# Perception and Automatic Recognition of Laughter from Whole-Body Motion: Continuous and Categorical Perspectives

**Published in:**
IEEE Transactions on Affective Computing

**Document Version:**
Peer reviewed version

**Queen's University Belfast - Research Portal:**
[Link to publication record in Queen's University Belfast Research Portal](#)

# Perception and automatic recognition of laughter from whole-body motion: continuous and categorical perspectives

Harry J. Griffin, Min.S.H Aung, Bernadino Romera-Paredes, Ciaran McLoughlin, Gary McKeown, William Curran and Nadia Bianchi-Berthouze

**Abstract** — Despite its importance in social interactions, laughter remains little studied in affective computing. Intelligent virtual agents are often blind to users' laughter and unable to produce convincing laughter themselves. Respiratory, auditory, and facial laughter signals have been investigated but laughter-related body movements have received less attention. The aim of this study is threefold. First, to probe human laughter perception by analyzing patterns of categorisations of natural laughter animated on a minimal avatar. Results reveal that a low dimensional space can describe perception of laughter "types". Second, to investigate observers' perception of laughter (hilarious, social, awkward, fake, and non-laughter) based on animated avatars generated from natural and acted motion-capture data. Significant differences in torso and limb movements are found between animations perceived as laughter and those perceived as non-laughter. Hilarious laughter also differs from social laughter. Different body movement features were indicative of laughter in sitting and standing avatar postures. Third, to investigate automatic recognition of laughter to the same level of certainty as observers' perceptions. Results show recognition rates of the Random Forest model approach human rating levels. Classification comparisons and feature importance analyses indicate an improvement in recognition of social laughter when localized features and nonlinear models are used.

**Index Terms** — H.5.mMiscellaneous; I.2.6.g Machine learning; I.5.4.d Face and gesture recognition; J.4.b Psychology

—————————— ◆ ——————————

## 1 INTRODUCTION

Recent efforts to create embodied conversational agents have realized the importance of using non-verbal components of human communication [1]. However, laughter – arguably the most important social signal for lubricating the flow of social interaction – has been largely overlooked. The association of laughter with humour, joviality, and frivolity has probably led to it being ignored as a topic for serious scientific enquiry; this is especially likely in human computer interaction (HCI) given the agenda of the Turing era in pursuing the more intellectual capacities of thought in computer-based intelligence [2]. However, laughter appears to be one of the most important elements in human conversation when it comes to developing rapport, guiding the direction of conversations and repairing conversations when they break down.

The increasing use of virtual agents and robots in entertainment, collaborative, and support roles places ever greater demands on their ability to detect users' emotional state from various modalities (body movements, facial expressions, speech) and produce emotional displays. This is particularly true in socially complex human-computer interactions such as education, rehabilitation and health scenarios. In these situations, emotionally expressive agents are much preferred by users [3].

Laughter is a ubiquitous and complex signal that remains relatively uninvestigated, in contrast to studies on other emotional expressions such as smiling [4]. This is particularly true with respect to body expressions of laughter. Due to the range of vocal and physical expressions of laughter, its detection and synthesis are very challenging. Laughter does more than express hilarity. It can convey negative and mixed emotions and act as an invitation to shared expression [5]. Hence, the ability to recognize and produce appropriate laughter at the correct level of intensity in response to a user's emotional signals – including laughter itself – would be a dramatic step forward in the realism and efficacy of virtual agents.

In this study we investigate the perception of laughter type from body movements and lay the groundwork for laughter type recognition from these cues. This study makes three contributions: first, by analyzing patterns of perceptual responses to whole-body animations of natural laughter on a minimal avatar, it increases our knowledge of visual laughter perception and how semantic categorical labels that are assigned to laughter may derive from continuous perceptual variation; second, by identifying body movements that are perceived as indicative of different types of natural and fake laughter, it informs more convincing animation of laughter in avatars, which will increase their perceived conversational authenticity and emotional range; third, it investigates if it is possible to automatically discriminate between different types of laughter by comparing a wide range of automated recognition methods. This work extends [6] by providing: (i) not only

————————————————

- *H.J.Griffin, M.S.H.Aung, B.Romera-Paredes, C.McLoughlin and N.Bianchi-Berthouze are with the UCL Interaction Centre, University College London, Gower Street, London, WC1E 6BT, United Kingdom*
- *G.McKeown and W.Curran are with the School of Psychology, Queen's University Belfast, BT7 1NN, United Kingdom*

an analysis of discrete categories of laughter but also an exploration of its dimensional components; (ii) a more comprehensive analysis of features for both upper and lower body; (iii) a more thorough analysis of recognition models for laughter perception and their relevant features.

## 1.1 The role of laughter

Laughter is often assumed to be simply the behaviour that occurs in response to a humorous event or stimulus; however, such a simplistic view misses many of the important roles and functions that laughter provides throughout the course of human interaction. It occurs very frequently in human interactions and is more often associated with simple conversational phrases than with explicit attempts at humour [7].

Laughter has evolutionarily ancient origins and appears to have been present in our last common ancestor with the great apes. Laughter-like phenomena are found in higher primates and reconstructions of phylogenetic trees based on the acoustic properties of these laugh-like signals mirror those based on genetic information [8]. The reported production of laugh-like social signals in rats [9] suggests that laughter may stretch even further back into the mammalian past. Human laughter seems to have arisen from the "relaxed open mouth display" found in many primate species. The open mouth display is associated with roughhousing style play, within which it seems to signal a social bonding play frame allowing physical boundaries to be tested and aggressive play to occur safely while minimizing the likelihood of actual violence and harm [10] (possibly highlighting the importance and role of the body in the expressions of laughter). Darwin [11] claimed that a continuum exists between human laughter and smiling, and Van Hooff [12] has taken this argument a stage further by arguing that what has now become a continuum in humans was preceded by separate evolutionary paths to smiling and laughter. Laughter also appears very early in human development, and is evident in most children at about four months of age. Gelastic seizures have been observed in neonates, suggesting that the apparatus to produce laughter exists at birth [13].

Laughter's relationship with humour is particularly intriguing. The evolutionarily ancient reflex-like breathing behaviour of laughter can be rapidly elicited in response to the often complex cognition required by humour. Given a joke that requires linguistic processing, an identifiable response to the humour occurs within 700-1000ms; laugh-related physical responses follow quickly after and are associated with differences in frontal cortex morphology [14], [15].

Its evolutionary ancient roots, early developmental appearance, and the speed with which it appears in response to humour suggest that laughter is a deeply embedded social signal within the repertoire of human affective behaviour and one that has a strong social nature. One research area that has studied laughter's role in social interaction in depth is conversation analysis. Laughter has received attention within this methodological approach from Jefferson, Sacks, and Schegloff [16] and, more recently, Glenn

and Holt [14], [5], [17] who view laughter as a social lubricant that facilitates the sequential dynamics of social interaction. It serves an almost punctuation-like role within human social conversation by facilitating topic change and termination, and conversational repair [5]. It achieves these goals by being ambiguous with respect to the propositional content of the conversation, while sending a social bonding signal. This creates a 'safe area' in a conversation that is detached from the content; topics can then be terminated or changed, or interlocutors can take divergent interpretations of meaning of the laughter before moving on to a new topic. Laughter also seems to function similarly to the great ape social signal in creating a play frame, but in a linguistic manner. For example issues of social impropriety can be safely explored and then easily retracted as non-serious. This may explain the important role laughter plays in flirtation and in banter surrounding negotiations. There also appear to be important sequential dynamics to laughter. Invitations to laugh can be accepted or denied; accepting an invitation leads to shared laughter that allows for changes in the direction of a conversation. In addition the sequential dynamics of laugh ordering appears to function differently in dyadic conversations, in which the speaker often laughs first, as opposed to groups of three or more where laughing first may be seen as self-aggrandizing [18].

An important issue in the function of laughter is the existence of different laugh categories. We can think of many words that can serve as qualifiers for laughter, such as joyful laughter, taunting laughter and schadenfreude laughter [19]; however it is far from clear that these mirror qualitatively different types of laugh and do not instead represent some blend of the listener's emotional state and the context in which a laugh occurs. Functionally, we may think of the two ends of the laughter and smiling continuum as distinct laughter types. Gervais and Wilson [20] amongst others have argued for two types; a hard-to-fake spontaneous signal that they term the Duchenne laugh (due to its inclusion of the FACS Action Unit 6 in a similar way to the Duchenne smile) and a non-Duchenne type of laughter. The former is more associated with spontaneous responses to humour, and the latter with more volitional social motives. However the clarity of the Duchenne distinction has been questioned [21] and it may be better to think of these two types of laughter as falling along a continuum as Darwin and van Hooff suggested [11], [12].

It is clear from the previous discussion that laughter has numerous important social roles within human interaction. Removing physical and nonverbal elements of communication, such as laughter, from human-computer interactions – leaving only the "intellectual" textual aspects of communication – will place the computer at a distinct communicative disadvantage rather than proving to be an advantage as Turing [2] surmised. The importance of including social signals in interactions between conversational avatars and humans becomes increasingly obvious; and perhaps the most important of these is the one that is often thought of as frivolous: laughter. If we are to create a computer interface that can interact with humans in the way humans interact with each other then we must under-

stand the role and function of this pervasive feature of human communication. We must be able to recognize laughter when it occurs, understand where it should be placed within conversational interactions, and be able to synthesize realistic auditory and visual behaviors related to laughter episodes, both at the level of whole body movement and in facial expressions.

## 1.2 Synthesis and recognition of laughter

There have been few studies on synthesizing laughter in virtual agents, most of which have focused on acoustics and the face [22], [23], or laughter accompanying videos of humanoid robots [24]. Urbain et al. present a laughter machine that is able to recognize laughter from sounds and give a response [25]. The distinctive respiration patterns of laughter have been widely corroborated [26] and integrated into anatomically inspired models of laughter [27]. Recently, the coordination of virtual agents' laughter respiration behaviour with other visual cues has been investigated [28]; however, this work is mainly based on hilarious laughter. A further difficulty for synthesis of laughter-related body movements is that stereotypical laughter actions, for example, clutching ones abdomen, rocking back and forth, slapping one's leg, are well known but may be seen as exaggerated and unnatural.

