# EXPLORING THE BODY AND HEAD KINEMATICS OF LAUGHTER, FILLED PAUSES AND BREATHS

Spyros Kousidis, Julian Hough and David Schlangen

Dialogue Systems Group, Bielefeld University, Germany
spyros.kousidis@uni-bielefeld.de

## ABSTRACT

We present ongoing work in the DUEL project, which focuses on the study of disfluencies, exclamations, and laughter in dialogue. Here we focus on the multimodal aspects of disfluent vocalizations, namely laughter and laughed speech, filled pauses, and breathing noises. We exemplify these phenomena in the rich multimodal *Dream Apartment Corpus*, a natural dialogue corpus, which, in addition to comprehensive disfluency and laughter annotation, comprises tracking data for the body and head. We discuss possible directions for developing models that can perceive as well as generate such multimodal behaviour.

**Keywords:** Laughter, disfluencies, multimodal

## 1. INTRODUCTION

In interaction research on both humans and artificial agents, there is a growing interest in phenomena that were previously avoided or ignored, namely disfluencies and laughter. The motivation comes both from the need for more precise perception of user behaviour, as well as the goal of modeling natural interaction and designing agents capable of it [15].

There are a number of improvements that modeling disfluencies and laughter can afford to human-agent interaction. For example, filled pauses at the beginnining of turns may act as a signal that more time is needed for production [2]. [1] demonstrated how this function of filled pauses could be exploited in human-robot interaction. Similar interaction management functions of audible breathing at turn boundaries can also be exploited. [7] show how the next speaker in multiparty interactions can be predicted using breathing and gaze features.

Laughed speech is problematic for ASR [3] and thus detection of laughed speech may help deal with possibly erroneous ASR output. [13] presented a classification of laughter, filled pauses, speech and silence, using audio features and a bigram model. Detection of filled pauses and repair disfluencies, which aids parsing spontaneous speech, has recently focussed on word-by-word incrementality [6], in order to allow the online interpretation of user speech beyond idealised, fluent utterances.

The multimodal aspects of disfluency and laughter phenomena have also had recent interest: for example laughter detection from facial features has proven quite successful [14, 4]. However, it is interesting to explore the behaviour in other modalities during such episodes. [10] found that human raters were highly confident in distingushing laughter from non-laughter, when observing a 3D virtual puppet performing body animations derived from raw human motion capture data. [5] classified different types of laughter (labeled by naive annotators on 3D animated avatar clips) using only the motion capture data. In both cases, the avatars were very simple, so that only the movements of the body were visible, while any other features such as body shape, height and face had been removed. The animations were based on episodes of elicited laugher, rather than laughter that occurs in interactive settings.

Using a multimodal corpus of natural dyadic interactions, we employ analytical methods to address the challenges outlined above, with the general hypothesis that a combined approach to detecting disfluencies and laughter phenomena, using multimodal data, will fare better than the previous approaches detecting these phenomena individually. As part of this ongoing work, we present our initial exploration in the area of laughter, filler and meaningful breath detection using kinematic features of the head and body.

## 2. MATERIAL AND ANNOTATIONS

In this work, we use the Dream Apartment Corpus [8, 9], a corpus of dyadic interactions between German speakers that features audio, video, body, head, eye and gaze tracking in a “minimally invasive” setting (no worn sensors, except for head-worn microphones). Body tracking was performed with Microsoft Kinect for Windows (version 1) and head/eye tracking was performed with Seeingmachines FaceLab.[1] An example scene from the corpus is shown in Figure 1. The Kinect data comprises 20 joints per

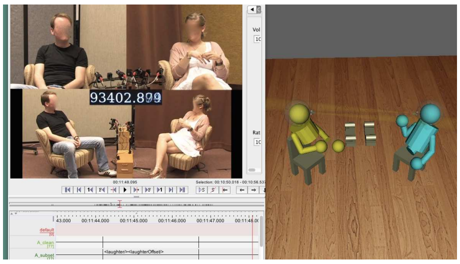

**Figure 1:** Scene from the Dream Apartment Corpus. The VR visualization shows the tracked body, (transparent) head, and gaze vectors

body at a sampling rate of 30 Hz, while the FaceLab data has twice that sampling rate (60 Hz).

