

# Introducing MULAI: A Multimodal Database of Laughter during Dyadic Interactions

Michel-Pierre Jansen\*, Khiet P. Truong\*, Deniece S. Nazareth\* †, Dirk K.J. Heylen\*

\*University of Twente - Human Media Interaction

{m.jansen-1, k.p.truong, d.k.j.heylen}@utwente.nl

†University of Twente - Psychology, Health and Technology

d.s.nazareth@utwente.nl

## Abstract

Although laughter has gained considerable interest from a diversity of research areas, there still is a need for laughter specific databases. We present the Multimodal Laughter during Interaction (MULAI) database to study the expressive patterns of conversational and humour related laughter. The MULAI database contains 2 hours and 14 minutes of recorded and annotated dyadic human-human interactions and includes 601 laughs, 168 speech-laughs and 538 on- or offset respirations. This database is unique in several ways; 1) it focuses on different types of social laughter including conversational- and humour related laughter, 2) it contains annotations from participants, who understand the social context, on how humorous they perceived themselves and their interlocutor during each task, and 3) it contains data rarely captured by other laughter databases including participant personality profiles and physiological responses. We use the MULAI database to explore the link between acoustic laughter properties and annotated humour ratings over two settings. The results reveal that the duration, pitch and intensity of laughs from participants do not correlate with their own perception of how humorous they are, however the acoustics of laughter do correlate with how humorous they are being perceived by their conversational partner.

**Keywords:** database, dyadic, interaction, audio, video, physiological, social laughter, acoustic properties, self-ratings, perceiver-ratings, humour

## 1. Introduction

Many theories exist about the exact origin and function of laughter (Gervais and Wilson, 2005), however most theorists agree that laughter plays a major role in day to day social interactions between humans. Laughter can be complex and displayed through multiple modalities including vocal expressions, facial expressions and body movements. Some research points out that (prosodic) laughter properties differ depending on the social function and situation of the laughter (Campbell, 2007; Tanaka and Campbell, 2011), whereas others (Curran et al., 2018; Owren and Bachorowski, 2003) argue that for most prosodic features the evidence is inconclusive and point towards social context for a better understanding of laughter. In addition, knowledge about how multimodal signals of expressive patterns of laughter interact with each other is still rather limited (Niewiadomski et al., 2013).

Hence, to study laughter and its expressive patterns within the social context, databases with social interactions are needed that contain annotated laughter sequences. Although several databases containing annotated laughter are available to the research community, they are often not specifically designed for the research of functionally different laughs within a social setting. The difficulty of studying laughter from such databases is that laughs often are elicited outside of a social context and therefore miss the specific nuances, different functional properties, and temporal sensitive nature of social laughter.

Laughter databases that focus on different forms of social laughter, frequently focus on a single modality such as audio- or visual data and lack other modalities that could be interesting to the community such as physiological signals. In addition, laughter annotation procedures and quality varies between and within currently available databases,

partly due to having other goals or the lack of an at the time widely accepted and used laughter annotation scheme and procedure. This variety in annotation quality makes it challenging to do analysis on more finely grained aspects of conversational laughter and comparing laughter data from different databases.

With the MULAI database, we introduce a database that focuses on different types of social laughter based on their function and social situation. Two of the prevalent forms of laughter that occur in the database are laughs that occurs in more open, conversational tasks and laughs that are elicited during joke telling tasks and seem to be more humour related. Humour annotations were collected from interlocutors who each reported how humorous they perceived themselves and how humorous they perceived their conversational partner after each task. One of the advantages of annotation method is that interlocutors can leverage their past experiences together with knowledge of the current social context and make more accurate judgements compared to third party annotators that often have limited knowledge of the context. Our previous work highlights the importance and benefits of having annotators with a good understanding of the context where the annotated behaviour is expressed (Dudzik et al., 2019).

The humour annotations from the interlocutors can be used to study the relationship between the characteristics of different types of laughter and perceived humour. The link between properties of laughter and perceived humour has had some attention (Gervais and Wilson, 2005; Petridis and Pantic, 2009; McKeown and Curran, 2015; Curran et al., 2018), however these studies used third party annotators that lack contextual knowledge. Petridis and Pantic (2009) investigated how perceived humour ratings related to if a laugh was primarily voiced or unvoiced and observed that

voiced laughter often is associated with higher humour ratings. McKeown and Curran (2015) also explored the relationship between laughter and perceived humour. They found a strong relationship between perceived humour and perceived laughter intensity by these raters.

Beside humour ratings from the participants, other available data can help with exploring the previously mentioned link between humour and expressive laughter patterns. The database provides access to synchronised physiological data modalities uncommon in other databases, which includes movement (IMU)<sup>1</sup>, heart-rate (ECG)<sup>2</sup> and skin conductance (GSR)<sup>3</sup> data of each participant. Personality profiles have been captured with multiple personality questionnaires and can give insight into individual differences of participants. Some scientific work highlights a link between personality and humour, for example Thorson and Powell (1993) demonstrate in an early study the link between sense of humour and personality attributes. This paper introduces the MULAI database, a new multimodal social laughter database that focuses on different forms of laughter including laughter from a conversational setting and laughter elicited through humour in a social setting. The main goal of this database is to provide researchers with data to study social expressive patterns of laughter and address the challenges previously mentioned. In this study, we use these annotations to further explore the link between the expressive patterns and properties of laughter and perceived humour in two different settings.

## 2. Related work

Databases can be divided in three groups based on the level of social interaction in which the laughter is recorded; single participant recordings, recordings of dyadic interactions between participants, and multi-party recordings often in the form of meetings focusing on a central theme. Social laughter frequently occurs in both dyadic and multi-party based recordings, but each serves its own purpose since interaction dynamics can be very different for both settings. Table 1 gives an overview of frequently accessed databases that contain laughter. We will now discuss each of these social settings separately.

