The co-use of laughter and head gestures across speech styles

Bogdan Ludusan[1], *Marin Schröer*[1], *Martina Rossi*[2], *Petra Wagner*[1]

[1]Phonetics Workgroup, Faculty of Linguistics and Literary Studies & CITEC, Bielefeld University, Germany
[2]ISFAS, Kiel University, Germany

{bogdan.ludusan, marin.schroeer, petra.wagner}@uni-bielefeld.de,
mrossi@isfas.uni-kiel.de

Abstract

The multimodal expression of laughter has been studied in various fields, with previous work focused on the characterization of the visual component during laughter. We investigate here the interaction between laughter and the gestures preceding it, considering that communicative gestures usually occur before their associated verbal message. Employing two German datasets of dyadic interactions, one containing narratives and the other dialogues, we analysed the distribution of head gestures preceding laughter. The results showed high individual variation in the production of head gestures, but also an effect of speech style and of addressee characteristics. Narratives contained more gestures and, in these materials, the familiarity between interlocutors decreased gesture incidence. Examining also the link between laughter and the corresponding co-gestures, we observed a significant effect of gesture presence on laughter form.

Index Terms: laughter, gestures, head movements, human-human interaction, multimodality

1. Introduction

Laughter is a frequent phenomenon in spontaneous interactions [1], being used to express both joy/humour (mirthful laughter) as well as an array of social functions (social/non-mirthful laughter) in human communication [2]. It may express both positive behaviours (affiliation, pleasantness) as well as negative ones (e.g., disaffiliation, by laughing at someone). Taking into account its multitude of functions in conversation, it is, thus, only natural for it to interact with other communication components, including other modalities.

The role of the visual modality in connection with laughter has been mainly investigated from the prism of human-machine interaction, with the aim of better modelling the multimodal behaviour of virtual agents. In terms of the frequency of co-occurrence between laughter and gestures, previous work has shown that there is a sizeable overlap between the two [3], which is more frequent than between non-laughter and gestures [4]. Examining a corpus of natural dyadic conversations, Kousidis and colleagues [3] quantified the amount of laughter overlapping with head gestures, finding that 40% of the identified laughter events co-occurred with one of the nine types of annotated head gestures. At the same time, [4] has shown that most people produce more hand gestures during laughter than during non-laughter, in spontaneous interactions. [5] took one step further, by analysing facial expressions, as well as head and upper-body motions, and correlating them with laughter function and form. Further evidence for the connection between laughter and the visual modality comes from the literature on laughter detection, in which visual features, in the form of body movements,

head movements or facial gestures, may be used to improve the discrimination between laughter and speech [6, 7, 8].

Despite the interest in the multimodal expression of laughter, little is known of the visual component not overlapping with the laughter events, i.e. its context. Its preceding context should play a particularly important role, since the cognitive and communicative connection that exists between gesture and speech [9, 10, 11] is not only apparent during simultaneous events in the two modalities, but especially when the onset of gestures appears before the spoken referents [12, 13].

We analyse in this study the co-use of laughter and gestures, by focusing on a type of movement shown to relate to laughter, head gestures, found immediately preceding laughter. We look into the frequency of occurrence of these gestures considering that their production may vary with social and dialogic aspects of conversation [14]. In particular, we investigate whether speech style (dialogic) or addressee characteristics (social) have an effect on gesture frequency. Moreover, as in face-to-face conversation gestures co-operate and interact with the vocal message [15, 16, 17], we examine also any possible interplay between produced gesture type and following laughter form and function.

2. Materials and methods

2.1. Corpora

The materials employed in this study are part of the ALICO [18] and the DUEL [19] corpora. They both contain recordings of interactions between pairs of native German interlocutors, all of them being university students.

ALICO is a multimodal corpus designed for the investigation of listener feedback and consists of face-to-face recordings of pairs of speakers. One of the conversation partners, designated as the storyteller, was asked to tell two vacation stories, while their interlocutor was instructed to actively listen to the stories, having the option to make comments or ask questions about them. In one of the two recordings, the listener was distracted by an ancillary task while listening to the story (press a button when hearing a word starting with /s/). Those materials were not used here. The corpus was transcribed, annotated for feedback expressions and for various prosodic components, as well as automatically segmented. The corpus was also manually annotated for head gestures, considering rotation and translation movements across three axes [18]. The complete annotation inventory included the following head gesture types: nod, jerk, tilt, bobble, turn, shake, protrusion, retraction, slide, shift and waggle.

