condition
HAPPYFOR	concern in happiness of someone else
HATE	Rejection of another person, intense dislike, or animosity
HOPE	Positive expectation for a person or special event in the future
LOVE	Affection for another person
PRIDE	Positive reflection of accomplishments and status
RELIEF	A worry/negative expectation that does not come to pass
REMORSE	Triggered by a negative event initiated by themselves
RESENTMENT	Response to perceived unfair treatment
SADNESS	Emotional pain after loss or misfortune
SATISFACTION	When expectations get fulfilled
SHAME	Dissatisfaction with one's own behavior, because norms are violated

objective errors were found and all warning messages were confirmed to be false positives. The following annotation properties were tested:

- start time had to precede end time (error)
- no character label (error)
- no onset cue label (error)
- missing arousal, valence, or direction value (error)
- conflicting simultaneous arousal annotation for the same character (error)
- emotion portrayed for longer than a minute (warning)
- annotation starts prior to a scene start or ends after a scene end (warning)
- unrecognized code/label for any variable (warning)
- no offset cue label (warning)

### Dataset content

The released data comprises four components: 1) raw annotations 2) aggregate inter-observer agreement (IOA) time series 3) emotion episode segmentations 4) the source code to generate all data derived from the raw annotations and all summary statistics presented in the Data Note.

**Raw annotations** In order to maximize the utility of the annotations, they are released in full rather than only as aggregate information. The annotations from each observer are available in an individual plain text file with comma-separated value (CSV) markup in the `raw` directory. This file format facilitates import and further processing in virtually any software application that supports data in tabular form. The file names indicate the annotated stimulus type (prefix `av` for the audio-visual movie and `ao` for the audio-only movie) as well as the observer identity.

Each file contains nine columns that correspond to the annotation properties described above. `start`, `end`) start and end time of an emotion episode reported as seconds from the start of the movie; `character`) a single character label from the set shown in [table 2](#) indicating the character portraying the emotion; `arousal`) state of arousal indicated with the labels `LOW` or `HIGH`; `valence`) emotional valence indicated with the label `POS` or `NEG`; `direction`) direction of the emotion indicated with the labels `SELF` or `OTHER`; `emotion`) a space-separated list of emotion labels from the set shown in [table 3](#); `oncue`) a space-separated list of onset indicators for a portrayed emotion from the set shown in [table 1](#); and `offcue`) a space-separated list of offset indicators from the set also shown in [table 1](#). With the exception of the columns `emotion` and `offcue`, all variables are considered mandatory and are present for all annotations. The remaining columns contain optional attributes and can be empty.

**Inter-observer agreement (IOA) time series** IOA time series reflect the consistency of observations at any particular time in the movie. The time series amplitude corresponds to the fraction of observers indicating the presence of a particular attribute of an emotion episode (interval  $[0, 1]$ ). Time series for the three bipolar

attributes *arousal*, *valence*, and *direction* were computed by subtracting the IOA with respect to the presence of the both extremes from each other. For example, the time series for *arousal* was computed by subtracting the time series of the IOA for low-arousal episodes from the time series of high-arousal episodes. This results in values from an interval of  $[-1, 1]$  — where  $-1$  corresponds to perfect agreement with respect to the presence of a low-arousal episode and  $+1$  indicates perfect agreement regarding high arousal.

IOA time series were computed for three different segment durations (i.e., agreement is considered when observers indicate the presence of a particular attribute anywhere in a segment): 1 s (temporal precision of the annotations), 2 s (corresponding sampling rate of the available brain imaging data), and the actual location and duration of individual shots in the movie (median duration  $\approx 5$  s). Individual time series for the three bipolar attributes, all 22 emotion categories, and the six emotion onset cues are available in the `timeseries` directory. The file names are encoded as follows: prefix `ioats`, followed by a label for the segment duration (e.g. `1s`), followed by the stimulus label (`av` or `ao`), and a character label (e.g. `forrest`, see [table 2](#)) or `allchar` for an aggregate time series across all characters. All files are in plain text format with CSV markup. The column headers indicate the corresponding emotion attribute.

**Emotion episode segmentation** Based on IOA time series, a segmentation of the movie into episodes of portrayed emotion was performed for each of the 35 character categories individually. IOA time series for 33 attributes (arousal (bipolar), valence and direction (separate for each state), 22 emotion category labels, and six onset cues) were binarized by thresholding with a minimum absolute IOA of 50%. An episode of portrayed emotion was defined as a time window with at least one super-threshold emotion attribute. The duration of an episode was determined by the number of consecutive movie segment without a change in the unique combination of super threshold attributes.

Emotion episodes for all characters are available in the `segmentation` directory in plain-text files with tab-separated value (TSV) markup. The file format follows the conventions required by the Advene software for “simple-structured content”, but are also compatible with generic spreadsheet applications. Each file contains three columns (without a header). The two columns contain the start and end time of an episode in seconds. The third column contains a space-separated list of key/value pairs. Five such properties are present for each episode: `char`) the label of the character portraying the emotion; `tags`) a comma-separated list of tags corresponding to all super-threshold emotion attributes; `arousal`, `val_pos`, and `val_neg`) these three properties each correspond to the median IOA value across the entire episode for arousal (bipolar) and positive and negative valence (each unipolar). The names of all tags typically correspond to the lower-case name of the attribute, with the exception of the three bipolar attributes where each extreme is coded with a dedicated label (arousal: `ha/la` (high/low), valence: `pos/neg`, direction: `self/other`).