Work on automatic recognition of laughter has also started to emerge but, as with the synthesis of laughter, has mostly focused on the acoustic modality e.g., [29], [30], [31], [32], [33], [34] and more recently on the combination of face and voice cues [35], [36], [37]. Fukushima et al. used electromyographic sensors to measure diaphragmatic activity, which drives laughter vocalisations, to detect laughter in people watching television [38]. Less attention has been given to body laughter expressions. Whole-body postural changes and peripheral gestures associated with different types of laughter remain almost uninvestigated.

More recently, there has been interest in creating automatic classifiers able to differentiate laughter types. To this end, motion descriptors based on energy estimates, correlation of shoulder movements and periodicity to characterise laughter have been investigated [39]. Using a combination of these measures a Body Laughter Index (BLI) was calculated. The BLIs of 8 laughter clips were compared with 8 observers' ratings of the energy of the shoulder movement. A correlation, albeit weak, between the observers' ratings and BLIs was found. There has been growing evidence supporting the possibility of automatically discriminating between different emotions from various modalities: acoustics [40], facial expressions [41] and body movements [42], [43], [44], [45], [46], [47]. Galvan et al. [48] investigated automatic discrimination of five types of acted laughter: happiness, giddiness, excitement, embarrassment and hurtful. Actors were asked to enact these five emotions using both vocal and facial expressions while they were video-recorded. The video clips were labelled by expert observers who were also made aware of the intention of the actors. The results showed that automatic recognition based only on the vocal features reach higher accuracy (70% correct recognition) than when using both facial

and vocal features (60% correct recognition) or facial features alone (40% correct recognition). While, on the basis of these results, the authors argue that vocal expressions carry more emotional information than facial expressions, it should be noted that the actors were asked to try to keep the head as still as possible so that it would be always frontal to the video camera. This may have constrained the way people expressed their laughter through their faces and head movements. In addition, the fact that the expressions were acted also raises the questions of how naturalistic they were. One could argue that we are better at acting an expression through our voice, since we can hear our own voice, whereas we cannot see our own face. This is particularly true when the actors are not professionals but lay people. It is also the case that, as the actors were given pre-ordained categories of laughs to generate, they could have made the laughter maximally distinguishable between categories imbuing them with extra cues related to the given category. This creates a circularity that does not exist in more naturalistic stimuli. In this study we investigate perception of laughter type from body movements and lay the groundwork for laughter type recognition from these cues.

## 2 MOTION DATA COLLECTION

Users' perception of laughter-related body movements were investigated in a forced-choice perceptual experiment. Body movements captured during different types of natural and acted laughter were used to animate an avatar. Observers categorized the animations as hilarious, social, awkward, fake, or non-laughter. Naive observers' categorizations were used to allow analysis of the perception of body movements in the absence of other modalities, for example, verbal, facial, and in the absence of knowledge of the eliciting stimulus and context.

### 2.1 Laughter collection

Nine pairs of participants took part in a motion capture recording session. The analysis of inter-participant laughter behaviour, though fascinating, was beyond the scope of this study as we had access to only one motion capture suit. The movements of one member of each pair (subjects - 3 male, 6 female, mean age 25.7) were captured using an inertial motion capture suit (Animazoo IGS-190). The suit was modified to maximize its sensitivity to spine and shoulder movements. Tasks to elicit laughter in both standing and sitting postures included word games such as tongue-twisters, collaborative games (Pictionary) and humorous videos [49]. Laughter also occurred during conversation during "rest" periods. The subjects also produced fake laughter on request.

### 2.2 Stimulus preparation

Using video recordings of the motion capture session, we identified 508 laughter segments, and 41 non-laughter segments that contained other behaviour such as talking or reading. The motion capture data from these segments were used to animate a minimal avatar defined by the positional co-ordinate triplets of 26 anatomical points over the whole body (Fig. 1). The anatomical proportions were

the same for all animations; joint angle values were transferred directly from all models to a standardised skeleton without correction for differences in segment lengths. Viewing angle was standardized to a slightly elevated ¾ viewpoint, although models were free to walk and turn in the standing tasks.
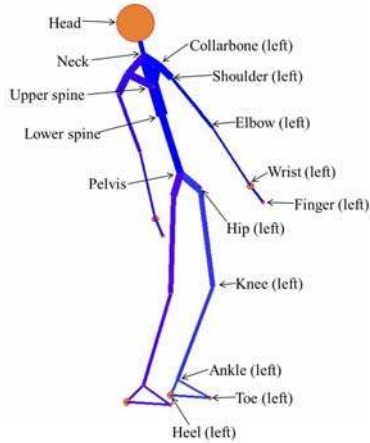


Fig. 1. Skeleton used to animate motion capture for the perceptual experiment. Nodes marked "left" are repeated on the right side of the avatar.

We eliminated animations with grotesque or impossible movements caused by motion capture errors or by disparities in body segment length between models and the standardised skeleton. One hundred and twenty six animations (112 laughter, 14 non-laughter, 69 standing, 57 sitting, mean duration = 4.1 s, SD = 1.8 s) were then selected as stimuli for the perceptual phase. Our aim was to produce sufficient segments in each observer-determined category to allow statistical analysis of body movements between categories, using the observers' categorisations as our ground truth (see Section 3). To this end, laughs were selected that spanned a wide range of experimenter-judged triggers, contexts and intensities. For an analysis of the relationship between observer judgements and experimenter judgements, which were made using such privileged contextual and multimodal information, see [50]. These stimuli are also available at the ILHAIRE laughter database: www.qub.ac.uk/ilhairelaughter [51].

## 3 PERCEPTUAL STUDY

Thirty-two observers (17 male, 15 female, mean age 33.0, SD = 14.1) viewed the clips of the animated avatar in random order and categorized each clip as *hilarious*; *social* (back-channeling, polite, conversational laughter); *awkward* (involving a negative emotion such as embarrassment or discomfort on another's behalf); *fake*; or *non-laughter*. These 5 categories are not assumed to be a universal, exhaustive taxonomy of laughter types, rather they were chosen as broad amalgamations of finer grained categories assigned by observers to a naturalistic video laughter database in a previous study [51]. Full definitions of the laughter categories given to the observers are in Appendix 1. No audio was presented with the animations.

The results of these perceptual judgments were used to group the animations by the most commonly chosen (*modal*) category. The number of animations per modal category was as follows: hilarious = 22, social = 41, awkward = 6, fake = 4, non-laughter = 44, tied = 9. All ties were two-way: 2 x hilarious/social, 1 x hilarious/awkward, 1 x social/awkward, 4 x social/non-laughter, 1 x fake/non-laughter. The mean number of observers who selected the modal category was 13.8 (SD = 4.3) with a maximum agreement of 29 of the 32 observers. Examples of frames from animations are shown in Fig. 2.
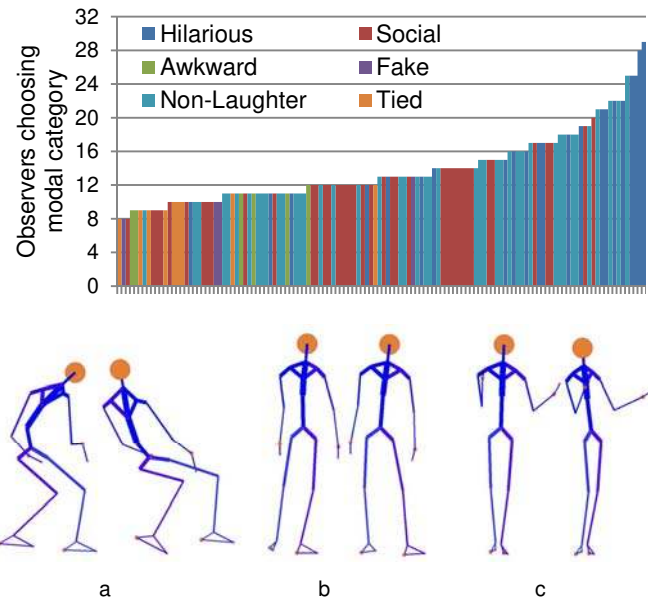


Fig. 2. Example frames from stimuli: (a) extremes of torso leaning in a sitting animation modally categorized as hilarious laughter; (b) foot shuffling/weight shifting in a standing animation modally categorized as awkward laughter; (c) whole body rotation while reading in a standing animation modally categorized as social laughter.

Fig. 3 shows the number of observers who chose the modal category for each animation. A clear pattern emerges of relatively high levels of agreement for stimuli modally categorized as hilarious or non-laughter, moderate agreement for social laughter and low agreement for awkward or fake laughter. A one-way ANOVA showed a significant effect of modal category on the number of observers who chose the modal category ($F(4,112)=9.57$, $p<.0001$); tied stimuli were excluded from this analysis, because, to be tied, half the observers or fewer must chose each modal category.

Fig. 3. Categorisations in modal category for each animation out of 32 observers. The colour of the column indicates the modal category.