The DUEL corpus contains task-based interactions, in which the interlocutors had to discuss one of three scenarios, while also being video-recorded. We make use of data from

Table 1: *Statistics regarding the materials analysed in this study. For each of the two corpora, it illustrates the speech style type contained, the duration of the materials (in minutes), the number of speakers, as well as the number of analysed laughter events (of which laughs in parenthesis) and preceding head gestures.*

Corpus	Speech style	Duration [min.]	#Speakers	#Laughter (#laughs)	#Head gestures
ALICO	narrative	43	7	60 (18)	36
DUEL	dialogue	105	12*2	377 (270)	115

two scenarios. In the first scenario, Dream Apartment, the interlocutors were asked to discuss the design, furnishing and decoration of a shared apartment presuming that they had a very large amount of money at their disposal. The second scenario, Film Script, consisted in coming up with the script for a movie based on an embarrassing moment, which could be based also on personal experience. The scenarios were recorded in three languages: French, German and Mandarin Chinese. We focus here on the German part of the corpus. The recordings were transcribed and manually annotated for conversational phenomena, such as laughter events (both laughs and speech-laughs), disfluencies and exclamations. Similarly to the ALICO corpus, DUEL contains metadata about the recorded pairs, including their gender composition and the familiarity (if they knew each other before or not) between interlocutors.

Although both corpora contain interactions between two speakers, there is an important difference with respect to their speech style. The speech in the ALICO corpus can be seen more like a narrative, in which one speaker leads the talk, with only a minimum input from the listening interlocutor, while the DUEL recordings exhibit the characteristics of strongly interactive dialogues, where a constant exchange and active turn taking between the conversation partners is present.

2.2. Data processing

As the ALICO corpus contains head gesture, but no laughter annotations, while the opposite is the case for the DUEL corpus, we annotated the two corpora with the missing types of annotations. First, the recordings from the ALICO corpus having storyteller gesture annotations, were manually labelled for laughter events by an expert annotator, similarly to those in the DUEL corpus (laughs and speech-laughs).

Then, the recordings belonging to the Film Script scenario (containing the most laughter instances among the three scenarios of the DUEL corpus), as well as three pairs from the Dream Apartment which produced a higher amount of laughter, were manually annotated for head gestures. The gesture annotation was carried out according to the M3D [20] guidelines for gesture type, on intervals of up to 10 min from each recording, containing the highest laughter density (see [21] for more details). As the gesture inventory employed in the annotation of the ALICO corpus (further called *original* set) is larger than the one proposed by the M3D recommendations, we mapped the former into the latter, for a more straightforward comparison between the two datasets. The employed subset of the ALICO corpus had two additional gesture types compared to the DUEL subset (shake and retraction), which were mapped into existing head movements performed along the same axis (turn and protrusion, respectively). Shakes were defined in ALICO as repeated turns, while retractions represent the same motion as protrusions, only going backwards instead of forwards. We further call this merged inventory the *common* set.

Next, we analysed the laughter events in the two databases.

In case there were multiple laughter events (a sequence of laughter events by the same speaker or overlapping laughter) as a result of one laughable, we took into account in our analysis only the first one of that sequence / the one of the speaker laughing first, to control for potential carryover effects. We then determined if these events were preceded or not by a head gesture. A gesture was defined as preceding a laughter event if it was produced within one second before the beginning of the event. In case multiple gestures were found within this time interval, the one closest to the beginning of the laughter event was considered to be the corresponding one.

Finally, we annotated the function of all laughter events included in our analysis. We employed a two-class inventory, discriminating between mirthful and non-mirthful laughter. The annotation was performed by one expert annotator, with a second annotator labelling two recordings from the ALICO corpus and two from the DUEL corpus, in order to determine the inter-annotator agreement (Cohen's kappa, κ). Descriptive statistics regarding the materials used in this study are given in Table 1. From ALICO (7 recordings), the data produced by each storyteller was considered, while from DUEL (12 recordings) we included data from both interlocutors. We further refer to the two datasets (ALICO, DUEL) as the narratives and the dialogues data, respectively.

2.3. Analyses

We investigated the frequency of occurrence of different types of head gestures preceding laughter, including how they vary by speaker and with the speech style. Moreover, we compared their distribution to the general distribution of head gestures and determined whether their production and type is influenced by external factors, such as addressee characteristics (addressee gender or pair familiarity) or speech style (narrative vs. dialogue). We also looked into whether there exists an interaction between head gestures and their corresponding laughter events, testing if the former has an effect on either the form or the function of the following laughter event.

As our data is made up of nominal variables (and their counts), we employed logistic (in case of two levels) and multinomial (for variables with more than two levels) regression models to test the effect of various predictors on the variable of interest. A sum-to-zero contrast was used for the nominal predictors of the models. We compared the fit of these models using the Akaike Information Criterion (AIC). A better model would have a lower AIC value than a worse fitting one. For assessing the similarity of gesture distributions, Fisher's exact tests were employed.