The task of each dyad is to discuss the layout of an appartment in which they would co-habit, given an extraordinary amount of funds. This task is in place mostly to induce spontaneity in the dialogue by giving a topic and a limited amount of time (15 minutes) to come up with a proper appartment plan. The entire corpus consists of 9 dyadic interactions (balanced genders, varying degrees of acquaintance) approximately 20 minutes each (3 hours in total).

Based on the audio, we have annotated all instances of disfluent and laughter phenomena in the corpus as shown in Table 1. Audible breath and laughter offset, a special type of loud inhalation that often (but not always) follows laughter episodes, as well as laughed speech and laughter are annotated using XML-like tags. Filled pauses are enclosed in curly brackets and also start with an *F* character accompanied by the filled pause transcription.

| phenomena | label |
|---|---|
| Laughter | <Laughter/> |
| Laughed Speech | <Laughter> speech </Laughter> |
| Filled Pause | {F ähm } {F äh} |
| Breath | <Breath/> |
| Laughter Offset | <LaughterOffset/> |

**Table 1:** Subset of Annotation Labels from the DUEL annotation manual.

## 3. BODY MOVEMENT DURING LAUGHTER

Laughter is accompanied by movements of the body and head, which motivates efforts to use this modality in order to identify laughter events in dialogues. [11] have proposed the use of a *Body Laughter Index* which uses features such as the kinetic energy of the head and shoulders, as well as the periodicity of the movements of these joints. We are interested in improving on such metrics using the available motion capture data in the DAP corpus. Figure 2a shows the kinetic energy per joint during a laughter episode. The kinetic energy of each joint is computed separately from a vector of displacement per unit time derived from the raw Kinect tracking data.

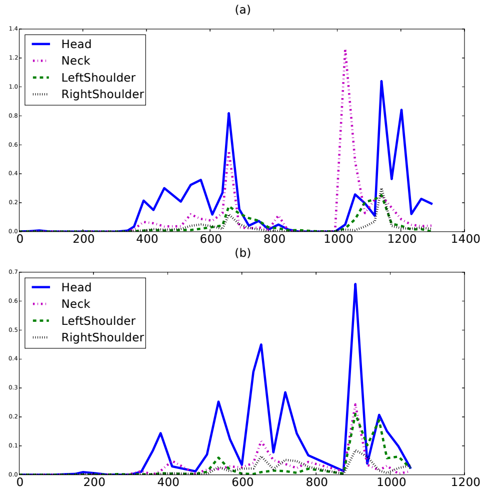

**Figure 2:** Energies of body joints during laughter episodes: (a) The head and neck perform much more pronounced movements than the shoulders (body is leaning forward and back). (b) periodic cycles of body movement during laughter.

A second area of interest is the periodicity of body movements during laughter episodes [12]. A second laughter episode is shown in Figure 2b. We observe the synchronous oscillation of the shoulders, neck and head, at a frequency of 8 Hz which is higher than the approximate 5 reported commonly in the literature. Of further interest is the study of periodicity features (frequency and phase) of laughter pulses and their comparison to periodicity features of the body and head.

## 4. HEAD GESTURES

Laughter and/or laughed speech is often accompanied by head gestures. The DAP corpus has already been annotated for communicative head gestures [8], by two annotators per dialogue. We consider only the overlapping annotated segments for which there is agreement on the type of gesture performed among the annotators, and compare these with carefully segmented occurences of laughter, filled pauses, and breaths. The results are shown in Table 2. We find that laughter episodes are accompanied by head gestures 39% of the time, which is quite frequent. The gesture types vary (the annotations in [8] distinguish 9 different types) and further work is required in order to determine whether some

| Gesture | Laughter | Fillers | Breaths |
|---|---|---|---|
| Nod | 20 | 7 | 7 |
| Turn | 7 | 6 | 6 |
| Tilt | 7 | 3 | 4 |
| Shake | 8 | 4 | 1 |
| Protrusion | 7 | 1 | 1 |
| Retraction | 3 | 1 | 0 |
| Shift | 3 | 2 | 0 |
| Bobble | 2 | 2 | 0 |
| Jerk | 0 | 0 | 1 |
| Co-occur % | 39.2 | 14.6 | 12.4 |

**Table 2:** Frequency of simultaneous head gesture and laughter/laughed speech, fillers and breaths per gesture type.

gestures accompany laughter more frequently than others. It is clear, however, that any automatic laughter detection that uses head motion features must take into account simultaneous communicative head gestures and their influence in head kinematics.