Fig. 4 shows the pattern of voting for animations. The different levels of certainty according to the modal category can be seen, but the allocation of choices to non-modal categories also clearly differs between stimuli, for example, stimuli modally categorized as social show a fairly even distribution of choices allocated to other categories, but stimuli modally categorized as hilarious show very few observers choosing non-laughter and vice-versa.
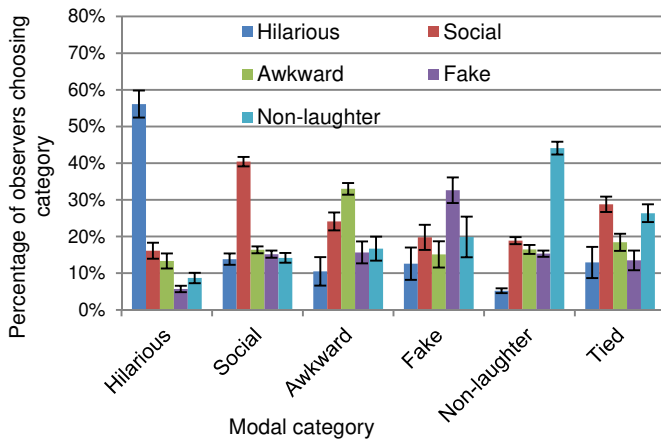


Fig. 4 - Observers choosing each category by modal category. Errors bars are ± 1 SE. Number of animations in each category: hilarious = 22, social = 41, awkward = 6, fake = 4, non-laughter = 44, tied = 9

# 4 DIMENSIONAL ANALYSIS OF LAUGHTER PERCEPTION

The uneven distribution of votes allocated to non-modal categories across stimuli implies that there may exist other perceptual categories or dimensions that were used by the observers for these categorisations. It may be that the categorisations reported by observers were the result of their fitting our, somewhat pragmatic, range of categories onto their fundamental perception of the laughter. Evidence for this hypothesis comes from a study by McKeown et al. using this same stimulus set [50]. In their paper, in addition to categorizing the stimuli according to the same 5 categories, observers also rated the intensity of the laughter in each animation. Modal category and intensity were found to be very strongly related, with intense laughs

categorized as hilarious, moderately intense laughs categorized as social and weaker laughs categorized as awkward or fake (non-laughs, by definition, had zero intensity). One explanation is that intensity is actually judged prior to category and categorization is then based on intensity according to whatever categorization schema is required. To test this hypothesis we applied Principal Component Analysis to our stimuli, treating each stimulus as a point in 5D space according to the proportion of choices allocated to the 5 categories. The results of the PCA are shown in Fig. 5 (Fig. 5-left – 1st and 2nd components; Fig. 5-centre - 1st and 3rd components; Fig. 5-right – 2nd and 3rd components). The amount of variance explained by the resulting components is: 1st component = 58.66%; 2nd component = 27.66%; 3rd component = 8.58%. The loadings of the original 5 dimensions on the new axes are shown in table 1.

TABLE 1 – Loading of 5 original laughter category dimensions on components emerging from Principal Component Analysis

| | 1st component | 2nd component | 3rd component |
|---|---|---|---|
| Hilarious | 0.809 | -0.355 | -0.139 |
| Social | -0.0354 | 0.693 | -0.484 |
| Awkward | -0.0612 | 0.130 | 0.811 |
| Fake | -0.147 | 0.131 | 0.0943 |
| Non-laughter | -0.565 | -0.599 | -0.282 |

We suggest that these results favour a dimensional interpretation of human laughter perception. From Fig. 5 and Table 1 we can see that the first component strongly mirrors the findings of McKeown et al. [50] and most likely reflects an intensity dimension ranging from non-laughter (zero intensity) to hilarious laughter (high intensity). Interpretation of further components is necessarily speculative since only percieved intensity was recorded by McKeown et al. The second component appears to separate social laughter from other categories. This may reflect a higher-level social interpretation of the avatar's movements. The low variance explained by the remaining components makes their interpretation yet more difficult, although the high loading of the awkward dimension on the 3rd component is noteworthy. Further work on this area is necessary to fully reveal the nature of these dimensions.
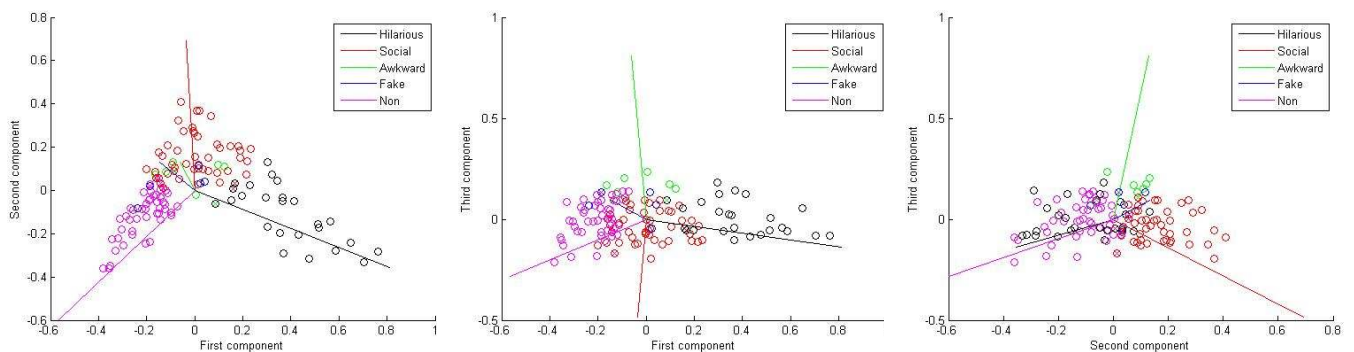


Fig 5 – Results of Principal Component Analysis of observer categorisation of animations. Each point represents an animation, colour coded by the modal category chosen by observers.

## 5 BODY MOVEMENT ANALYSIS

The modal laughter category selected by observers acted as a ground truth for statistical analysis of body movement features [42]. Segments tied for the modal category were excluded from the body movement analysis, as were segments for which the modal category was selected by less than ⅓ of observers (<11/32). Too few awkward (N=6) and fake (N=4) remained so these were excluded from further analysis. Fifty-three standing segments (6 hilarious, 23 social, 24 non-laughter) and 41 sitting segments (15 hilarious, 9 social and 17 non-laughter) were entered into the analysis of body movements.

The number of movement features that can be analyzed is large, unmanageably so if interactions of multiple features are considered. Therefore, our selection of features was based on previous studies [27], [52] and observers' comments in post-experiment interviews on which features they used in categorizing laughter. These included postural changes e.g., angle of the spine at each node and collectively across all spine nodes, and gestures e.g., moving a hand toward the face or abdomen. This feature set improves on Griffin et al. [6] by adding lower-limb features, energy [39] and smoothness measures, and refining shoulder movement measures. Fifty-seven features, including duration, were entered into the analyses.

The full list of features is presented in Table 2 (first column). Distances (maximum, minimum and range) are Euclidean distances calculated in 3D space. The angle at a node is the inner angle between the 3D vectors that define the two segements adjoining at the node. Energy is calculated as the sum of the square of the angular velocity for a given node [45]. Smoothness of a motion trajectory is calculated as the length of the arc produced by the Fourier magnitude spectrum of the trajectory's speed profile [53]. In addition, the independent directional components in the Anterior-Posterior (*A-P*) and Lateral body planes of each of the spine and head segments are also calculated. This is to measure the directional contribution to the overall posture by each segment independently. The power spectral density of the shoulder motion in the Superior-Inferior (*S-I*) body plane between 4-6Hz is also calculated as movements around 5Hz have been shown to occur in muscular and vocal patterns in laughter [52].

Feature analysis of the motion capture data was based on position coordinate triplets of the anatomical nodes (Fig. 1). The 57 chosen features for hilarious, social and non-laughter segments were entered into separate one-way ANOVAs for standing and sitting segments (independent variable was observers' modal categorization). If Levene's test for homogeneity of variance returned $p < .05$, Welch's correction for nonhomogeneity was applied. Planned contrasts tested differences between laughter and non-laughter (hilarious and social vs. non-laughter) and between laughter types (hilarious vs. social).

Many features in all body areas showed significant results and there is overlap in features signicant for sitting

and standing animations (Table 2). Only main effects for angles at the upper and lower spine node, the sum of spine angles and the shoulder energy survive correction for multiple comparisons in sitting and standing laughter.

Posture appears to influence which movements are most relevant for laughter recognition, most noticeably vertical shoulder movements and rotations are relevant for discrimination of sitting laughter, but largely not for standing laughter. Closer inspection of the data reveals a strikingly consistent pattern. For all significant effects on maximum and range values of features, the value for hilarious laughter was greater than that for social laughter which was greater than that for non-laughter. For minimum values that were significant this pattern is reversed, such that the value was lower for hilarious than for social laughter, which was lower than that for non-laughter. These results indicate that for animations that were perceived as laughter, and specifically hilarious laughter, movements were larger: hands were extended further, the spine was extended through greater angles, the shoulders were rotated and moved vertically more, and there was more energy in the joints where this was measured. The pattern for minima indicates that for hilarious laughter hands were brought closer to the head and that the spine was bent further forward. Interestingly, significant effects for minimum values only appeared for standing laughter.

For sitting laughter, there was a significant effect of stimulus duration and stimuli perceived as laughs were longer than those perceived as non-laughter. Mean (SD) durations for sitting stimuli were: hilarious = 5.08s (1.42); social = 4.12s (1.45); non-laughter = 2.88s (1.05). For standing laughter no main effect of modal category on duration appeared, a marginally significant trend appeared for stimuli perceived as laughter to be longer than those perceived as non-laughter. Mean (SD) durations for standing stimuli were: hilarious = 4.68s (1.75); social = 4.28s (1.45); non-laughter = 3.40s (1.83). These results are concordant with the perception of laughter, particularly hilarious laughter, resulting from stimuli with greater movement since large movements, for example, a large spine bend often required some seconds to be completed.

## 6 AUTOMATIC RECOGNITION

Due to there being relatively few studies on automatic recognition of laughter category from body movement information, we implement wide ranging experiments with a broad feature set, several linear and nonlinear models based on various learning paradigms each evaluated using four metrics. Moreover, we seek to understand relative feature importance using white box models as well as methods to analyze black box models. This differs from separate statistical tests in that feature importance estimates are relative to each other within the feature set.

Nine supervised learning models were implemented for comparative evaluations in the automatic recognition of laughter type. Among this set, three methods have functionalities that quantify relative importance between all of the features. All models are trained targeting the outcomes of all the observer's categorizations instead of targeting the

modal category alone.