Emotion episode segmentation was performed separately for the audio-visual and the audio-only movie—once using IOA time series sampled with a 1 s segment size and a second time

sampled with the actual duration and location of shots in the audio-visual movie. File names encode these conditions as follows: prefix `emotions` followed by the stimulus label (`av` or `ao`) followed by the segment size label (`1s` or `shots`).

**Source code** The full source code for all descriptive statistics and figures included in this paper is available in `descr_stats.py` (Python script). Moreover, this script also contains all functionality to generate the IOA time series as well as the emotion episode segmentation from the raw annotations in the data release. The required information on the timing of movie scenes and shots is contained in two additional CSV files: `movie_scenes.csv` and `movie_shots.csv`.

### Time stamp conversion

To be able to use the provided emotion annotations for an analysis of functional data (fMRI, cardiac trace, etc.) acquired using the previously described procedure<sup>1</sup>, timing information has to be converted because data were recorded in eight sessions using partially overlapping segments of the movie. **Table 4** specifies the location of these segments with respect to the timing of the unsegmented movie used for emotion annotation. Subtracting movie segment start times from annotation time stamps yields valid relative time stamps with respect to the time series of the available functional data.

### Dataset validation

**Characters portraying emotions** Annotations for a total of 34 and 27 different character categories were created for the audio-visual movie (AV) and the audio-only movie (AO) respectively. In the AV case, these comprise all possible characters except the narrator of the audio-description. There are 9 (AV) and 8 (AO) characters for which there is a 50% minimum inter-rater agreement on the presence of a portrayed emotion for at least five episodes of one second or longer throughout the entire movie (arbitrary thresholds). The number of emotion episodes for the AV/AO union of these characters sets is shown in **figure 1** (top). As expected, the majority of all portrayed emotions are associated with the main movie characters Forrest, Dan, and Jenny.

**Portrayed emotion categories** For the audio movie, 97.5% of all annotations include at least one of the 22 explicit emotion labels (see **table 3**), whereas only 68.4% of annotations for the audio-visual

movie contain such a label. This is an indication that emotion cues in the visual movie are more diverse and ambiguous in comparison to the audio-only stimulus. **Figure 1** (bottom) shows the frequency and display duration for individual emotions. For both stimulus types, the majority of all emotional episodes involve the five categories anger/rage, fear, happiness, love, and sadness. Some of our pre-determined emotion categories were not attributed to any emotion episodes in the movie, were used very infrequently, or were used only by a small subset of observers. Such emotions include resentment, gratification, fears confirmed, or satisfaction (see **table 5**). It remains unknown whether these emotions are not portrayed in the movie or if the respective categories were not appropriately defined.

**Emotion onset cues** **Figure 2** shows the frequency of individual emotion onset cues for all six distinguished cue types. When only considering annotations with a minimum inter-rater agreement of 50% with respect to the type of onset cue, the two dominating emotion cues are facial expressions (AV only) and verbal cues (for both stimulus types), followed by gestures and body language (AV only). **Figure 3** (bottom row) shows a rather uniform distribution of verbal emotion cues across the entire duration of the movie for both stimulus types. This suggests that a comparative analysis of verbal emotion cue processing during AV vs. AO stimulation, as suggested above, may be feasible.

#### Dataset 1. Annotations of portrayed emotions in the movie "Forrest Gump"

<http://dx.doi.org/10.5256/f1000research.6230.d45328>

See main body of text for detailed description of dataset.

### Observer agreement as a probabilistic indicator

It can be assumed that observers in this study have used individual intensity thresholds for determining the start, end, or quality of an emotional episode. Therefore, the IOA with respect to a particular aspect of portrayed emotions at any given time in the movie could be seen as a measure of the relative probability of perceiving an emotion feature. In order to assess the reliability of such a measure, we generated time series for the IOA with a sampling rate of 1 Hz, as described above, and evaluated their reliability and validity in terms of their correlation across observers, stimulus types, and with each other.