### 2.1. Single participant, non-interaction databases

Single participant databases like the AVL- (Urbain et al., 2010), MANHOB- (Petridis et al., 2013), MMI-V- (Valstar and Pantic, 2010) and the Montreal database (Belin et al., 2008) often do not contain social laughter since no interactions are recorded. During the recordings, the participants in these databases are usually asked to either produce laughter on command (posed laughter) or watch some stimuli in the form of pictures or movies aimed at inducing laughter. In this category only the MANHOB-database created by Petridis et al. (2013) contains other modalities outside of the usual audio- and video modalities.

---

<sup>1</sup>Inertial Measurement Units

<sup>2</sup>Electrocardiogram sensors

<sup>3</sup>Galvanic Skin Response

### 2.2. Dyadic interaction databases

Dyadic conversational databases are the most prevalent form of databases, examples include the DUEL- (Hough et al., 2008), AVIC- (Schuller et al., 2009), DD- (Cohn et al., 2009), DiapixUK- (Baker and Hazan, 2011), HCRC- (Thompson et al., 1993) and the SEMAINE database (McKeown et al., 2007). Laughter inducing tasks, scenarios and conversational topics are often deployed in this category of databases. Interestingly, the DD database used natural recordings of real clinical interviews with participants. However, due to the focus of the DD database, recorded laughter is relatively uncommon. The database we introduce in this paper, the MULAI database, also contains recordings of dyadic interactions.

### 2.3. Multi-party interaction databases

Multi-party databases contain laughter that is produced in a social setting with multiple participants. Examples of multi-party databases that contain laughter annotations are, MMLI (Niewiadomski et al., 2013), ICSI-MRDA (Shriberg et al., 2004) and the AMI database (Mccowan et al., 2005). The creators of these databases often record natural, meetings on specific topics or asked participants to play out a scenario. The MMLI database takes a different approach, here participants were asked to watch funny video clips or play social games in groups, in addition some dyadic interactions were recorded.

Table 1 reveals that not many databases that are specifically focused on laughter include other modalities outside of the audio- and visual modalities. Only the MMLI (Niewiadomski et al., 2013), DUEL (Hough et al., 2008), ILHAIRE (McKeown et al., 2012) and MANHOB (Petridis et al., 2013) database provide other modalities to study when researching laughter. With the exception of the MMLI, these databases focus on body movement. The MULAI database is in this way uniquely positioned since it provides body movement, ECG and GSR data. These less commonly available modalities can be leveraged by interested researchers.

## 3. The MULAI Database

The MULAI database contains 13 sessions of dyadic human-human interactions with social conversational laughter and humour induced laughter, totalling 357 minutes of recorded video-, audio- and physiological data streams. The annotated part of the MULAI database spans 134 minutes and is rich in laughter related events with 601 annotated laughter bouts, 168 annotated speech-laugh and 538 laughter related events (mostly related to on-/offset breaths). This database provides researchers with the data to study the expressive patterns of conversational laughter.

### 3.1. Participants

Students from a course were asked to find a participant from outside the course and participate together in a data collection session. In total 32 participants participated over a span of 16 sessions. From this group, 6 participants are not included in the database as they did not give consent for incorporating their data in the MULAI database. This results in a database of 26 participants (age  $M = 24$ ,  $SD = 2.3$

Table 1: Existing databases containing laughter annotations. From left to right: database, elicitation method or source of laughter, recorded modalities (A= audio, AV = audio-visual, BM = body movement, Resp = respiration, TM = thermal camera), reference.

Database	Elicitation procedure	Recorded modalities	Reference
<b>No interaction</b>			
AVLC	Watching funny video clips	AV	Urbain et al. (2010)
MANHOB	Watching funny video clips and posed laughter	AV, TM	Petridis et al. (2013)
MMI-V	Watching funny audio and video clips	AV	Valstar and Pantic (2010)
Montreal	Laughter during posed emotional vocalizations	A	Belin et al. (2008)
<b>Dyadic interaction</b>			
<b>MULAI</b>	<b>Conversational and humour elicited laughter</b>	<b>AV, BM, ECG, GSR</b>	-
DUEL	Conversational laughter	AV, BM	Hough et al. (2008)
AVIC	Laughter during played out scenarios	AV	Schuller et al. (2009)
DD	Laughter during clinical interviews	AV	Cohn et al. (2009)
DiapixUK	Conversational laughter	A	Baker and Hazan (2011)
HCRC	Conversational laughter	A	Thompson et al. (1993)
SEMAINE	Laughter during conversations with agents	AV	McKeown et al. (2007)
<b>Multi-party interaction</b>			
MMLI	Watching funny video clips and social games	AV, BM, Resp	Niewiadomski et al. (2013)
ICSI MRDA	Recorded meetings on several topics	A	Shriberg et al. (2004)
AMI	Recorded meetings and scenario driven meetings	AV	Mccowan et al. (2005)
ILHAIRE	Selection of multiple other databases	AV, BM	McKeown et al. (2012)

years), consisting of 14 male and 12 female participants. Most of the participating pairs were at least to some extent familiar with each other. A majority of the participants are Dutch (N=17) and in addition, some international students participated in the study. All participants spoke English.

### 3.2. Measurements

All participants were equipped with a similar set of microphones, cameras and sensors. Visual data of the face and upper body was captured with two Panasonic HC-V180EG-K cameras. Videos were shot with a resolution of 1920 x 1080 at 50 hz. Audio data was recorded with a Shure BLX14E-M17 wireless microphone set (in combination with a Zoom H6) to capture high quality audio at 48 kHz.

Physiological data was captured using Shimmer sensor units at a sampling rate of 512 hz. Participants wore three different units, respectively around both wrists and one on the chest, figure 1 shows how participants wore the different sensors.