TABLE 2 – List of body movement features and ANOVA analysis results. See Table legend in next column.

| Features returning 3 values  (Indices) | Max | Min | Range |
|---|---|---|---|
| Distance between hands (1,2,3) | | | a,B x,y |
| Distance between hands and hip (4,5,6) | x,Y | | A,b x,y |
| Distance between hands and head (7,8,9) | | x,y | A,B x,y |
| Angle at upper spine joint (10,11,12) | a,b | x,y | A,B,c X,Y,z |
| Angle at lower spine joint (13,14,15) | | x,y | A,B,c X,Y |
| Angle at neck joint (16,17,18) | a,b | y | A,B x,y |
| Sum of all spine angles (19,20,21) | | x,y | A,B,c X,Y |
| A-P component, lower spine → upper-spine segment direction (25,26,N/A) | x,y | | A,B X,Y |
| Lateral component, lower spine → upper-spine segment (27,28,N/A) | a,b x,Y | | A,B,c |
| A-P component, upper-spine → neck segment direction (29,30,N/A) | x,y,z | | A,B X,Y,z |
| Lateral component, upper-spine → neck segment (31,32,N/A) | x,y | | A,B,c x,y |
| A-P component, neck → head segment direction (33,34,N/A) | a,b x,y | | A ,B,c X,Y |
| Lateral component, neck → head segment (35,36,N/A) | a,b x,y | | A ,B,c x,y |
| Rotation of shoulders relative to hip line (37,38,39) | a,b | X,y,z | A,B,c |
| **Features returning single value for each animation** | | | |
| Energy at elbow joint (max of left and right) (22) | | | |
| Energy at shoulder joint (max of left and right) (23) | | | A,B,c X,y,Z |
| Energy at neck joint (24) | | | A,B,c x,y |
| Azimuthal rotation of shoulders in global space (40) | | | A,B |
| Energy at hip joint: max of left and right (41) | | | a,b,c |
| Energy at knee joint: max of left and right (42) | | | b,c |
| Energy at upper spine joint (43) | | | a,b x,y |
| Energy at lower spine joint (44) | | | a,b,c x |
| Ankle trajectory distance: max of left and right (45) | | | a,B,c |
| Knee trajectory distance: max of left and right (46) | | | A,B,C x,y,z |
| Smoothness of shoulder trajectory relative to upper spine: mean of left and right (47) | | | a |
| Range, S-I shoulder displacement: mean of left and right (48) | | | A,B |
| Correlation, left and right shoulder S-I displacement (49) | | | |
| Power (4-6Hz), S-I shoulder displacement: mean of left and right (50) | | | |
| Duration | | | A,B y |

TABLE 2. – Legend.
Significant results are indicated by letter. Lower case indicates $p < .05$, upper case indicates lower than threshold corrected for multiple comparisons $p < .05/57$ i.e., $p < .000877$. A-P = anterior-posterior, S-I = superior-inferior. Body movement feature indices are given in parentheses for reference to automatic recognition results (section 6, Fig. 7); features marked "N/A" were included in this analysis but not used in the automatic recognition.

A/a – main effect of modal category for sitting laughter
B/b – laughter vs. non-laughter contrast for sitting laughter
C/c – hilarious vs. social laughter contrast for sitting laughter
X/x – main effect of modal category for standing laughter
Y/y – laughter vs. non-laughter contrast for standing laughter
Z/z – hilarious vs. social laughter contrast for standing laughter

Doing this provides two advantages: first, a measure of uncertainty for each stimulus is preserved. For example, Fig. 6 shows a near uniform number of categorizations between all five laughter types for stimulus 1; in contrast stimulus 2 shows a high value for hilarious compared to the other laughter categories and can be deemed as perceptually less ambiguous. This therefore avoids the need to make weak assumptions regarding the validity of the ground truth for these ambiguous cases (e.g. Fig. 6 – Stimulus 1) from a supervised learning standpoint. Secondly, all 126 stimuli can be utilised since no instance needs to be discarded due to tied modal categories or due to a weakly modal category e.g., chosen by fewer than 11 of the 32 observers. All features listed in Table 2 are used; except for stimulus duration – which is experimenter-determined, rather than inherent in the avatar movement - and the ranges of the directional components (marked *N/A* in table 2); since these features are normalized, the small amout of information that the corresponding ranges would add does not justify the extra dimensionality that would result from their inclusion. This leads to a full feature set with $d = 50$ dimensions.
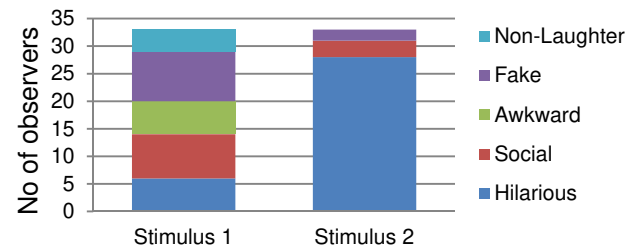


Fig 6 – Examples of categorization patterns of stimuli showing low (left) and high (right) agreement.

## 6.1 Supervised learning models

### 6.1.1 Standard models

Formally, the problem consists of a set of $T = 5$ supervised regression tasks, one for each type of laughter (including

non-laughter). We denote by $x^i \in \mathbb{R}^d$, the vector of instances describing instance $i$. We define the matrix of all of the training instances as $X = \mathbb{R}^{d \times m}$, where $m$ is the number of training instances and $d$ being the dimensionality of the data. A distinct label $y_t^i$ is provided for each task $t \in \{1..T\}$ for each instance $i$, taken from the frequency of observations. We denote $Y_t \in \mathbb{R}^m$ as the vector label $t$ for all instances. We also denote the corresponding model predicted output as $\hat{y}_t^i$.

### 1)      $k$-Nearest Neighbour ($k$-NN)

This is a simple model which assigns the value of the predicted output based on the $K$ nearest training instances in the data space. We attain the necessary multiple outcome vector by using the means of the labels from the $K$ nearest neighbours $N_K(x) \subset \{1,2..m\}$ of a given instance $x$. For a test instance $x$, the prediction is calculated by:

$$\hat{y}_t^i = \frac{1}{K} \sum_{i \in N_K(x)} y_t^i.$$

### 2)      Multi Layer Perceptron with Softmax (MLP)

The MLP is a widely used feed forward neural network that can be naturally applied to learn multiple regression tasks. For our purposes we further constrain the sum of the network outputs to 1 by using the softmax activation function [54]. This is an extension of the logistic function given by:

$$\hat{y}_t^i = \frac{\exp(q_t^i)}{\sum_{s=1}^T \exp(q_s^i)}$$

where $q_t^i$ is the activation value for the output node for task $t$ and input $i$.

### 3)      Linear and Kernel Ridge Regression (RR, KRR)

This is a baseline regression approach. In the linear form, RR is based in solving the optimization problem:

$$\min_{w_t} \|X^\mathrm{T} w_t - Y_t\|_2^2 + \lambda \|w_t\|_2^2$$

where $w_t$ represents the weight vector of the linear model $f_t(x) = \langle w_t, x \rangle$, and $x, w_t \in \mathbb{R}^d$, for task $t \in \{1, .., T\}$. For convenience we denote $\|\cdot\|_2$ as the $\ell_2 - \mathrm{norm}$ of a vector. One can extend this approach to nonlinear models by applying the kernel trick. In this case we have chosen the Gaussian kernel:

$$\mathcal{K}(x, t) = \exp\left(\frac{-1}{\sigma^2} \|x - t\|_2^2\right)$$

### 4)      Linear and Kernel Support Vector Regression (SVR, KSVR)

We also implement the widely used Support Vector Regression to predict the degree of belief of each of the laughter type based on the frequency of the ratings for each instance. In the linear form, SVR is based on the optimization

of the following problem:

$$\min_{w_t, \xi} \frac{1}{2} \|w_t\|^2 + C \sum_{i=1}^m \xi^i$$

$$s.t \begin{cases} |y_t^i - \langle w_t, x^i \rangle| & \leq \varepsilon + \xi^i \\ \xi^i & \geq 0 \end{cases}$$

In that, $\varepsilon \geq 0$ is the deviation allowed from the ground truth labels $y_t^i$. This constraint is weakened in some points by adding an extra margin $\xi_i$. The number of deviations larger than $\varepsilon$ is accounted by the second hyper-parameter $C \geq 0$. Similar to KRR, a nonlinear variant KSVR is also used in the comparison, employing also the Gaussian kernel.

### 6.1.2 Models with Feature Importance Estimation

In addition to the six aforementioned models we also include a further three methods, each of which has a model specific process to quantify the relative importance within the given feature set. These particular methods were selected as they are based on different paradigms and their approaches for relative feature importance evaluation are based on different principles and criteria. In doing this, we seek to gain further insight into which features contribute to the discriminatory power of each of these three models.

### 1)      Random Forest (RF) with permutation based feature importance testing.

We test the Random Forest algorithm [55] which generates an ensemble of decision trees, using the mean of the ensemble as the final outcome. Each of these trees has access to a set of $\delta$ attributes, randomly chosen when each tree is created. In the experiments conducted here, we vary the number of trees from 100 to 300, and the number of attributes considered for each tree is $\delta = \lfloor \sqrt{d} \rfloor = 7$, as suggested in [56]. We also test $2\delta$ attributes as a comparison to investigate the effect of allowing more attributes to be selected.

The relative importance of each feature can be estimated by calculating the change in the error from the random permutation of each feature within a testing set. Larger increases in the error value for a particular permutated feature demonstrate its greater importance within the whole ensemble. Features which give rise to little to no difference in error are deemed less important for the model.

### 2)      Least Absolute Shrinkage and Selection Operator (LASSO)

This method is a linear regression model which encourages sparsity within the weight vectors for regularization. This is done by the addition of an $\ell_1$-norm regularizer term leading to the following optimization problem:

$$\min_{w_t} \|X^\mathrm{T} w_t - Y_t\|_2^2 + \lambda \|w_t\|_1$$

where $\lambda \geq 0$ is a hyper-parameter that controls the degree of sparsity in the solution.