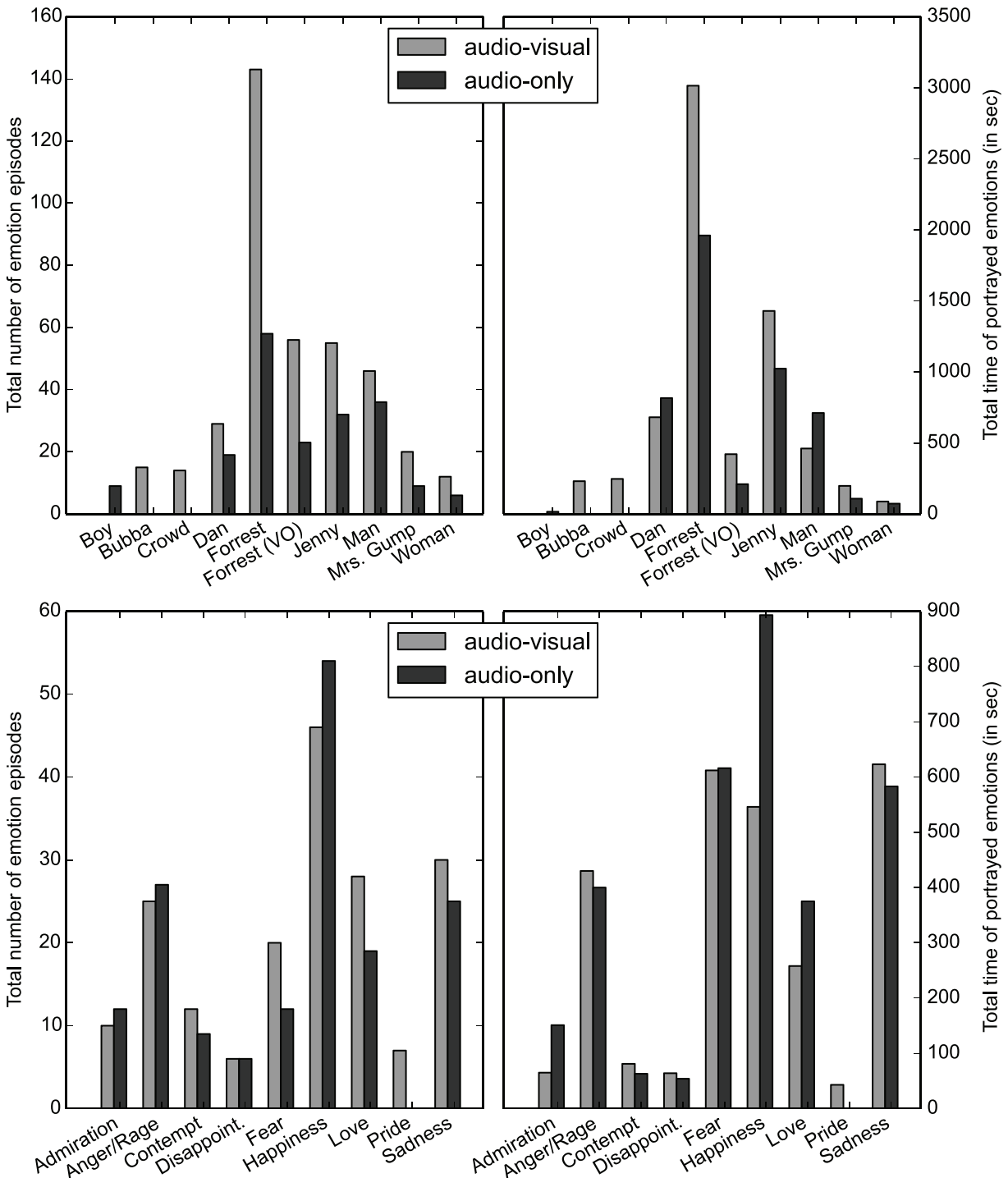
**Intra-stimulus inter-observer consistency** The reliability of IOA time courses was estimated by repeatedly splitting the set of observers into two groups of approximately equal size. Within each group, IOA time courses were computed and correlated across the observer subgroups. Split-half correlations were computed for all possible ways of combining observers into two groups (without replacement). **Table 5** shows the average time series correlation across all splits and the width of the 95% confidence interval of the mean.

In general, IOA reliability is larger for the AV movie annotations than for the AO movie. This finding is likely a result of the lower number of observers for the latter. Intra-stimulus IOA reliability estimates suggest that a subset of emotion attributes or categories is not contained in the AO movie stimulus. The most prominent

**Table 4. Position of the fMRI stimulus segments with respect to the movie annotation time.** The last column lists the position of the scene boundary used as a reference for the segment transition. All times are in seconds and refer to the full (unsegmented) stimulus, which is a shorted version of the original movie<sup>1</sup>.

#	Start	End	Duration	Boundary
0	0.0	902.0	902.0	891.2
1	886.0	1768.0	882.0	1759.2
2	1752.0	2628.0	876.0	2618.8
3	2612.0	3588.0	976.0	3578.5
4	3572.0	4496.0	924.0	4488.0
5	4480.0	5358.0	878.0	5349.2
6	5342.0	6426.0	1084.0	6418.2
7	6410.0	7086.0	676.0	7085.5