Figure 1: An example of participants interacting during one of the sessions of the experiment, all visible sensors are encircled. Each color represents a different kind of sensor (indigo = over-the-shoulder cameras, turquoise = EXG+ units, red = regular IM units, green = GSR+ units and orange = lavalier microphones)

All Shimmer units capture inertial movement (IM) with 9 degrees of freedom, which is obtained through accelerometers, a gyroscope and a magnetometer. Additionally the Shimmer GSR+ unit obtains data on the Galvanic Skin Response (skin conductance response) and was placed on the wrist of the dominant hand, the Shimmer EXG+ unit retrieved electrocardiograph data from electrodes placed on each participants torso.

Besides sensor data, questionnaires were filled in by the participants to retrieve information on how humorous they perceived themselves and how humorous they perceived their conversation partner. These questionnaires contained a self-rating scale with the statement ‘I think I was funny’ and a perceived-rating scale with the statement ‘the other was very funny’. Both are five point Likert scales and have an answer range from ‘completely disagree’ to ‘completely agree’. In addition, the Likert scales were accompanied with emoticons reflecting different states of amusement to make answering the statements more intuitive for the participants.

Finally, questionnaire data has been retrieved that facilitates researchers interested in participant profiles. In order to learn more about participants’ personalities and preferences, participants filled in the Ten Item Personality Inventory (TIPI) (Gosling et al., 2003) and the International Personality Item Pool (IPIP-50-R) (Goldberg et al., 2006). Both questionnaires have reasonably to good psychometric qualities across different cultures (Jonason et al., 2011; Ypofanti et al., 2015; Zheng et al., 2008; Romero et al., 2012). The TAS-20 (Bagby et al., 1994a; Bagby et al., 1994b) was deployed to make predictions about the introspective qualities of participants. A demographic questionnaire was also presented to the participants. All questionnaires were filled in directly after participants signed an informed consent form and before further data collection.

### 3.3. Recording Set-up

After completing the consent form and questionnaires, participant pairs were led into the experiment room. They were seated opposite of each other at a table with letter indications corresponding to the participants letter ID (A or B). Cameras were placed on both sides of the table at fixed locations, pointing towards the participant sitting on the opposite side of the table. See Figure 2 for an example of how participants were placed during the sessions. Camera height and zoom was manually adjusted to record the most optimal frame. Each participant wore a lavalier microphone around the neck. As described in the previous section, each participant was equipped with three separate Shimmer devices; the EXG+ unit, the GSR+ unit and the IMU unit. These were placed respectively, on the torso, the wrist of the dominant hand and the wrist of the non-dominant hand.

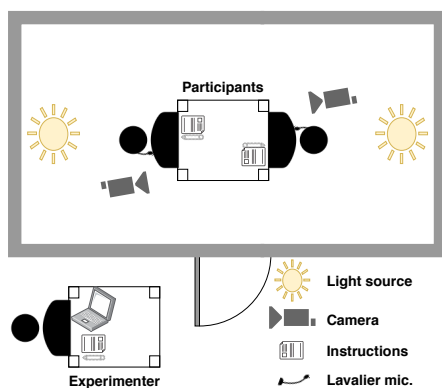


Figure 2: Recording setup displaying the positions of the participants, light sources, recording equipment and experimenter during the sessions.

### 3.4. Recording Protocol and tasks

All participants performed several task-based scenarios during the data collection. These tasks were chosen on the basis of their potential to induce social-, conversational- and mirthful laughter through both open and structured tasks. Participant specific task instructions and questionnaires were handed out before the start of the recording. Special effort was made by the experimenters to rehearse task instructions verbally, to make sure that participants understood all instructions since some tasks were more ambiguous in nature than others. After each task, the participants filled in a questionnaire containing questions on whether they laughed during the task and humour rating scales.

#### 3.4.1. Survival task

The first task consisted of a modified 'survival'-task and was used to habituate participants to the setting and experiment protocol. This task is often deployed in the field of psychology and was first deployed by NASA to test and foster collaboration skills. Participants were instructed to imagine that they were stranded on an uninhabited island and needed to survive for a couple of months. The participants were then asked to construct a list of 10 items they

wanted to take with them and both agreed on. The task ended after 3 minutes. Participants often constructed a list in less than the prespecified amount of time, which gave them some residual time for spontaneous chatter. During and after the task, spontaneous and conversational laughter often occurred. The experimenter would come in after three minutes and would tell the participants they were now allowed to proceed to the next task. Data of this task is missing in three sessions.

#### 3.4.2. Make the other laugh task

The 'make the other laugh'-task was developed by the experimenters with the goal of eliciting humour related and social laughter. Participants had either the instruction to make the other person laugh as much as possible or the instruction to follow the conversation. No further guidance or suggestions were given to the participants. These general instructions give room to more unstructured conversations and a diversity of laughter eliciting techniques. This task contains two 2 minute rounds, where the participants would switch roles between rounds. Participants were kept blind to their conversation partners instructions for each round and were asked not to discuss their instructions.

#### 3.4.3. Joke telling task

The third task involved three identical rounds where both participants told each other a joke. This 'joke telling'-task was designed to be a more controlled alternative to the 'make the other laugh'-task. It is designed to elicit a lot of humour related laughter in a more controlled setting with less conversations. At the start of the third task, the participants were instructed to tell jokes to each other. Participants selected three jokes from their stack of jokes that were collected beforehand by the researchers and were distributed randomly into two stacks.

A bonus round was introduced to the participants in which they were instructed to either tell a joke they had prepared in advance, look up a joke and prepare it on the spot or select a joke that was left in their stack. They were asked to pick something that they thought was funny and would make the other participant laugh. This bonus round gave participants the opportunity to use a joke that tailors more to their preferences and knowledge about what their conversation partner might enjoy.