By way of sweeping a range of $\lambda$ values the relative importance of the features can be assessed. At each value of $\lambda$ the weight values in the solution associated with a particular feature can be inspected; non zero values show features that are in use and therefore important to the model. Features which return non-zero weights at higher levels of sparisty can be deemed as more important to the model than features that return non zero weights at low levels of sparsity [57].

### 3) Multi Layer Perceptron and Automatic Relevance Determination (MLP-ARD)

This model is a Bayesian extension of the standard MLP which considers probability distributions over the network weight space instead of single values. Prior initialization of these functions are set to zero mean Gaussian distributions. Posterior distributions are calculated according to Bayes' theorem given some training data [58]. As part of the Bayesian criterion it is necessary to optimize the model evidence. To this end a set of hyper-parameters $\alpha_j, j = 1, \ldots, d$ are used, each associated with the $j^{\text{th}}$ input node which in turn is associated with a feature. This provides a way to soft prune less important features during training; a process known as Automatic Relevance Determination (ARD) [59].

Inspecting the final $\alpha_j$ values after training can reveal the relative impact of each feature. Each $\alpha_j$ value reflects the variance within the weight values fanning out from each input node $j$. Smaller values can be deemed as more relevant than higher values where the associated weights have been restricted to values closer to zero and therefore less relevant in the network.

### 6.1.3 Evaluation Metrics

In order to robustly evaluate the multiple outcomes of the models against the distribution of the observers' categorisations, as suggested in [47], we apply four well established multi-score metrics over $M$ instances

1. Mean Square Error: this is the standard loss function which is computed as:

$$MSE := \frac{1}{MT} \sum_{i=1}^{M} \sum_{t=1}^{T} \left( y_t^i - \hat{y}_t^i \right)^2$$

2. Cosine Similarity: finds the cosine of the angle between two vectors resulting in a maximum of 1 when the vectors are fully aligned.

$$CS := \frac{1}{M} \sum_{i=1}^{M} \frac{\langle y^i, \hat{y}^i \rangle}{\|y^i\|_2 \|\hat{y}^i\|_2}$$

3. Top Match Rate: evaluates the number of times the predicted top ranked label is the same as the top ranked label for the ground truth.

$$TMR := \frac{1}{M} \sum_{i=1}^{M} 1_{\left\{ \underset{1 \leq t \leq T}{\operatorname{argmax}} \, y_t^i = \underset{1 \leq t \leq T}{\operatorname{argmax}} \, \hat{y}_t^i \right\}}$$

where $1_A$ is a function on condition A where:

$$1_A = \begin{cases} 1, & A \ is \ true \\ 0, & A \ is \ false \end{cases}$$

4. Ranking Loss: this metric calculates the average fraction of label pairs that are reversely ordered for an instance. By ordering the label outcomes as:

$$\left( y_{l_1}^i \geq y_{l_2}^i \geq \cdots \geq y_{l_T}^i \right)$$

The ranking loss predicted outputs can be calculated by:

$$RL := \frac{1}{M} \sum_{i=1}^{M} \frac{\sum_{j=1}^{T} \sum_{k=j+1}^{T} 1_{\left\{ \hat{y}_{l_j}^i < \hat{y}_{l_k}^i \right\}}}{T \times (T-1)/2}$$

where $1_A$ is the same function on condition A as for TMR.

## 6.2 Recognition results

All of the models described in section 6.1 were trained on three sets of data named: 'All-Set' for the full 126 instance dataset, 'Sit-Set' for the 57 instances of sitting stimuli and 'Stand-Set' for the 69 instances of standing stimuli.

For the All-Set, we evaluate all of the models using a leave one laugher out (*LOLO*) validation approach. This ensures that instances from the same individual laughers are not concurrently present in training, validation and test sets. We split the subjects into three groups: $n$ training laughers, 1 validation laugher to tune model parameters and 1 testing laugher to assess performance. For each model this procedure is repeated 72 times (9 test laughers × 8 validation laughers, accounting for all combinations) and the average results are reported. Parameter values were tuned over a set range for each of the models, the appropriate ranges were determined in initial experiments. The parameters adjusted are as follows: for *k*- NN: $k = \{5, 6, .., 9\}$; RR: $\lambda = \{10, 100,.., 10^5\}$; SVR: $C = \{10^{-5}, 10^{-4},..,100\}$; KSVR: $C = \{10^{-4}, 10^{-3},..,1\}$, $\sigma = \{10^{-7}, 10^{-6},..,10^{-3}\}$; KRR: $\lambda = \{10^{-6}, 10^{-5},..,10\}$, $\sigma = \{10^2, 10^3, 10^4\}$; LASSO: $\lambda = \{10, 100,.., 10^5\}$; RF: $\delta = \{7, 14\}$, $n_{trees} = \{100, 200, 300\}$ (total number of trees in the ensemble); MLP and MLP-ARD: $n_{hidden} = \{20, 40\}$ (the number of hidden layer nodes).

For the Sit-Set and Stand-Set a grouped LOLO validation was performed due to some laughers having very few instances where they are sitting or standing. Laughers with less than 3 instances were grouped into one fold. This still maintains that no instances from test laughers are present in the training set while providing the evaluation and test sets with a more balanced number of instances. For the Sit-Set this led to one grouping containing two laughers and for the Stand-Set this led to one grouping containing three laughers. Overall, for both Sit-Set and Stand-Set, this leads to six fold cross validation with (6 × 5) 30 runs for all combinations. The same parameter tunings were done here as for All-Set.

TABLES 3-5 – Comparisons of recognition performances. Each column provides the mean (standard devation) of one of the four evaluation metrics. ↑ indicates higher values correspond to better performance and ↓ indicates the opposite. The first nine rows correspond to the automatic recognition models, the last row (IR) indicates the mean level of agreement between observer groups

TABLE 3
PREDICTION MODEL COMPARISONS FOR ALL DATA.

|          | MSE ↓          | CS ↑        | TMR ↑       | RL↓         |
|----------|----------------|-------------|-------------|-------------|
| k-NN     | 0.015 (0.005)  | 0.88 (0.04) | 0.52 (0.21) | 0.30 (0.05) |
| RR       | 0.015 (0.005)  | 0.89 (0.03) | 0.44 (0.20) | 0.29 (0.07) |
| KRR      | 0.017 (0.008)  | 0.87 (0.05) | 0.45 (0.21) | 0.31 (0.07) |
| SVR      | 0.014 (0.005)  | 0.89 (0.03) | 0.50 (0.18) | 0.28 (0.08) |
| KSVR     | 0.015 (0.006)  | 0.89 (0.04) | 0.49 (0.15) | 0.30 (0.05) |
| LASSO    | 0.015 (0.005)  | 0.89 (0.03) | 0.43 (0.22) | 0.30 (0.08) |
| MLP      | 0.017 (0.007)  | 0.87 (0.05) | 0.50 (0.18) | 0.32 (0.05) |
| ARD      | 0.016 (0.006)  | 0.88 (0.04) | 0.50 (0.17) | 0.31 (0.05) |
| RF       | **0.010 (0.003)** | **0.92 (0.03)** | **0.59 (0.18)** | **0.26 (0.06)** |
| *IR*     | 0.022 (0.003)  | 0.95 (0.01) | 0.85 (0.03) | 0.10 (0.01) |

TABLE 4
PREDICTION MODEL COMPARISONS FOR SIT DATA

|          | MSE ↓          | CS ↑        | TMR ↑       | RL↓         |
|----------|----------------|-------------|-------------|-------------|
| k-NN     | 0.017 (0.007)  | 0.88 (0.05) | 0.61 (0.22) | 0.29 (0.08) |
| RR       | 0.017 (0.005)  | 0.88 (0.04) | 0.48 (0.23) | 0.30 (0.09) |
| KRR      | 0.017 (0.007)  | 0.88 (0.04) | 0.60 (0.19) | 0.30 (0.10) |
| SVR      | 0.017 (0.006)  | 0.88 (0.04) | 0.54 (0.19) | 0.29 (0.09) |
| KSVR     | 0.017 (0.007)  | 0.88 (0.05) | 0.62 (0.17) | 0.28 (0.09) |
| LASSO    | 0.017 (0.005)  | 0.88 (0.02) | 0.46 (0.17) | 0.28 (0.07) |
| MLP      | 0.019 (0.007)  | 0.87 (0.04) | 0.53 (0.15) | 0.31 (0.08) |
| ARD      | 0.017 (0.007)  | 0.87 (0.05) | 0.52 (0.17) | 0.29 (0.08) |
| RF       | **0.011 (0.004)** | **0.92 (0.03)** | **0.66 (0.14)** | **0.24 (0.08)** |
| *IR*     | 0.019 (0.004)  | 0.95 (0.01) | 0.84 (0.04) | 0.11 (0.01) |

TABLE 5
PREDICTION MODEL COMPARISONS FOR STAND DATA

|          | MSE ↓          | CS ↑        | TMR ↑       | RL↓         |
|----------|----------------|-------------|-------------|-------------|
| k-NN     | 0.014 (0.007)  | 0.88 (0.49) | 0.51 (0.17) | 0.29 (0.05) |
| RR       | 0.014 (0.005)  | 0.89 (0.03) | 0.42 (0.13) | 0.28 (0.05) |
| KRR      | 0.017 (0.007)  | 0.87 (0.04) | 0.41 (0.19) | 0.30 (0.05) |
| SVR      | 0.016 (0.006)  | 0.87 (0.04) | 0.37 (0.15) | 0.29 (0.04) |
| KSVR     | 0.015 (0.006)  | 0.88 (0.04) | 0.44 (0.20) | 0.28 (0.06) |
| LASSO    | 0.017 (0.007)  | 0.87 (0.05) | 0.39 (0.20) | 0.30 (0.06) |
| MLP      | 0.017 (0.008)  | 0.86 (0.05) | 0.49 (0.15) | 0.31 (0.06) |
| ARD      | 0.015 (0.006)  | 0.88 (0.04) | 0.48 (0.17) | 0.30 (0.05) |
| RF       | **0.012 (0.005)** | **0.91 (0.03)** | **0.57 (0.14)** | **0.26 (0.05)** |
| *IR*     | 0.024 (0.004)  | 0.94 (0.01) | 0.85 (0.04) | 0.10 (0.01) |

### 6.2.1 Predictive performances

Tables 3-5 show comparisons between the performances of all of the models using the four multi-score metrics for All-Set, Sit-Set and Stand-Set respectively. The results show mean (and standard deviation) of each evaluation measure after 72 runs for All-Set and 30 runs for Sit-Set and Stand-Set.