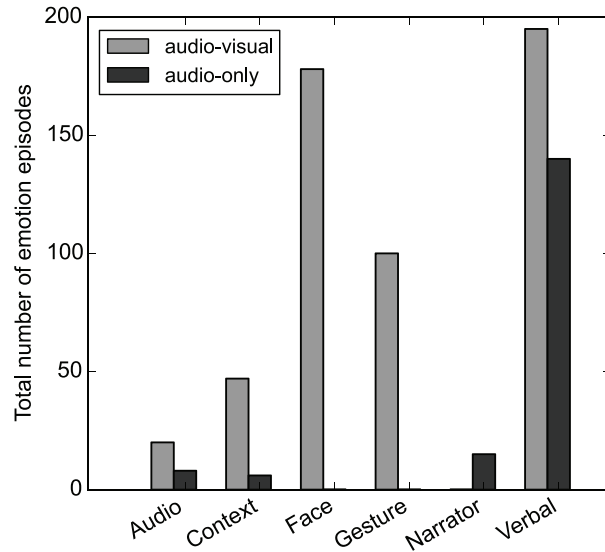




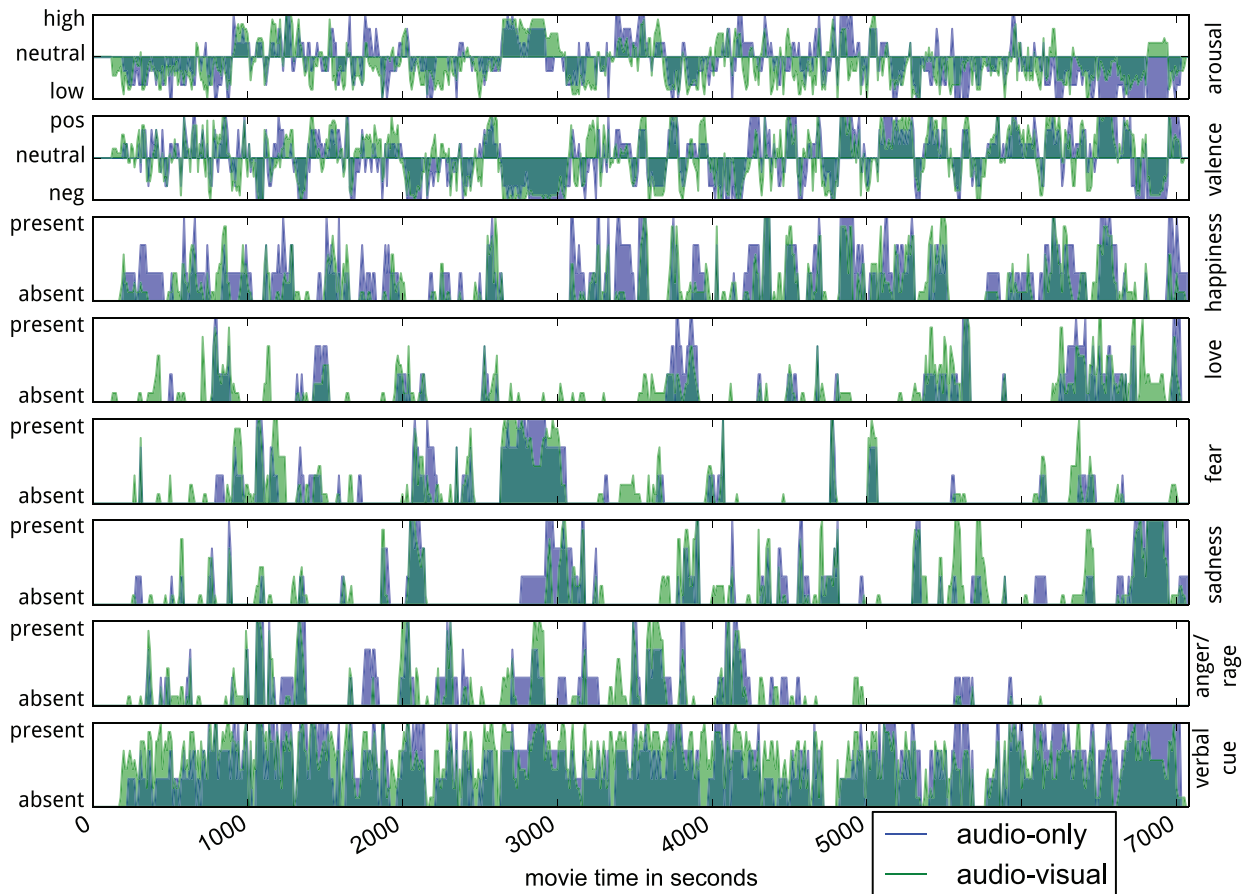
**Figure 1. Annotation frequencies by character and emotion.** The minimum criterion for counting emotion episodes is a 50% inter-observer agreement. Only categories/characters with a minimum of five such episodes are shown. The left panels show the number of episodes for both movie variants; the right panels show the cumulative duration of emotion display across all considered episodes. The top panels show the annotation frequency by movie character; the bottom panels show the corresponding frequencies for emotion categories.

**Table 5. Intra-stimulus inter-observer consistency.** All values are Spearman correlations for 1 Hz modulations of the fraction of IOA (as, for example, depicted in figure 3) with respect to a particular emotion attribute across the entire duration of the movie. The specified range corresponds to the width of the 95% confidence interval of the mean correlation for all possible combinations of partitioning observers into two sub-groups (audio-visual 4 vs. 5; audio-only 1 vs. 2). Higher correlations indicate higher consistency of agreement modulations across observer sub-groups. Correlations higher than 0.5 (arbitrary threshold) are depicted in bold for visualization purposes. The *all characters* column indicates the agreement for a particular emotion attribute over time irrespective of the annotated character. The columns on the right show the corresponding correlations for the two main characters. *n/a* fields indicate an insufficient number of annotations to compute the consistency measure.