After the initial data recording, participants participated in a brief emotion induction task involving pictures from the IAPS database (Lang et al., 1997). This will not be included in the current database since it is outside of the scope of our current goals.

### 3.5. Segmentation and Synchronization of Data Streams

Since the audio-, video- and physiological data are captured with multiple devices, manual synchronization was required. To ease this labour, the participants of the session were instructed to produce a simultaneous clap after counting down together before and after several tasks during each session. Participants would count down and then clap their hands at the same time. These simultaneous handclaps ensures a multimodal peak signal over multiple devices worn

by both participants. The post-experiment segmentation and synchronization of the data was done in several steps. First, audio- and video data were segmented by a research assistant using Adobe Premiere Pro, using hand-claps as an audiovisual marker for the start of tasks. This resulted in, after editing, 240 video clips (interlocutor A, interlocutor B and combined) and 152 audio clips (interlocutor A and interlocutor B). A log was kept for all the relative start- and end times of the video clips.

We then synced the video-, audio- and physiological data. A self-developed R code was deployed to find possible hand-clap signatures in the inertial movement data that was collected by the shimmers attached to the wrists of the participants. The resulting list of hand-clap signatures and the log with hand-clap interval timings were then used to manually identify and select the correct data points in the physiological signal. This method ensured synchronization with relatively high accuracy. The physiological data was further segmented in accordance with the audio-video segments. The shimmer devices we use are equipped with real time clock (RTC) stamps, therefore researchers interested in ECG- or GSR data will be able to use the already identified markers and synchronize these formats for their research purposes with relative ease.

### 3.6. Annotations

Several challenges in annotating laughter have been identified by the community. These challenges are partly the result of heterogeneous terminology used among laughter researchers with different backgrounds, and related to this, a lack of universally agreed upon definition of laughter that can be operationalized and applied by researchers interested in laughter as a social signal (Trouvain, 2003; Truong and Trouvain, 2012; Truong et al., 2019). This hampers the comparison of results and the development of theories on laughter and its expressive patterns.

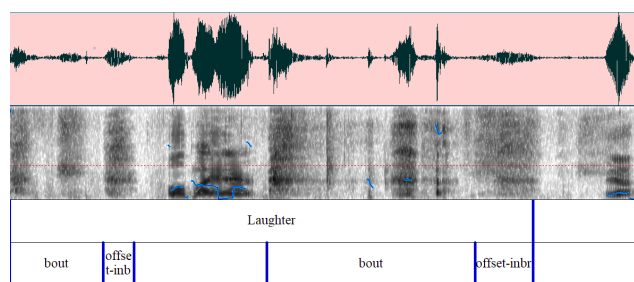


Figure 3: Example of laughter annotations in Praat. The first tier represents an initial rough annotation imported from ELAN while the second tier shows a finer grained annotation of laughter bouts and offset respiratory activity.

During the annotation of this data collection we use elements of a laughter annotation scheme which we recently introduced to make an effort towards a more universally and operationalized annotation of laughter (Truong et al., 2019). More specifically we use elements from the third tier of this annotation scheme. Laughter bouts and speech-laughes were annotated together with in- and out breaths that

were linked to the bouts start (onset) or end (offset). Similar as in Truong et al. (2019) we define a laughter bout as a laugh-like syllable or a sequence of laugh-like syllables that are produced in one exhalation phase. A bout can be, and often is separated by an inbreath at the start or end of the laugh bout. Speech-laugh are defined as a stretch of articulated speech with laughter interspersed. Truong et al. (2019) also defines smiled speech in this tier, however due to the interests of the authors at the time, this is not annotated. We used the Praat software package (Boersma and Weenink, 2020) to annotate the audio data. Instances where the annotation scheme still could not fit the data were resolved through several discussions among the authors. Examples of laughter annotations can be seen in figure 3.

The ‘make the other laugh’-task and the ‘joke telling’-tasks were annotated resulting in a total of 601 annotated laughter bouts, 168 annotated speech-laughes and 538 other on- or offset respiration events. For these annotated laugh events we extracted acoustic features commonly researched in relation to laughter, including duration, pitch and intensity features. Figure 4 gives a pairwise overview of the quantity and distribution of three important acoustic properties of laughter. Interestingly, this figure indicates diversity in how much and how laughter is produced within each conversation pair, which is in line with observations made by the experimenters. Some conversation pairs seem to produce much more or less laughter than others, while there also seems to be acoustic variability within pairs and differences between pairs.

As mentioned in the introduction, the annotated laughter segments in combination with the humour annotations from the participants can be used to study laughter in different settings. In the next section we will explore the relationship between perceived humour and the acoustic properties of laughter in two tasks, the more open and conversational ‘make the other laugh’-task and a more structured humour specific ‘joke telling’-task.

## 4. Preliminary analysis: The relationship between the acoustic properties of social laughter and humour perception

The relationship between perceived humour and acoustic properties of laughter has been researched before (Gervais and Wilson, 2005; McKeown and Curran, 2015). McKeown and Curran (2015) for example, explored the relationship between how intense participants perceived pre-selected laughter segments and how much these laughs were related to humour. Interestingly, they found a strong relationship between perceived humour and laughter intensity by these raters. In contrast to the data used by McKeown and Curran (2015), the MULAI database is more focused on laughter in a social setting and benefits from annotations from the participants in these settings.