TABLES 6-8 – F1-score and recall in parentheses for each model based on the most frequent observer labels for the three categories with a significant number of instances.

TABLE 6
CLASSIFICATION PERFORMANCE FOR ALL DATA.

|          | Hilarious    | Social       | Not a Laugh  |
|----------|--------------|--------------|--------------|
| k-NN     | 0.43 (0.38)  | 0.46 (0.47)  | 0.57 (0.67)  |
| RR       | 0.57 (0.56)  | 0.30 (0.25)  | 0.58 (0.76)  |
| KRR      | 0.54 (0.52)  | 0.37 (0.34)  | 0.57 (0.70)  |
| SVR      | **0.63** (0.62) | 0.32 (0.25) | 0.61 (**0.84**) |
| KSVR     | 0.57 (0.55)  | 0.40 (0.35)  | 0.59 (0.76)  |
| LASSO    | 0.53 (0.56)  | 0.30 (0.25)  | 0.58 (0.75)  |
| MLP      | 0.49 (0.45)  | 0.50 (0.51)  | 0.56 (0.62)  |
| MLP-ARD  | 0.49 (0.49)  | 0.49 (0.49)  | 0.60 (0.66)  |
| RF       | 0.60 (**0.64**) | **0.58 (0.59)** | **0.76** (0.81) |

TABLE 7
CLASSFICATION PERFORMANCE FOR SIT DATA

|          | Hilarious    | Social       | Not a Laugh  |
|----------|--------------|--------------|--------------|
| k-NN     | 0.71 (0.66)  | **0.46 (0.41)** | 0.61 (0.76) |
| RR       | 0.77 (0.77)  | 0.13 (0.94)  | 0.50 (0.66)  |
| KRR      | 0.77 (0.80)  | 0.27 (0.19)  | 0.58 (0.76)  |
| SVR      | **0.80 (0.80)** | 0.12 (0.08) | 0.54 (0.75) |
| KSVR     | 0.74 (0.78)  | 0.34 (0.25)  | 0.60 (0.77)  |
| LASSO    | 0.79 (0.82)  | 0.17 (0.14)  | 0.42 (0.51)  |
| MLP      | 0.65 (0.58)  | 0.39 (0.37)  | 0.51 (0.61)  |
| MLP-ARD  | 0.67 (0.63)  | 0.33 (0.29)  | 0.54 (0.67)  |
| RF       | 0.69 (**0.84**) | 0.15 (0.09) | **0.78 (0.94)** |

TABLE 8
CLASSFICATION PERFORMANCE FOR STAND DATA

|          | Hilarious    | Social       | Not a Laugh  |
|----------|--------------|--------------|--------------|
| k-NN     | 0.20 (0.11)  | 0.56 (0.64)  | 0.58 (0.65)  |
| RR       | **0.28 (0.17)** | 0.40 (0.39) | 0.55 (0.71) |
| KRR      | 0.27 (**0.17**) | 0.46 (0.50) | 0.49 (0.56) |
| SVR      | 0.13 (0.09)  | 0.34 (0.31)  | 0.56 (**0.74**) |
| KSVR     | 0.00 (0.00)  | 0.45 (0.48)  | 0.54 (0.66)  |
| LASSO    | 0.05 (0.03)  | 0.45 (0.48)  | 0.48 (0.58)  |
| MLP      | 0.21 (0.14)  | 0.51 (0.55)  | 0.53 (0.56)  |
| MLP-ARD  | 0.20 (0.14)  | 0.54 (0.59)  | 0.58 (0.64)  |
| RF       | 0.00 (0.00)  | **0.66 (0.84)** | **0.71** (0.69) |

In addition, we also seek to understand the recognition rates between human observer groups to set this as a benchmark. This calculation provides a quantitative context when assessing the rates given in Tables 3-5. Using a simplified version of the approach proposed in [43], the raters were split randomly into two groups of 16 observers and the collective predictions of each group were computed. The same four measures used for evaluating the systems were applied to measure the agreement between these two predictions. We repeated this process 10,000 times and computed the averages (and standard deviation). These results are reported in the last row of Tables 3-5 as *IR*.
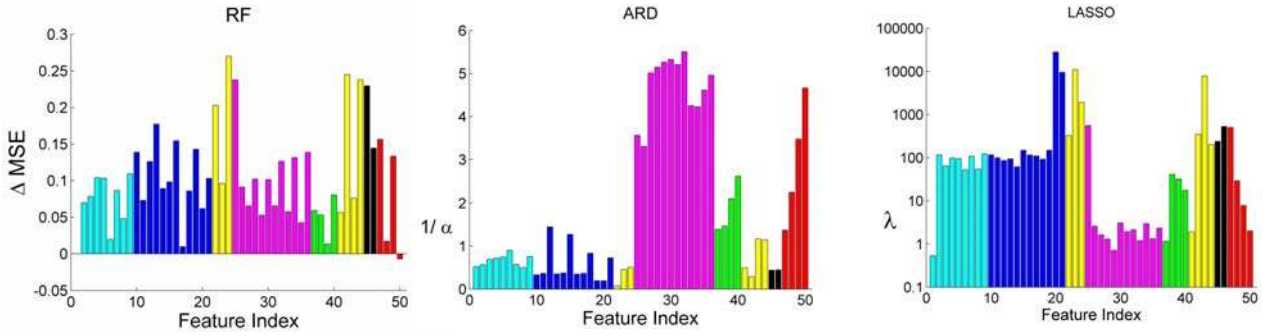
Fig. 7 – Feature importance histograms for RF (left), MLP-ARD (middle) and LASSO (right). The horizontal axes indicate the features index in the order indicated in Table 9. The vertical axes indicate the relative feature importance for the corresponding model. It can be seen that the Anterior-Posterior and Lateral component features (magenta) are more important to the nonlinear MLP-ARD model in contrast to the joint energies (yellow), joint angles(blue) and hand distances (cyan) which are more important to the linear LASSO model.

TABLE 9
FEATURE GROUPINGS BY TYPE WITH CORRESPONDING INDI-
CES. THE INDICES ARE GIVEN IN TABLE 2

| Column Index | Colour | Feature Type |
|---|---|---|
| 1-9 | Cyan | Hand distances |
| 10-21 | Blue | Spine Joint and Neck Angles |
| 22-24 & 41-44 | Yellow | Upper and Lower body joint energies respectively |
| 25-36 | Magenta | Anterior-Posterior and Lateral spine, head and neck components |
| 37-40 | Green | Azimuthal shoulder rotation local and global |
| 45-46 | Black | Ankle & Knee trajectory distances |
| 47-50 | Red | Localised S-I shoulder displacement information local shoulder trajectory smoothness |

In tables 3-5 it can be seen that the RF model outperforms all other methods across all four metrics for all three data sets. When considering MSE and CS, it can be seen that RF performs comparably with the IR benchmarks, but is lower for TMR and RL. In the All-Set, RR and SVR outperform their respective nonlinear variants KRR and KSVR, for the *MSE*, *CS* and *RL* metrics. A paired t-test for each metric over all validation runs of SVR against KSVR gave p = 0.0134 (MSE), p = 0.0072 (CS) and p = 0.0209 (RL). For KRR against RR we obtained: p = 0.0001 (MSE), p = 0.0009 (CS) and p = 0.004 (RL) Significant effects were not found for *TMR*. For the Sit-Set the nonlinear models are marginally better performing (except for *MSE* between KRR and RR) but with no statistical significance. For the Stand-Set RR significantly outperforms KRR p = 0.0005 (MSE), p = 0.0004 (CS) and p = 0.004 (RL). KSVR outperforms SVR only with regard to *MSE* (p = 0.014). In the direct comparison between MLP and MLP-ARD, there is a trend for the latter outperforming its non Bayesian variant in all cases except for the TMR in the 'Stand-Set'and 'Sit-Set'

In addition we calculate the F1-score and recall of the classifications from each of the models by assuming the most frequent observer label as the ground truth and the highest model output as the prediction (Tables 6-8). This can be viewed as treating the data as a classification problem. Within the 126 instances there were only 6 instances where awkward was the most frequent label and 5 instances for fake, whereas the number of instances for hilarious, social, and non-laughter were 25 , 46, and 44  respectively. Moreover, for some of the laughers these classes do not occur if ground truth is considered in this way. Since we use LOLO based cross validation the classification performance cannot be measured, therefore we show F1 and recall scores for only the categories: hilarious, social and non-laughter. In this case the benchmark of 0.2 can be considered as the chance level.

The outcomes shown in tables 6-8 reiterate the efficacy of the RF modelthough *k*-NN delivers the best performance in classifying social for the Sit-Set. It must be noted that all models do not classify hilarious in the Stand-Set above chance level; however this is attributed to there only being 6 modally hilarious instances in this subset. It can be seen that the nonlinear kernerlised models (KRR and KSVR) tends to outperform their linear counterparts in predicting social, this is not the case for 'Hilarious. This is also reflected in the difference between the linear LASSO model and the nonlinear MLPs in that the nonlinear methods classify social better than LASSO, again this is the opposite case for hilarious.