	All characters		Forrest-only		Jenny-only	
	Audio-visual	Audio-only	Audio-visual	Audio-only	Audio-visual	Audio-only
<i>Dimensional emotion attributes</i>						
Arousal	<b>0.676</b> ±0.006	0.305 ±0.078	<b>0.600</b> ±0.008	0.443 ±0.078	<b>0.688</b> ±0.006	0.330 ±0.076
Valence	<b>0.851</b> ±0.004	<b>0.564</b> ±0.044	<b>0.784</b> ±0.006	0.440 ±0.054	<b>0.858</b> ±0.003	<b>0.598</b> ±0.079
Direction	<b>0.555</b> ±0.011	0.068 ±0.050	<b>0.543</b> ±0.012	-0.059 ±0.072	<b>0.576</b> ±0.012	0.053 ±0.145
<i>Emotion categories</i>						
Admiration	0.392 ±0.006	0.343 ±0.094	0.306 ±0.015	n/a	n/a	<b>0.643</b> ±0.197
Anger/Rage	<b>0.719</b> ±0.005	0.377 ±0.042	<b>0.693</b> ±0.006	0.141 ±0.100	<b>0.809</b> ±0.007	<b>0.626</b> ±0.109
Compassion	0.337 ±0.010	0.182 ±0.013	0.232 ±0.019	n/a	0.459 ±0.018	0.363 ±0.042
Contempt	0.433 ±0.008	0.157 ±0.043	n/a	n/a	0.404 ±0.020	n/a
Disappoint.	0.408 ±0.008	0.367 ±0.095	0.431 ±0.007	0.456 ±0.129	n/a	n/a
Fear	<b>0.749</b> ±0.003	<b>0.546</b> ±0.108	<b>0.729</b> ±0.004	<b>0.504</b> ±0.072	<b>0.633</b> ±0.005	0.392 ±0.013
Fears conf.	0.048 ±0.009	n/a	n/a	n/a	n/a	n/a
Gloating	<b>0.644</b> ±0.014	0.045 ±0.038	n/a	n/a	n/a	n/a
Gratification	0.155 ±0.010	n/a	0.242 ±0.017	n/a	n/a	n/a
Gratitude	0.317 ±0.012	n/a	0.267 ±0.024	n/a	0.165 ±0.015	n/a
Happiness	<b>0.618</b> ±0.006	0.289 ±0.107	<b>0.561</b> ±0.007	0.381 ±0.060	<b>0.738</b> ±0.005	<b>0.532</b> ±0.108
Happy-for	0.478 ±0.013	n/a	0.408 ±0.028	n/a	n/a	n/a
Hate	0.441 ±0.011	0.097 ±0.070	0.478 ±0.027	n/a	n/a	n/a
Hope	0.333 ±0.010	0.217 ±0.060	0.202 ±0.008	n/a	0.439 ±0.026	<b>0.804</b> ±0.052
Love	<b>0.568</b> ±0.008	0.396 ±0.032	0.472 ±0.008	0.327 ±0.092	<b>0.660</b> ±0.009	0.432 ±0.040
Pride	0.473 ±0.011	0.156 ±0.054	<b>0.536</b> ±0.009	n/a	n/a	n/a
Relief	0.353 ±0.019	-0.003 ±0.001	<b>0.505</b> ±0.028	n/a	n/a	n/a
Remorse	<b>0.557</b> ±0.022	0.323 ±0.107	n/a	0.453 ±0.364	0.464 ±0.006	n/a
Resentment	n/a	n/a	n/a	n/a	n/a	n/a
Sadness	<b>0.737</b> ±0.005	0.464 ±0.023	<b>0.750</b> ±0.005	<b>0.522</b> ±0.006	<b>0.695</b> ±0.009	0.228 ±0.040
Satisfaction	0.112 ±0.007	n/a	n/a	n/a	n/a	n/a
Shame	0.316 ±0.006	0.185 ±0.112	0.290 ±0.006	0.034 ±0.036	0.201 ±0.015	n/a
<i>Emotion onset cues</i>						
Audio	0.418 ±0.008	0.037 ±0.051	0.354 ±0.013	n/a	0.203 ±0.007	n/a
Context	0.384 ±0.010	n/a	0.424 ±0.011	n/a	<b>0.547</b> ±0.010	n/a
Face	<b>0.713</b> ±0.006	n/a	<b>0.687</b> ±0.010	n/a	<b>0.880</b> ±0.004	n/a
Gesture	<b>0.619</b> ±0.010	n/a	<b>0.532</b> ±0.011	n/a	<b>0.769</b> ±0.009	n/a
Narrator	n/a	0.251 ±0.047	n/a	0.202 ±0.045	n/a	<b>0.578</b> ±0.286
Verbal	<b>0.637</b> ±0.008	0.368 ±0.107	<b>0.565</b> ±0.007	0.455 ±0.042	<b>0.783</b> ±0.005	<b>0.626</b> ±0.168



**Figure 2.** Number of emotion episodes for all distinguished onset cue categories (minimum 50% inter-rater agreement) for both stimulus types.