Similarly, although less often, we observe co-occurrence of breaths, laughter offsets and filled pauses with head gestures. The relative frequency of head turns as opposed to other gestures is higher in comparison to laughter episodes. We attribute this to instances of one interlocutor looking away from the other while audibly breathing or uttering a filler, in order to gain time to think. This analysis can be made more precise by considering events at the end of vocalisations, and also looking at the gaze vector, in addition to the head movement.

## 5. OUTLOOK

Our initial explorations in a rich multimodal corpus indicate several potential contributions and improvements on the currrent state of the art in the area of body-motion-based laughter and disfluency detection. As part of the DUEL project, data collections similar to the DAP corpus in three languages (German, French, Chinese) are planned, in order to explore cross-linguistic aspects of disfluencies and laughter in elicited spontaneous dialogues.

## 6. REFERENCES

[1] Bohus, D., Horvitz, E. 2014. Managing human-robot engagement with forecasts and... um... hesitations. *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM 2–9.

[2] Clark, H. H., Tree, J. E. F. 2002. Using uh and um in spontaneous speaking. *Cognition* 84(1), 73–111.

[3] Dumpala, S. H., Sridaran, K. V., Gangashetty, S. V., Yegnanarayana, B. 2014. Analysis of laughter and speech-laugh signals using excitation source information. *Acoustics, Speech and Signal Processing (ICASSP), 2014*. IEEE 975–979.

[4] Eyben, F., Petridis, S., Tzimiropoulos, G., Zafiriou, S., Pantic, M. 2011. Audiovisual Classification Of Vocal Outbursts In Human Conversation Using Long-Short-Term-Memory Networks. *Acoustics, Speech and Signal Processing (ICASSP)* Prague, Chech Republic. IEEE 5844 – 5847.

[5] Griffin, H. J., Aung, M. S., Romera-Paredes, B., McLoughlin, C., McKeown, G., Curran, W., Bianchi-Berthouze, N. 2013. Laughter type recognition from whole body motion. *Affective Computing and Intelligent Interaction (ACII), 2013*. IEEE 349–355.

[6] Hough, J., Purver, M. 2014. Strongly incremental repair detection. *Proceedings of EMNLP 2014*.

[7] Ishii, R., Otsuka, K., Kumano, S., Yamato, J. 2014. Analysis of respiration for prediction of who will be next speaker and when in multi-party meetings. *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM 18–25.

[8] Kousidis, S., Malisz, Z., Wagner, P., Schlangen, D. 2013. Exploring annotation of head gesture forms in spontaneous human interaction. *Proceedings of the Tilburg Gesture Meeting (TiGeR 2013)*.

[9] Kousidis, S., Pfeiffer, T., Schlangen, D. 2013. Mint. tools: Tools and adaptors supporting acquisition, annotation and analysis of multimodal corpora. *Proceedings of Interspeech 2013*.

[10] Mancini, M., Hofmann, J., Platt, T., Volpe, G., Varni, G., Glowinski, D., Ruch, W., Camurri, A. 2013. Towards automated full body detection of laughter driven by human expert annotation. *Affective Computing and Intelligent Interaction (ACII), 2013*. IEEE 757–762.

[11] Mancini, M., Varni, G., Glowinski, D., Volpe, G. 2012. Computing and evaluating the body laughter index. In: *Human Behavior Understanding*. Springer 90–98.

[12] Niewiadomski, R., Mancini, M., Ding, Y., Pelachaud, C., Volpe, G. 2014. Rhythmic body movements of laughter. *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM 299–306.

[13] Salamin, H., Polychroniou, A., Vinciarelli, A. 2013. Automatic detection of laughter and fillers in spontaneous mobile phone conversations. *IEEE conference on Systems, Man, and Cybernetics (SMC), 2013*. IEEE 4282–4287.

[14] Scherer, S., Schwenker, F., Campbell, N., Palm, G. 2009. Multimodal laughter detection in natural discourses. In: *Human Centered Robot Systems*. Springer 111–120.

[15] Urbain, J., Niewiadomski, R., Bevacqua, E., Dutoit, T., Moinet, A., Pelachaud, C., Picart, B., Tilmanne, J., Wagner, J. 2010. AVLaughterCycle. *Journal on Multimodal User Interfaces* 4(1), 47–58.

[1] http://www.eyetracking.com/Hardware/Eye-Tracker-List