The MULAI database thus equips us with data to further explore the work of McKeown and Curran (2015) from a different angle, by looking directly at certain acoustic properties of laughter that have been shown to influence the perception of laughter (Kipper and Todt, 2001) instead of relying on perceived laughter properties. It also takes humour annotations from annotators within the social context

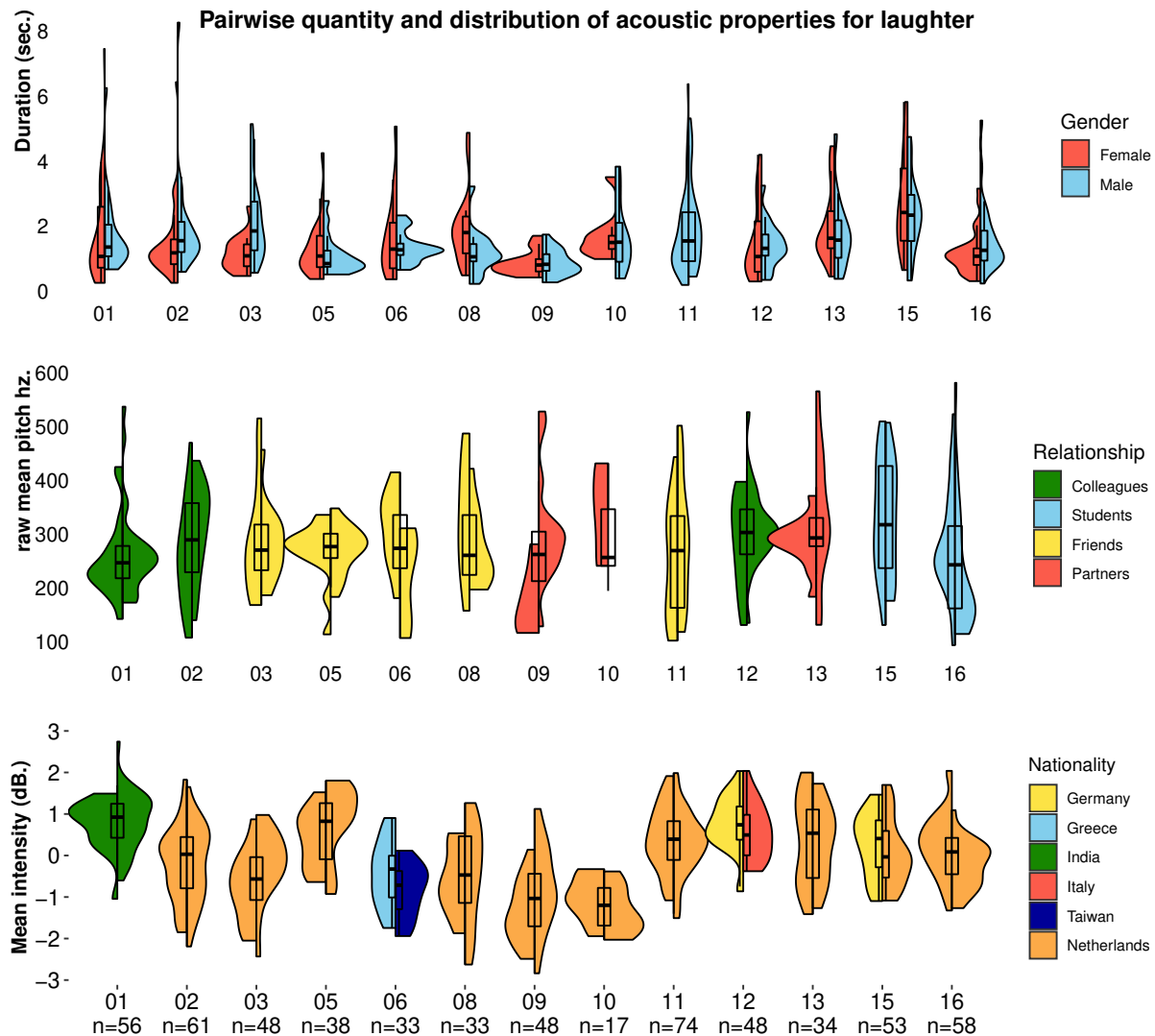


Figure 4: Violin plots showing the quantity of laughs and a distribution on three different acoustic properties of laughter for each of the 13 conversational pairs in the MULAI corpus. Each number on the x-axis represents a specific participant pair in our database and in addition the number of laughs for the pair is shown on the last row, the y axis represents an acoustic property depending on the row (duration, mean pitch, mean intensity). The left and right part of the violin plot represent the distributions of laughs for individual participants of the each pair respectively. The legends show information on important participant characteristics of each pair.

into account, who arguably have a better understanding and knowledge of this context than third party annotators. For the rest of this section we therefore explore the relation between acoustic properties of laughter and the perceived humour scores of the interlocutors. We will explore three questions:

1. How do self-rated humour and perceived humour by conversational partners correlate?
2. How do acoustic properties (pitch, intensity, duration) of laughter correlate with self-rated humour?
3. How do acoustic properties (pitch, intensity, duration) of laughter correlate with perceived humour?

We will exclude speech-laughter from the analysis due to its complex and ambiguous nature and focus on what we identify as ‘laughter bouts’ in the annotation section.

#### 4.1. Methods and statistical analysis

To investigate the relationship between participant humour self-ratings and the ratings they received from their conversation partner during the conversational tasks and the joke-telling tasks, we performed two Pearson Rank Correlation tests using R (Team, 2012).

Furthermore we investigated the relationship between humour self-annotation and acoustic properties of social laughter. We carried out a separate analysis for the two different task settings since they are different in their contents and context. Using Praat, we extracted three acoustic prop-

erties from annotated laughter; mean pitch, mean intensity and duration. We then performed a linear mixed effects analysis of the relationship between the extracted acoustic properties of laughter and the self annotated humour using the lme4 package (Bates et al., 2015) and lmerTest package (Kuznetsova et al., 2017). In our model we used mean pitch, mean intensity and duration as predictors. Since our observations and visualisations indicate that acoustic laughter properties vary between pairs, subjects and genders, we wanted to control for this and therefore included pairs, subject and gender as random intercepts for our model. To answer the question if perceived humour ratings of conversational partners are correlated with acoustic properties of produced laughter we used the same model, replacing self ratings with the ratings of conversational partners.