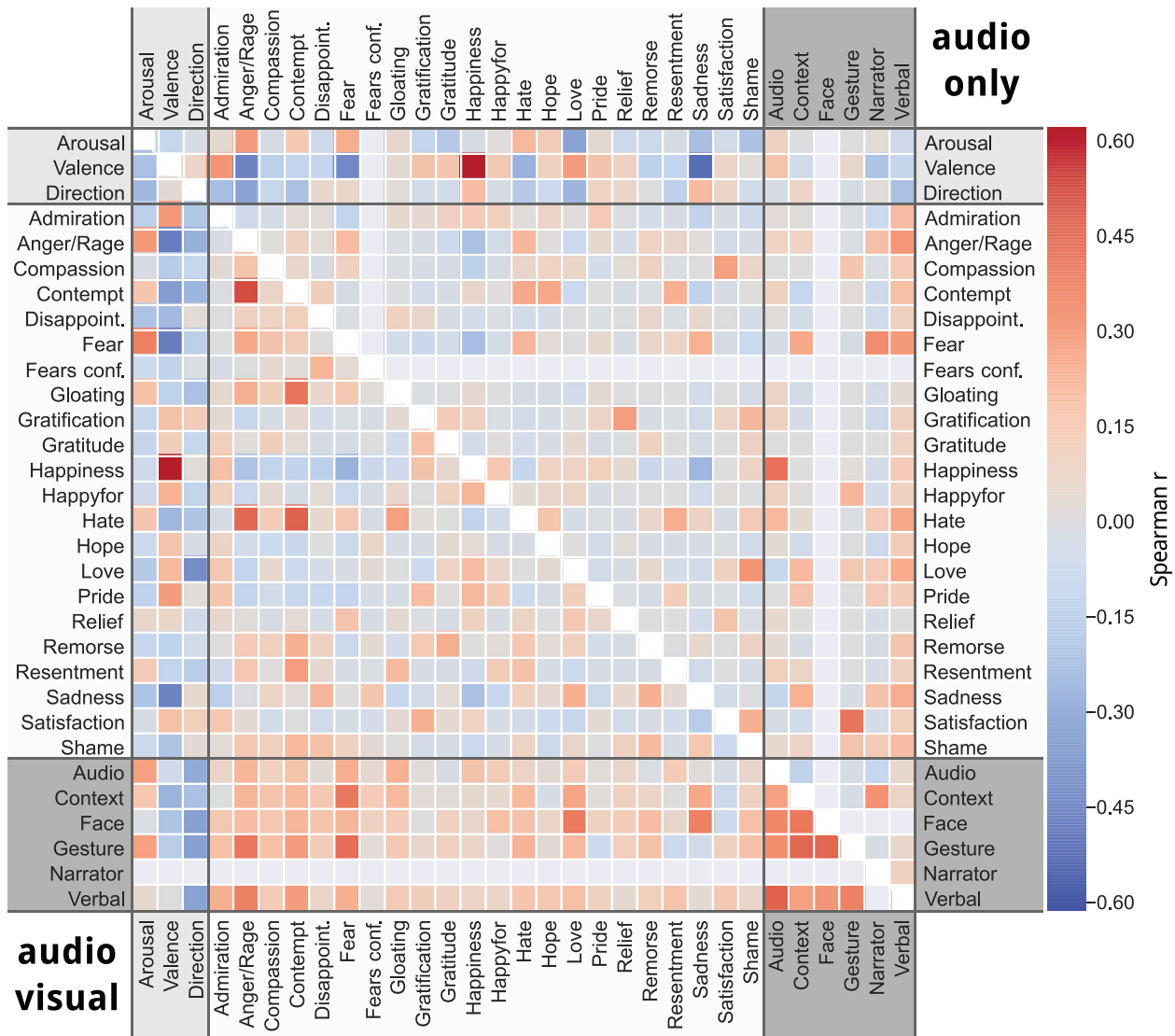


**Figure 3.** Inter-observer agreement (IOA) time courses for selected emotion attributes across the full duration of the movie. Extreme values indicate perfect agreement with respect to presence or absence of a particular emotion attribute at a given time. IOA time series for the two movie types are overlaid on top of each other. For the purpose of visualization, the depicted time series reflects IOA computed using a 10 s segment size — in contrast to 1 s used for all other statistics presented in this paper.

example is the direction of an emotion. Emotion categories like hate, shame, or relief also show large reliability differences between stimulus types. However, in these cases the small number of emotion episodes may multiply the negative effect of the small number of observers on the interpretability of this estimate.

Uniformly high reliability estimates were found for the two main dimensional emotion attributes arousal and valence, as well as for the five most frequently annotated emotion categories anger/rage, fear, happiness, love, and sadness.

**Intra-stimulus inter-variable correlations** In order to assess possible confounds in the annotations, we computed the correlation of intra-stimulus IOA time courses (across all observers) for all pairwise combinations of emotion attributes and categories. The correlation matrices for the AV and the AO movie stimulus are depicted in figure 4. The results reveal an expected correlation pattern between individual emotion categories and the two main dimensional attributes arousal and valence — with stronger correlations being observed for more primary and more frequently portrayed emotions. Interestingly, audio and verbal cues in the



**Figure 4. Intra-stimulus indicator correlation.** The lower triangular matrix depicts the correlations between the IOA time courses for the three primary bipolar emotion attributes (arousal [high/low], valence [pos/neg], and direction [self/other]), the 22 emotion categories, and the six emotion onset indicators for the audio-visual movie. The upper triangular matrix shows the corresponding correlations for the audio-only movie. There were no observations for the emotion “fears confirmed”, or the facial expression onset cue in the audio-only movie. Likewise, there were no observations for the “narrator” onset cue in the audio-visual movie. The corresponding undefined correlations are depicted as zeros.

AV stimulus seems to be more associated with a high state of arousal and emotions directed towards other people than in the AO stimulus. In the AV stimulus, facial expressions seem to have a tendency to communicate negative emotions rather than positive ones.

**Inter-stimulus consistency** To estimate the similarity of the two stimulus types, we computed the correlation of IOA time courses

for the AV and the AO movie with respect to emotion attributes and categories. **Table 6** shows the respective correlations and their 95% confidence intervals. Again, the three dimensional attributes as well as the five most frequent emotion categories show significant correlations. Notably, verbal cues are the most consistent indicator of emotions across both types of stimuli. This is evidence for an expected substantial semantic similarity with respect to the emotional content in the two variants of the “Forrest Gump” movie.

**Table 6. Inter-stimulus indicator correlation.** All values are Spearman correlations of 1 Hz modulations of the fraction of IOA (as, for example, depicted in [figure 3](#)) for the two stimulus types (audio-visual and audio-only movie). The specified range corresponds to the width of the 95% confidence interval of the correlation (computed via Fisher transformation). Correlations higher than 0.5 (arbitrary threshold) are depicted in bold for visualization purposes. *n/a* fields indicate an insufficient number of annotations to compute the consistency measure.