TABLE 10
CONFUSION MATRIX WITH NUMBER OF INSTANCES FOR RF
CLASSIFICATIONS ON THE ALL-SET -  ROWS SHOW GROUND
TRUTH (GT) AND COLUMNS SHOW PREDICTED (P)

| | Hilarious -P | Social - P | Not a Laugh - P |
|---|---|---|---|
| Hilarious - GT | **129** | 65 | 9 |
| Social - GT | 64 | **220** | 54 |
| Not a Laugh - GT | 7 | 83 | **289** |

### 6.2.2 Inspection of relative feature importance

Using the methods described in 6.1.2 a measure of relative feature importance is calculated respective to each corresponding model. For the RF model each feature within an unseen sample set is permuted. The difference in error ($\Delta MSE$) produced by each permutation is calculated at

each training run. Similary, for MLP-ARD we extract the final $1/\alpha_j$ values associated with each feature after each training run. Finally, for the LASSO model we sweep the range $\lambda = \{10^{-1}, 10^{-0.98}, ..., 10^5\}$ and extract the first occurance of a non zero weight value for each feature.

Fig. 7 left and middle show the mean of $\Delta MSE$ and $1/\alpha_j$ for the All-Set after the 72 runs. Fig. 7 (right) shows the $\lambda$ values at which each feature has a non zero weight after training on all instances in the All-Set. The interpretation of Fig. 7 is further informed by Table 9 which contains a brief description of the feature group types and the corresponding histogram indices.

The relative feature importance measures from ARD and LASSO shows distinct differences among the feature type groups. ARD clearly favours the Anterior-Posterior and Lateral component information for the spine, neck and head segments (index: 25-36) along with the localized shoulder motion correlation (index: 49) and the power spectral density of the vertical (superior-inferior) shoulder motion in the 4-6Hz band (index: 50). In contrast, hand distance information (index: 2-9), spine and neck angles (index: 10-21), 'Energy' (index: 22-24 & 41-44) and lower limb distance information (index: 45 &46) are important to LASSO. In the importance histogram returned by RF this distinction between groups of features is not as succinct; this shows importance attributed to specific subset of features within each group. Fig. 7 (left) shows high use of shoulder smoothness and correlation (index: 47 & 49), spinal contribution in the A-P plane (index: 25) as well as elbow, neck, spine and knee energies (index: 22, 24, 42 & 44), spine angle information (index: 13) and ankle trajectory distance (index: 45).

# 7 DISCUSSION

## 7.1 Models of laughter perception

Our historical understanding of laughter has included the existence of laughter "types" i.e., different expressions of laughter reflecting and caused by the laugher's underlying emotional state. This has been shown by the use of a wide range of adjectives to describe laughter either in terms of emotion e.g., nervous laughter, in addition to straightforward morphology e.g., hearty laughter. Our results, in parallel with those from a separate set of observers by McKeown et al. [50], suggest that categorical models may emerge from a discretisation of primary perception which is based on continuous dimensions. This suggestion is, at one level, trivial, as the perception of laughter starts, like any other perception, with low level features whose magnitude are (at a macroscopic level) continuous rather than categorical. Categorization occurs later in the perceptual process. This is in keeping with the continuum of laughter suggested by Darwin [11]; but more so with the incomplete continuum arising from two evolutionary pathways suggested by van Hooff [12].

The nature of these categories cannot be conclusively inferred here, but results from analysis of body movements and parallel datasets [50] suggest that the primary dimension is one of intensity, ranging from non-laughter to hilarious laughter. Although social laughter occupies the middle of this primary dimension it is relatively distinct on a second dimension indicating, perhaps, that even a pared-down laughter-type taxonomy should consider social laughter as separate from hilarious.

These conclusions are preliminary due to the limited stimulus set, which was constrained by practical considerations of observer fatigue. Furthermore, the number of categories available as observer responses was constrained by the pragmatic requirements of the perceptual experiment. If a set of categories could be sufficiently optimized, for example by including a more wide ranging taxonomy of categories and sub-categories, it is possible that a dimensional model would offer no additional explanatory weight.

Further work is therefore required to reveal whether dimensional descriptions of laughter can sufficiently describe the perceptual and social complexities of laughter perception. Observer reporting of further continuous dimensions, in addition to intensity, will generate valuable information; however, the low agreement on awkward and fake laughter suggests that the inclusion of other modalities, such as facial expression and vocalisations, may be necessary to generate models that encompass the perception of emotionally and socially complex laughter.

## 7.2 Body movements during laughter

Our findings on body movements related to the perception of laughter build on the work by Griffin et al. [6] by expanding the set of features analysed and refining some feature definitions. The results described here are correspondingly similar to those previously reported. We find that movements across many different body areas influence the perception of laughter. Which movements are indicative to observers of laughter differ between sitting and standing postures, but certain features are common to both, specifically those which we might intuitively associate with laughter i.e., bending of the spine and movement of the shoulders.

Remarkably clear patterns emerge in the data showing that large movements produce the perception of social laughter while the largest movements produce the perception of hilarious laughter. This result is extremely supportive of our interpretation of the dimensional analysis of laughter perception, which indicates that the dimension that explains the greatest percentage of variance corresponds to a perception of laughter intensity ranging from low/no-intensity non-laughter to high-intensity hilarious laughter.

Conversely the relative absence of strong effects from the planned contrast comparing hilarious and social laughter suggests that these may be distinguished on the basis of a high-level, more complex combinatorial interpretation of movement features. Social laughs, by definition, require reference to another person, for example, as a backchannelling response to a comment. The observers' inference that the avatar's movements show actions that reflect such an interpersonal reference is unlikely to be based on the value of a single feature. This is supported by the finding that non-linear supervised learning models yield the best results for recognition of social laughs. Future work which

considers how features interact, including temporal dependencies of movement features, is necessary to draw further conclusions on this issue.

## 7.3 Automatic recognition of laughter

Direct comparisons between linear models and their nonlinear counterparts reveal two effects: first, social laughter is better classified by the nonlinear variants with hilarious being better predicted by linear models.

In comparison with the findings in Griffin et al. [6] Table 6 reveals a notable improvement in F1-score and recall for every model when classfiying social laughter. This is attributed to the expanded and refined feature set, as all other factors are unchanged. Upon further inspection and in line with the nonlinearity of the social category the largest improvements are in KRR, KSVR and MLP. This lends itself to the notion that the new features introduced in this study generate further nonlinearity in the feature space which improves saliency in describing social laughter.

This is supported by the importance histogram returned by the nonlinear MLP-ARD Fig. 7 (middle) which shows high relevance for two of the three feature types not used in Griffin et al. [6] These being highly localised information such as: the Anterior-Posterior and Lateral components contributed by the spine, head and neck and also the local Superior-Inferior shoulder movement information. In contrast the feature type groups shown to be important for the linear LASSO model Fig. 7 (right) returns high importance to the more coarse level feature types used in Griffin et al. [6]: hand distances, spine angles and joint energies.

Overall it can be seen that the RF model outperforms all other models and when considering MSE and CS it is comparable with human rating levels (IR). Its feature importance evaluation returns mixed values when feature types are considered. This suggests a combination of both coarse level and highly localized features can lead to more effective recognition of laughter categories.

## 8 CONCLUSION

Laughter remains a relatively uninvestigated behavior in HCI, despite being a key social signal in positive and naturalistic social interaction. This study represents a multifaceted attempt to elucidate how body movements of laughter expression are perceived and how laughter can be automatically recognized from this modality. Our stimuli are necessarily a subsample of the vast number of possible natural human laughs, leaving open the possibility, as in all such studies of natural behaviour, that the effects found are specific to these stimuli. While we find strong evidence for differences between hilarious, social and non-laughter, future studies using larger stimulus sets are therefore necessary, for example, to determine whether awkward and fake laughter are signalled by perceptually distinguishable body movements.

Our results suggest that, as in many other emotional behaviours [38], body movements are a rich source of information and can provide a compelling and specific percept of laughter, even in very basic animations; however,

other modalities are likely to be necessary to drive the perception of laughs with more emotionally complex content such as awkwardness.

We show promising results in automatic detection of laughter, achieving high performance across a range of models that use different types of features to predict observer categorisations. This is encouraging for automatic real-time recognition as it implies that a suitable model could be chosen according to the format and range of body movement data available.

Finally we have produced evidence from perceptual results, along with data from a second set of observers [50] and mirrored by results from automatic recognition, suggesting that categorical descriptions of laughter may be the final outcome of a perceptual process that can be more accurately described by a dimensional model. If the nature of these dimensions can be clarified it may inform us not only of how laughter is perceived, but of the function and potential applications of laughter itself.

## REFERENCES

[1] M. Schröder, E. Bevacqua, R. Cowie, F. Eyben, F., H. Gunes, D. Heylen, et al. "Building Autonomous Sensitive Artificial Listeners," *IEEE Trans. Affective Computing*, vol. 3, no. 2, pp. 165–183. 2012.

[2] A. Turing, "*Computing machinery and intelligence*", *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[3] C. Creed and R. Beale, "*User interactions with an affective nutritional coach*," Interacting with Computers, vol. 24, no. 5, pp. 339–350, 2012.

[4] M. Ochs, R. Niewiadomski, P. Brunet, and C. Pelachaud, "Smiling virtual agent in social context," *Cognitive Processing*, vol. 13, iss. 2, pp. 1–14, 2012.

[5] E. Holt, "The last laugh: Shared laughter and topic termination," *J. Pragmatics*, vol. 42, pp. 1513–1525, 2010.

[6] H.J. Griffin, et al., "Laughter Type Recognition from Whole Body Motion." in *Proc. 5th Annu. Conf. Affective Computing and Intelligent Interaction*, Geneva, Switzerland, 2013, pp. 349-355.

[7] R. R. Provine, *Laughter: a scientific investigation*. London, UK: Faber and Faber, 2000.

[8] M. Davila Ross, J. Owren, M., and E. Zimmermann, "Reconstructing the evolution of laughter in great apes and humans." *Curr. Biol.* vo'. 19, no. 13, pp. 1106–1111, 2009.