## 4.2. Results

First we explored the relationship between how people judged themselves and how their conversation partner judged them, Figure 5 shows a relative distribution of the difference scores between self ratings and those from conversational partners. A difference score of 0 indicates agreement among self- and perceived ratings. The histogram shows a larger percentage of participants agreeing on the humour ratings for the conversational tasks compared to those of the joke-telling tasks. Considering the joke-telling task, people tended to rate themselves less humorous compared to how their conversation partners scored them.

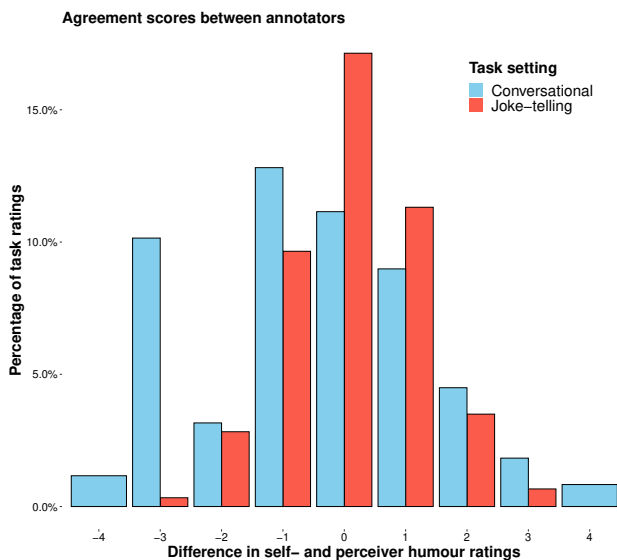


Figure 5: A histogram showing the difference in humour ratings between self ratings and perceiver ratings for both conversational and joke-telling tasks. A score of 0 means that self-raters and perceivers agreed upon how funny a person is. A positive difference score indicates a higher self rating were a negative difference score indicates a lower self rating compared to the perceiver.

Results from a Pearson rank test show a non-significant small positive correlation between the scores ( $\rho = 0.03$ ,  $p = 0.527$ ) during the conversational tasks. Humour annotation scores ( $\rho = 0.22$ ,  $p < .001$ ) for the joke-telling tasks show

a significant positive relationship. This shows that participants in joke-telling tasks seemed to score themselves similarly with how others scored them with respect to humour, where no such significant correlation was found for participants in the conversational task setting.

To answer if there is a correlation between humour self ratings and the duration, pitch and intensity of laughs we performed a linear mixed model analysis described in the previous section. The results of this model can be viewed in table 2 and show that in our sample there was no significant correlation between how participants rated themselves and the acoustic properties of the laughter they produced during the interactions in either of the two task settings. Worth mentioning is that the mean pitch of laughter within conversational tasks is close to being significantly correlated with humour self ratings.

Table 2: Estimates of the fixed effects (all acoustic properties of annotated laughter) on humour self-annotation using a Linear Mixed Model with subject, gender and task as random effects.

	$\beta$	SE	CI	P
<b>Conversation</b>				
Duration	-0.0797	0.0684	-0.2169/0.0531	0.2447
Pitch	-0.0011	0.0006	-0.0022/0.0006	0.0619
Intensity	-0.0093	0.0081	-0.0259/0.0065	0.2543
<b>joke-telling</b>				
Duration	-0.0106	0.0335	-0.0850/0.0525	0.7500
Pitch	0.0006	0.0005	-0.0004/0.0017	0.2300
Intensity	0.0014	0.0072	-0.0129/0.0155	0.8500

Second, we explored whether acoustic properties of laughter produced by a participant correlated with how their conversation partner judged their sense of humour. The results can be viewed in table 3. Interestingly, in both task settings there are correlations between how people laugh and how humorous they were perceived. In the conversational task setting, mean pitch and intensity of the laugh were significantly correlated and showed a positive correlation with the rating by the conversation partner. So the higher the pitch and intensity of laughs, the higher the humour ratings of the conversation partner during the conversational tasks. For the joke-telling task we see a significant negative correlation between the duration of laughs and the conversation partner humour ratings and a significant positive correlation between the intensity of laughter and conversation partner humour ratings. The longer the laughter surrounding the jokes, the lower the scores of conversation partners were. In contrast, the more intense the laughs were, the higher the perceived humour ratings by the conversation partners were. In the next section we will further reflect upon these results.

## 5. Discussion and future work

In this paper we introduced the MULAI database, a dyadic laughter database designed for researching the expressive patterns of social laughter in several different contextual settings. These settings include more open, conversational tasks and more structured joke telling tasks. The database

Table 3: Estimates of the fixed effects (all acoustic properties of annotated laughter) on humour annotation scores by the conversation partner using a Linear Mixed Model with subject, gender and task as random effects.

	$\beta$	SE	CI	<i>P</i>
<b>Conversation</b>				
Duration	0.0956	0.0527	-0.008/0.1993	0.0707
Pitch	0.0010	0.0004	0.0001/0.0019	<b>0.0239</b>
Intensity	0.0234	0.0062	0.0110/0.0358	<b>0.0002</b>
<b>Joke-telling</b>				
Duration	-0.0675	0.0327	-0.1320/-0.0024	<b>0.0413</b>
Pitch	0.0003	0.0005	-0.0007/0.0013	0.5712
Intensity	0.0163	0.0071	0.0023/0.0302	<b>0.0226</b>

offers modalities that are unique among laughter databases. In addition the MULAI database offers data on individual differences of the participants in the form of personality questionnaires and it offers perceived and experienced humour ratings from participants.