	All characters	Forrest-only	Jenny-only
<i>Dimensional emotion attributes</i>			
Arousal	0.474 ±0.018	0.278 ±0.021	<b>0.518</b> ±0.017
Valence	<b>0.726</b> ±0.011	<b>0.557</b> ±0.016	<b>0.693</b> ±0.012
Direction	0.425 ±0.019	0.295 ±0.021	<b>0.509</b> ±0.017
<i>Emotion categories</i>			
Admiration	0.260 ±0.022	0.149 ±0.023	0.197 ±0.022
Anger/Rage	0.485 ±0.018	0.307 ±0.021	0.478 ±0.018
Compassion	0.232 ±0.022	-0.018 ±0.023	0.357 ±0.020
Contempt	0.201 ±0.022	n/a	0.041 ±0.023
Disappointment	0.372 ±0.020	0.409 ±0.019	n/a
Fear	<b>0.673</b> ±0.013	<b>0.525</b> ±0.017	<b>0.799</b> ±0.008
Gloating	0.231 ±0.022	n/a	n/a
Gratification	0.062 ±0.023	-0.004 ±0.023	-0.003 ±0.023
Gratitude	0.201 ±0.022	0.071 ±0.023	0.444 ±0.019
Happiness	<b>0.532</b> ±0.016	0.499 ±0.017	<b>0.645</b> ±0.013
Happy-for	0.289 ±0.021	0.118 ±0.023	0.064 ±0.023
Hate	0.266 ±0.022	0.279 ±0.021	-0.001 ±0.023
Hope	0.181 ±0.022	0.123 ±0.023	0.464 ±0.018
Love	<b>0.618</b> ±0.014	0.444 ±0.019	<b>0.592</b> ±0.015
Pride	0.203 ±0.022	0.194 ±0.022	n/a
Relief	0.062 ±0.023	-0.007 ±0.023	n/a
Remorse	0.376 ±0.020	<b>0.578</b> ±0.015	0.316 ±0.021
Resentment	<b>0.585</b> ±0.015	-0.001 ±0.023	n/a
Sadness	<b>0.545</b> ±0.016	<b>0.537</b> ±0.016	<b>0.653</b> ±0.013
Satisfaction	0.038 ±0.023	-0.007 ±0.023	n/a
Shame	0.254 ±0.022	0.401 ±0.019	0.104 ±0.023
<i>Emotion onset cues</i>			
Audio	0.236 ±0.022	0.319 ±0.021	0.245 ±0.022
Context	0.267 ±0.021	0.322 ±0.021	0.316 ±0.021
Verbal	<b>0.582</b> ±0.015	<b>0.560</b> ±0.016	<b>0.839</b> ±0.007

## Data availability

F1000Research: Dataset 1. Annotations of portrayed emotions in the movie “Forrest Gump”, [10.5256/f1000research.6230.d45328](https://doi.org/10.5256/f1000research.6230.d45328)<sup>14</sup>

## Author contributions

AKP, AL, BH, FRS, HS, LNR, MB, MGE, MGO, MP, NH, SK, and TR designed the data collection procedure, performed the annotation, and contributed to the manuscript; MH conceived the study, aggregated data from individual observers into the released form, performed the data analysis, and contributed to the manuscript. All authors have agreed to the final content of the manuscript.

## Competing interests

No competing interests were disclosed.

## Grant information

This research was, in part, supported by the German Federal Ministry of Education and Research (BMBF) as part of a US-German collaboration in computational neuroscience (CRCNS; awarded to James Haxby, Peter Ramadge, and Michael Hanke), co-funded by the BMBF and the US National Science Foundation (BMBF 01GQ1112; NSF 1129855). Michael Hanke was supported by funds from the German federal state of Saxony-Anhalt, Project: Center for Behavioral Brain Sciences.

## Acknowledgements

We are grateful to Roland Nigbur for his advice during all stages of this study. Additionally, we thank Alex Waite and Florian Baumgartner for their feedback on the manuscript.

## References

- Hanke M, Baumgartner FJ, Ibe P, *et al.*: **A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie.** *Scientific Data*. 2014; **1**. [Publisher Full Text](#) | [Free Full Text](#)
- Hasson U, Honey CJ: **Future trends in Neuroimaging: Neural processes as expressed within real-life contexts.** *Neuroimage*. 2012; **62**(2): 1272–1278. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gross JJ, Levenson RW: **Emotion elicitation using films.** *Cognition & Emotion*. 1995; **9**(1): 87–108. [Reference Source](#)
- Ekman P: **Facial expressions of emotion: New findings, new questions.** *Psychological Sci*. 1992; **3**(1): 34–38. [Publisher Full Text](#)
- Ethofer T, Van De Ville D, Scherer K, *et al.*: **Decoding of emotional information in voice-sensitive cortices.** *Curr Biol*. 2009; **19**(12): 1028–1033. [PubMed Abstract](#) | [Publisher Full Text](#)
- Skerry AE, Saxe R: **A common neural code for perceived and inferred emotion.** *J Neurosci*. 2014; **34**(48): 15997–16008. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gray EK, Watson D: **Assessing positive and negative affect via self-report.** In J. A. Coan and J. J. B. Allen, editors, *Handbook of emotion elicitation and assessment*. Oxford University Press, Oxford. 2007; 171–183. [Reference Source](#)
- Ekman P: **An argument for basic emotions.** *Cognition & Emotion*. 1992; **6**(3–4): 169–200. [Publisher Full Text](#)
- Scherer KR: **What are emotions? And how can they be measured?** *Social Sci Inf*. 2005; **44**(4): 695–729. [Publisher Full Text](#)
- Russell JA: **A circumplex model of affect.** *J Personality Social Psychol*. 1980; **39**(6): 1161–1178. [Publisher Full Text](#)
- Coan JA, Gottman JM: **The specific affect coding system (SPAFF).** In J. A. Coan and J. J. B. Allen, editors, *Handbook of emotion elicitation and assessment*. Oxford University Press, Oxford. 2007; 267–285. [Reference Source](#)
- Aubert O, Prié Y: **Advene: active reading through hypertext.** In *Proceedings of ACM Hypertext*. 2005; 235–244. [Publisher Full Text](#)
- Ortony A: **The cognitive structure of emotions.** Cambridge university press, 1990. [Reference Source](#)
- Dataset 1. Annotations of portrayed emotions in the movie “Forrest Gump”.** *F1000Research*. 2015. [Data Source](#)