[9] J. Panksepp, "Psychology: Beyond a Joke: From Animal Laughter to Human Joy?" *Science*, vol. 308, no. 5718, pp. 62–63, 2005.

[10] Matsusaka, T., "When does play panting occur during social play in wild chimpanzees?" *Primates*, vol. 45, no. 4, pp. 221–229, 2004.

[11] C. Darwin, *The descent of man and selection in relation to sex*. London, UK: John Murray, 1871.

[12] J. A. van Hooff, (1972). "A comparative approach to the phylogeny of laughter and smiling," in *Non-verbal Communication*. R. A. Hinde Eds, Cambridge, UK: Cambridge U. Press.

[13] E. Shahar, D. Goldsher, J. Genizi, S. Ravid, and Z. Keidar, "Intractable gelastic seizures during infancy: ictal positron emission tomography (PET) demonstrating epileptiform activity within the hypothalamic hamartoma." *J. Child Neurology*, vol. 23 no. 2, pp. 235–239, 2008.

[14] G. Juckel, et al., "Is evaluation of humorous stimuli associated with frontal cortex morphology? A pilot study using facial micro-movement analysis and MRI." *Cortex*, vol. 47, no. 5, pp. 569–574, 2011.

[15] K. Marinkovic, S. Baldwin, M.G. Courtney, T. Witzel, A. M. Dale, and E. Halgren, "Right hemisphere has the last laugh: neural dynamics of joke appreciation. *Cognitive, Affective, & Behavioral Neuroscience*, vol. 11, no. 1, pp. 113–130, 2010.

[16] G. Jefferson, H. Sacks, and E. Schegloff, "Notes on laughter in the pursuit of intimacy. *Talk and social organisation. Intercommunication series*, vol. 1, pp. 152–205, 1987.

[17] E. Holt, "On the nature of "laughables": laughter as a response to over-done figurative phrases." *Pragmatics*, vol. 21, no. 3, pp. 393–410, 2011.

[18] P. J. Glenn, *Laughter in Interaction*, Cambridge UK: Cambridge U Press, 2003.

[19] D. P. Szameitat, et al., "Differentiation of emotions in laughter at the behavioral level." *Emotion*, vol. 9, no. 3, pp. 397, 2009.

[20] M. Gervais, and D. Wilson, "The evolution and functions of laughter and humor: A synthetic approach," *Q. Rev. Biol.*, vol. 80, no. 4, pp. 395–430, 2005.

[21] E. G. Krumhuber, and A. S. R. Manstead, "Can Duchenne smiles be feigned? New evidence on felt and false smiles," *Emotion*, vol. 9, no. 6, pp. 807–820, 2009.

[22] D. Cosker and J. Edge, "Laughing, crying, sneezing and yawning: Automatic voice driven animation of non-speech articulations," in *Proc. 22nd Annu. Conf. Computer Animated Social Agents*, Amsterdam, 2009.

[23] R. Niewiadomski, J. Urbain, C. Pelachaud, and T. Dutoit, "Finding out the audio and visual features that influence the perception of laughter intensity and differ in inhalation and exhalation phases," in *Proc. 4th Int'l Workshop Corpora Research Emotion, Sentiment and Social Signals, LREC*, Istanbul, 2012.

[24] C. Becker-Asano, T. Kanda, C. Ishi, and H. Ishiguro. "Studying laughter in combination with two humanoid robots," *AI & Soc.*, vol. 26, no. 3, pp. 291-300, 2011.

[25] J. Urbain, et al., "AVlaughtercycle. enabling a virtual agent to join in laughing with a conversational partner using a similarity-driven audio-visual laughter animation," *J. Multimodal User Interfaces*, vol. 4, pp. 47–58, 2010.

[26] M. Filippelli, et al., "Respiratory dynamics during laughter," J. Applied Physiol., vol. 90, pp. 1441–1446, 2001.

[27] Z. V. DiLorenzo, P. and B. Sanders, "Laughing out loud: control for modeling anatomically inspired laughter using audio," in *ACM Trans. Graphics*, vol. 27, p. 125, 2008.

[28] R. Niewiadomski and C. Pelachaud, "Towards multimodal expression of laughter," in *Intelligent Virtual Agents*. Springer, 2012, pp. 231–244.

[29] C.-H. Chou, C.-H. Li, B.-W. Chen, J.-F. Wang, and P.-C. Lin, "A realtime training-free laughter detection system based on novel syllable segmentation and correlation methods," in *Proc. IEEE 4th Int'l Conf. on Awareness Science and Technology*, pp. 294–297, 2012.

[30] K. Laskowski, "Contrasting emotion-bearing laughter types in multiparticipant vocal activity detection for meetings," in Proc. IEEE Int'l Conf.Acoustics, Speech and Signal Processing, pp. 4765–4768, 2009.

[31] M. Miranda, J. A. Alonzo, J. Campita, S. Lucila, and M. Suarez, "Discovering emotions in filipino laughter using audio features," in *Proc. IEEE 3rd Int'l Conf. Human-Centric Computing*, pp. 1–6, 2010.

[32] K. P. Truong, and D. A. Van Leeuwen, "Automatic discrimination between laughter and speech." *Speech Communication, vol. 49*, no. 2, pp. 144-158, 2007.

[33] M. T. Knox, and N. Mirghafori. "Automatic laughter detection using neural networks." in *Proc. Annu. Conf. Int'l Speech Communication Association INTERSPEECH*, pp. 2973-2976. 2007.

[34] L. S. Kennedy, and D. P. W. Ellis, "Laughter detection in meetings." in NIST ICASSP Meeting Recognition Workshop, Montreal, 2004, pp. 118-121.

[35] S. Petridis, and M Pantic. "Audiovisual discrimination between speech and laughter: Why and when visual information might help." IEEE Trans Multimedia, vol. 13, no. 2, pp. 216-234, 2011.

[36] S. Petridis, M. Leveque, M. Pantic. "Audiovisual Detection of Laughter in Human-Machine Interaction" in *Proc. 5th Annu. Conf. Affective Computing and Intelligent Interaction,* Geneva, 2013, pp. 129 - 134.

[37] M. T. Suarez, J. Cu, and M. Sta. "Building a Multimodal Laughter Database for Emotion Recognition," in *LREC*, pp. 2347-2350. 2012.

[38] S. Fukushima, Y. Hashimoto, T. Nozawa, and H. Kajimoto, "Laugh enhancer using laugh track synchronized with the user's laugh motion," in *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, New York, 2010, pp. 3613–3618.

[39] M. Mancini, G. Varni, D. Glowinski, and G. Volpe, "Computing and evaluating the body laughter index," in *Human Behavior Understanding*. Springer, 2012, pp. 90–98.

[40] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotionrecognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[41] Z. Zeng, M. Pantic, G. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 39–58, 2009.

[42] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: a survey," *IEEE Trans. Affective Computing*, vol. 4, pp. 15–33, 2013.

[43] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, "Automatic recognition of non-acted affective postures," IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 41, no. 4, pp. 1027–1038, 2011.

[44] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *Affective computing and intelligent interaction*. Springer, 2007, pp. 71–82.

[45] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," in *Affective Computing and Intelligent Interaction*. Springer, 2007, pp. 59–70.

[46] H. Gunes and M. Piccardi, "Automatic Temporal Segment Detection and Affect Recognition from Face and Body Display", *IEEE Trans. on Systems, Man, and Cybernetics - Part B, Special Issue on Human Computing*, Vol. 39, No. 1, pp. 64-84, Feb. 2009.

[47] H. Meng, A. Kleinsmith, and N. Bianchi-Berthouze, "Multi-score learning for affect recognition: the case of body postures," in *Affective Computing and Intelligent Interaction*. Springer, 2011, pp. 225–234.

[48] C. Galvan, D. Manangan, M. Sanchez, J. Wong, and J. Cu, "Audiovisual affect recognition in spontaneous filipino laughter," in *Proc. IEEE 3rd Int'l Conf. Knowledge and Systems Engineering*, Hanoi, 2011, pp. 266–271.

[49] G. McKeown, W. Curran, C. McLoughlin, H. J. Griffin, and N. Bianchi-Berthouze, "Laughter induction techniques suitable for generating motion capture data of laughter associated body movements," in *Proc. 10th IEEE Int'l Conf. Automatic Face and Gesture* Recognition, Shanghai, 2013, pp. 1-5.

[50] G. McKeown, et al., Human Perception of Laughter from Context-free Whole Body Motion Dynamic Stimuli. in *Proc. 5th Annu. Conf. Affective Computing and Intelligent Interaction*, Geneva, Switzerland, 2013, pp. 306-311.

[51] G. McKeown, R. Cowie, W. Curran, W. Ruch, and E. Douglas-Cowie, "Ilhaire laughter database," in *Proc. 4th Int'l Workshop on Corpora for Research on Emotion, Sentiment & Social Signals, LREC*, 2012, pp. 32–35.

[52] W. Ruch and P. Ekman, "The expressive pattern of laughter," in *Emotion, qualia, and consciousness*, 2001, pp. 426–443,

[53] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet. "A Robust and Sensitive Metric for Quantifying Movement Smoothness", *IEEE Trans. Biomedical Engineering,* vol. 59, no. 8, pp. 2126-2136. 2012.

[54] J. S. Bridle, "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition," in *Neurocomputing*. Springer, 1990, pp. 227–236.

[55] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5–32, 2001.

[56] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and qsar modeling," *J. Chemical Information and Computer Sciences*, vol. 43, no. 6, pp. 1947–1958, 2003.

[57] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc B*, vol. 58, no. 1, pp 267-288. 1996

[58] C. M. Bishop, *Neural networks for pattern recognition*. Oxford University Press, UK, 1995.

[59] D. J. C. MacKay, Bayesian methods for backpropagation networks, in *Models of Neural Networks* III, E. Domany, J. L. van Hemmen, K. Shulten Eds., Springer-Verlag, New York,:1994.