With its unique contents, the database lends itself for several research interests including but not limited to exploratory research on individual differences in laughter expressions, inter-individual laughter synchronization and mimicry and the link between perceived humour and properties of social laughter.

## 5.1. Discussion of results

The link between perceived humour and properties of social laughter was explored using the MULAI database, we tested whether specific laugh properties are related to how people judge their own humourousness and how their conversation partner judged them. The results show that for our sample, there was no significant correlation between the duration, pitch and intensity of laughs and how people judged themselves. However, the intensity of the person’s laugh does seem to correlate with how his or her conversation partner judges them in both task settings and is in line with the results of some of the previous research (McKeown and Curran, 2015). A difference between both studies is how the intensity ratings are obtained. Where the aforementioned research focuses on human annotation of intensity focuses through crowd-sourcing techniques, our study uses the acoustic intensity extracted from the speech signal. The pitch of laughter in the conversational task setting was positively and significantly correlated with how humourous conversational partners rated their interlocutor, which is in line with other studies who used pitch as an indicator for how mirthful laughter is (Petridis and Pantic, 2009).

There was no significant correlation for pitch and perceived humour in the joke-telling task. The duration of laughter is also significantly correlated to the perceived humour from conversation partners, the longer the laughs were, the lower perceived humour ratings seemed to be in our sample. This was only the case in the joke telling task, although for the conversational task the (positive) correlation is close to the significance threshold. We did not find an explanation for the correlation between laughter duration and perceived humour in literature.

Two general trends that seem to appear in our results are particularly interesting, First we see a difference between self ratings and perceived ratings and how these correlate with acoustic properties of laughter. One possible explanation for this is that self-raters have access to specific information and knowledge to determine their own ratings whereas his/her conversational partner does not have access to this information and knowledge, and therefore rely more on observable factors for their ratings. A different explanation of these observations can be found in the (modified) affect induction theory of laughter (Curran et al., 2018), which proposes that one of the main purposes of laughter is to induce positive affect in others. In addition they suggest that laughter signals on themselves are fairly ambiguous and perceivers use social context and the intensity of laugh signals to interpret the meaning of these laugh signals.

Second, it is interesting to see differences in the results are dependent on the task setting. The results seem to suggest that the correlations between acoustic properties of laughter and perceived humour ratings are task dependent, which could be pointing towards the social context-sensitive nature of laughter and other communicative expressions (Curran et al., 2018; Jansen, 2019).

We observe that the significant correlations in general are small, it could be the case that other unexplored factors show stronger correlations. A contextual factor that could be further investigated is the content of the jokes or the strategy that participants deployed to make their conversation partner laugh.

## 5.2. Future work

Several additions are planned as part of our future work with the database and are in part depended on available resources. First we plan to provide speech transcripts, annotations of strategies used by participants to make his/her conversational partner laugh and degrees of familiarity between participant pairs. These additions can be beneficial to researchers interested in leveraging these forms of context to study laughter and humour. Second, we plan to expand annotation for the ‘survival’-task, this collaboration task is rich in conversational laughter and includes some open non-task-related interactions which could be interesting to researchers looking to study laughter in natural conversations. Finally, since the authors are interested in the function and context of individual laughs, we are considering to add new annotations describing these aspects.

With the MULAI database we contribute to the research of humour perception, social- and contextual laughter and hope to encourage other researchers to further push our understanding of these topics.

## 6. Availability

The MULAI database will be made available to the research community after official publication. An individual license is needed for access to the questionnaire-, video-, audio- and physiological data of the MULAI database. Please contact the authors for updates and availability of the database.