# Open Peer Review

Current Referee Status:



Version 1

Referee Report 27 April 2015

doi:10.5256/f1000research.6684.r8301



**Ariel Rokem**

Department of Psychology, Stanford University, Stanford, CA, USA

This paper provides a description and analysis of a publicly released data-set: an annotation of the emotions expressed by characters in different variants of the movie Forrest Gump, in either audio or audio-visual formats. The data-set is interesting in its own right, and will be eminently useful in conjunction with the fMRI data that this group has previously released, and will continue to release. This data paper goes well beyond a standard description of the data-set, by providing analysis and interpretation of the data.

I have just one comment with regards to providing direction for eventual interpretation of the data by users: The conflation of gender differences in perceiving emotions with audio vs. visual distinctions is a real concern, and the authors acknowledge this, but the current description does not provide enough direction on how to deal with this potential confound. I think that it would be helpful if a few sentences about gender differences in perceiving emotion would be added, and some evaluation of what the effects of these differences might be in this data. For example, do differences between AO and AV annotations coincide with predictions from the gender difference literature?

It is commendable (and still sadly unusual) that that authors make software that is so easy to use, which reproduces pretty much every single statement they make about the data and produces all the figures. That said, I do have a few comments for improvements of this software:

1. Though not strictly necessary, in this day and age, it would be nice to make the scripts provided compatible with python 3 (from a quick look, it looks like the addition of some parentheses should do the trick, and see more below).
2. Making the code PEP8 compatible (e.g. line-wrapping) would make it easier to read.
3. The authors might disagree, but pymvpa seems like an onerous dependency just to supply some utility functions used in the analysis. I would add ``plot_bars``, ``xunique_combinations`` and ``unique_combinations`` to the code distributed with the data. Here's a gist of the file I used when my computer refused to compile libsvm: <https://gist.github.com/arokem/cd4270a78cbad58c2cd4>
4. As a conclusion from the previous point, a (github) repository with the code used here would be good, for future users of this data/code to be able to make their proposals for improvements to the code via pull requests. The state of the code for this publication could be a tag/release in this repo.

Minor:

Table 3, the description of HAPPYFOR is: 1) not clear and 2) not capitalized (in contrast to the other descriptions). I am guessing you meant something like: "Like HAPPINESS, but as reflected off another character", or some-such.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

Referee Report 21 April 2015

doi:[10.5256/f1000research.6684.r8300](https://doi.org/10.5256/f1000research.6684.r8300)



**Dylan Wagner**

Department of Psychology, Ohio State University, Columbus, OH, USA

This data note describes a dataset consisting of time indices and annotations for the emotional content of scenes from the movie Forest Gump. The primary intent of this dataset is to encourage analyses of real-world cognition during the viewing of naturalistic stimuli and, as such, it is meant to accompany a previously released dataset of functional neuroimaging data using the same movie stimulus as used here.

I have only minor quibbles:

- The authors may wish to indicate which observers were familiar with the movie prior to annotation.
- I recommend adding a section in which the authors discuss some of the potential issues that may limit the correspondence between the emotion annotation dataset and the neuroimaging dataset of the same movie. Specifically, the use of random scene order during annotation (but not during movie viewing) or the gender split among observers who annotated the audio movie and those that annotated the audiovisual movie.
- At the expense of looking a gift horse in the mouth, an ipython notebook capable of generating all figures and descriptive data would have been marginally preferable to a script.

Finally, I would like to note that the release of this dataset along with the associated neuroimaging data previously released by members of this group represents an extremely generous sharing of resources. I commend the effort and look forward to seeing what emerges from researchers looking into this treasure trove of data.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

Author Response 22 Apr 2015

**Michael Hanke**, Otto-von-Guericke-Universität Magdeburg, Germany

Thanks a lot for the swift and positive feedback!



re familiarity: I think it is fair to say that by the time when observers started the actual annotation, all of them were somewhat familiar with the movie, due to the pre-tests we ran on the feasibility of particular annotation approaches -- independent of whether they saw the movie before. We'll add a corresponding statement.

re correspondence limitation: we could prepare a dedicated summary of aspects that may have an impact. However, other than the two aspects mentioned by the reviewer it may only be a question of "detection probability" which should be reflected in the inter-observer agreement. The issue of scene order randomization should emphasize the portrayed emotions as opposed to lower-frequency mood modulations induced by the movie (see procedure description). The gender split is an issue, but its impact is unclear. Most likely it causes the annotations to be less representative for the opposite gender. However, the annotation procedure encouraged a repeated, rather analytic dissection of the stimulus, in contrast to a single-pass introspection of how a scene "felt". This may compensate for an initial gender bias in the perception of the stimulus.

re ipython notebook: great suggestion! We'll convert the script into a notebook, and will provide both.

Thanks again!

**Competing Interests:** No competing interests were disclosed.

---