## 7. Bibliographical References

- Bagby, M., Parker, J. D. a., and Taylor, G. J. (1994a). The Twenty-item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic research*, 38(1):23–32.
- Bagby, R. M., Taylor, G. J., and Parkers, J. D. A. (1994b). The twenty-item Toronto Alexithymia scale-II. Convergent, discriminant and concurrent validity. *Journal of psychosomatic research*, 38(1):33–40.
- Baker, R. and Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43(3):761–770.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using (lme4). *Journal of Statistical Software*, 67(1):1–48.
- Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The Montreal Affective Voices: A validated set of non-verbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, 40(2):531–539.
- Boersma, P. and Weenink, D. (2020). Praat: doing phonetics by computer.
- Campbell, N. (2007). Changes in Voice Quality due to Social Conditions. *ICPhS XVI*, (August):2093–2096.
- Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., Zhou, F., and De La Torre, F. (2009). Detecting depression from facial actions and vocal prosody. *Proceedings - 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*, (October).
- Curran, W., McKeown, G. J., Rychlowska, M., André, E., Wagner, J., and Lingenfelter, F. (2018). Social context disambiguates the interpretation of laughter. *Frontiers in Psychology*, 8(JAN):1–12.
- Dudzik, B., Jansen, M.-p., Burger, F., Kaptein, F., Broekens, J., Heylen, D. K. J., Hung, H., Neerincx, M., and Truong, K. P. (2019). Context for Emotion Perception in Audiovisual Databases for Automatic Affect Detection: A Survey. In *International Conference on Affective Computing and Intelligent Interaction (ACII)*, page 8. IEEE.
- Gervais, M. and Wilson, D. S. (2005). The Evolution and Functions of Laughter and Humor: A Synthetic Approach. *The Quarterly Review of Biology*, 80(4):241–277.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., and Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1):84–96.
- Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6):504–528.
- Hough, J., Tian, Y., Ruiters, L. D., Betz, S., Kousidis, S., Schlangen, D., and Ginzburg, J. (2008). DUEL : A Multi-lingual Multimodal Dialogue Corpus for Disfluency, Exclamations and Laughter. pages 1784–1788.
- Jansen, M. P. (2019). Communicative signals and social contextual factors in multimodal affect recognition. *ICMI 2019 - Proceedings of the 2019 International Conference on Multimodal Interaction*, pages 468–472.
- Jonason, P. K., Teicher, E. A., and Schmitt, D. P. (2011). The TIPI’s validity confirmed: Associations with sociosexuality and self-esteem. *Individual Differences Research*, 9(1):52–60.
- Kipper, S. and Todt, D. (2001). Variation of sound parameters affects the evaluation of human laughter. *Behaviour*, 138(9):1161–1178.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13):1–26.
- Lang, P., Bradley, M., and Cuthbert, B. (1997). International Affective Picture System ( IAPS ): Technical Manual and Affective Ratings.
- Mccowan, I., Carletta, J., Kraaij, W., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., Kadlec, J., Karaiskos, V., Lathoud, G., Lincoln, M., Lisowska, A., Post, W., Reidsma, D., and Wellner, P. (2005). The AMI Meeting Corpus. In *International Conference on Methods and Techniques in Behavioral Research.*, page 4.
- McKeown, G. and Curran, W. (2015). The relationship between laughter intensity and perceived humour. In *Proceedings of the 4th Interdisciplinary Workshop on Laughter and Other Non-verbal Vocalisations in Speech*, number April, pages 27–30.
- McKeown, G., Valstar, M., Cowie, R., Pantic, M., and Schröder, M. (2007). The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, 6(1):1–14.
- McKeown, G., Cowie, R., Curran, W., Ruch, W., and Douglas-Cowie, E. (2012). ILHAIRE laughter database. In *Proceedings of the LRECWorkshop on Corpora for Research on Emotion Sentiment and Social Signals*, number October.
- Niewiadomski, R., Mancini, M., Baur, T., Varni, G., Griffin, H., and Aung, M. S. H. (2013). MMLI: Multimodal multiperson corpus of laughter in interaction. In *International Workshop on Human Behavior Understanding*, pages 184–195.
- Owren, M. J. and Bachorowski, J.-A. (2003). Reconsidering the Evolution of Nonlinguistic Communication: The Case of Laughter. *Entomologia Experimentalis et Applicata*, 103(3):239–248.
- Petridis, S. and Pantic, M. (2009). Is this joke really funny? Judging the mirth by audiovisual laughter analysis. *Proceedings - 2009 IEEE International Conference on Multimedia and Expo, ICME 2009*, pages 1444–1447.
- Petridis, S., Martinez, B., and Pantic, M. (2013). The MAHNOB Laughter database. *Image and Vision Computing*, 31(2):186–202.
- Romero, E., Villar, P., Gómez-Fraguela, J. A., and López-Romero, L. (2012). Measuring personality traits with ultra-short scales: A study of the Ten Item Personality Inventory (TIPI) in a Spanish sample. *Personality and Individual Differences*, 53(3):289–293.
- Schuller, B., Müller, R., Eyben, F., Gast, J., Hörnler, B., Wöllmer, M., Rigoll, G., Höthker, A., and Konosu, H.

- (2009). Being bored? Recognising natural interest by extensive audiovisual integration for real-life application. *Image and Vision Computing*, 27:1760–1774.
- Shriberg, E., Dhillon, R., Bhagat, S., Ang, J., and Carvey, H. (2004). The ICSI Meeting Recorder Dialog Act ( MRDA ) Corpus. *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, pages 97–100.
- Tanaka, H. and Campbell, N. (2011). Acoustic Features of Four Types of Laughter. In *Proceedings ICPHS*, pages 1958–1961.
- Team, R. C. (2012). R: A Language and Environment for Statistical Computing.
- Thompson, H. S., Anderson, A. H., Bard, E. G., and Doherty- . . . , G. (1993). The HCRC Map Task corpus: natural dialogue for speech recognition. *Proceedings of the workshop on Human Language Technology*, pages 25–30.
- Thorson, J. A. and Powell, F. C. (1993). Sense of humor and dimensions of personality. *Journal of Clinical Psychology*, 49(6):799–809.
- Trouvain, J. (2003). Segmenting phonetic units in laughter. *Proceedings ICPHS*, pages 2793–2796.
- Truong, K. P. and Trouvain, J. (2012). Laughter Annotations in Conversational Speech Corpora - Possibilities and Limitations for Phonetic Analysis. *Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, pages 20–24.
- Truong, K. P., Trouvain, J., and Jansen, M.-p. (2019). Towards an annotation scheme for complex laughter in speech corpora. In *20th Annual Conference of the International Speech Communication Association*, page 5.
- Urbain, J., Bevacqua, E., Dutoit, T., Moinet, A., Niewiadomski, R., Pelachaud, C., Picart, B., Tilmanne, J., and Wagner, J. (2010). The AVLaughterCycle Database. In *Proceedings -The International Conference on Language Resources and Evaluation*, number Section 11, pages 2996–3001.
- Valstar, M. and Pantic, M. (2010). Induced Disgust, Happiness and Surprise: an Addition to the MMI Facial Expression Database. In *proceedings -The International Conference on Language Resources and Evaluation*, pages 65–70.
- Ypofanti, M., Zisi, V., Zourbanos, N., Mouchtouri, B., Tzanne, P., Theodorakis, Y., and Lyrakos, G. (2015). Psychometric properties of the International Personality Item Pool Big-Five personality questionnaire for the Greek population. *Health Psychology Research*, 3(2).
- Zheng, L., Goldberg, L. R., Zheng, Y., Zhao, Y., Tang, Y., and Liu, L. (2008). Reliability and Concurrent Validation of the IPIP Big-Five Factor Markers in China: Consistencies in Factor Structure between Internet-Obtained Heterosexual and Homosexual Samples. *Pers Individ Dif.*, 45(7):649–654.