All statistical analyses were performed using the functions provided by the R software [22]. The multinomial models were fitted using the *multinom* function from the *nnet* package [23], while the post-hoc tests following the Fisher's exact test were performed by means of the *row_wise_fisher_test* from the *rstatix* package [24], and applying Bonferroni correction.
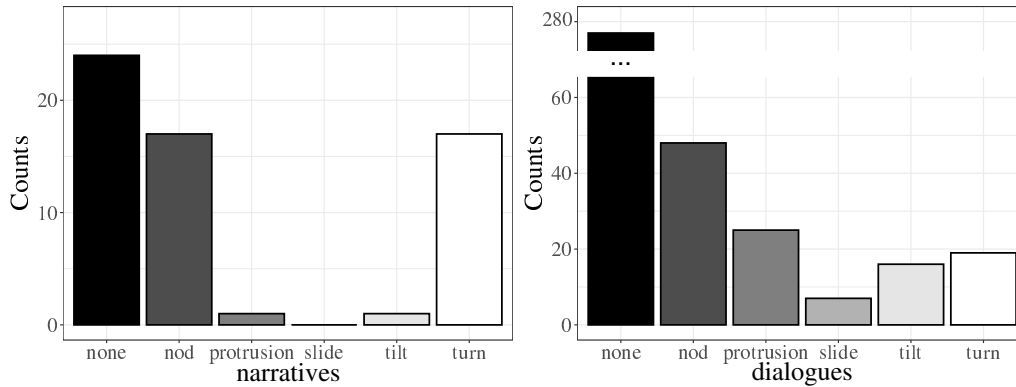
Figure 1: *Distribution of gestures preceding laughter in the narratives (left panel) and in the dialogues (right panel) data. We include, besides the five gesture types considered in the analysis (nod, protrusion, slide, tilt and turn) also the case when the laughter event was not preceded by a gesture (none). Please note the interrupted vertical axis in the right panel, used to illustrate the difference in frequency between the none case and the other cases.*

3. Results

We first compared, for each dataset separately, the distribution of head gestures (the original set, for the narratives data) occurring before laughter to that of all head gestures present in the dataset, using Fisher's exact tests. For the narratives data, it showed a significant difference between the two conditions ($p = 4.8e^{-6}$), with a post-hoc test showing that this effect is driven mainly by a higher proportion of head shakes occurring before laughter ($p = 8.7e^{-7}$). Also for the dialogue data, the difference in proportions between these two cases was significant ($p = 0.008$), being driven by a higher head protrusion production before laughter ($p = 0.006$).

For a more straightforward comparison of gesture production between the two corpora, the common set of gestures was used for all subsequent analyses. Figure 1 illustrates the distribution of head gestures preceding laughter (*none* denotes a lack of gesture). We can see that in the dialogue data a smaller amount of laughter events are preceded by gestures – 30.5% than in the narratives data – 60% (overall gesture rate being 34.6%). There are proportionally more nods than any other type of gesture in the dialogue data, with the other gestures occurring with a frequency of 10-50% of that of nods. Turns and nods appear equally frequent before laughter in the narratives data, making up almost the entire repertoire of employed gestures. We tested the difference in the distribution of gestures (excluding *none*) in the two datasets by means of a Fisher's exact test. It showed significant differences between the datasets, overall ($p = 1.6e^{-4}$), as well as in the production rate of protrusions ($p = 0.048$) and that of turns ($p = 0.003$). There is also considerable individual variation, within each dataset, with the per-speaker standard deviation being 25.4% and 21.8%, in the narratives and dialogue data, respectively. The rate of laughter preceded by head gesture varies between 0% (for two speakers) and 66.7% in the dialogue data, and between 37.5% and 100% (for two speakers) in the narratives data (see Figure 2).

We then checked, by means of a logistic regression model whether dialogic (speech style: narrative vs. dialogue) or social (addressee gender or pair familiarity) factors have an effect on the production of gestures before laughter. The latter was represented by the odds of gestures, i.e., the number of laughter events preceded vs. not preceded by head gestures, while the predictors were: speech style, addressee gender, pair fa-

miliarity, as well as the interactions between speech style and each of the other two factors. An ANOVA (type III) analysis of the model revealed a significant effect of speech style ($\chi^2 = 8.59, p = 0.003$) and an interaction between speech style and familiarity ($\chi^2 = 9.00, p = 0.003$). Keeping all other factors constant, the narrative increased the odds of using head gestures before laughter by 89.1% (95% confidence interval - CI [0.241, 1.904]), while in the case of the narrative materials, the familiarity between speakers decreased the use of head gestures preceding laughter by 39.3% (95% CI [0.157, 0.569]).

Having seen that the frequency/presence of gestures preceding laughter varies with the speech style and interlocutor (individual variation), we investigated whether these factors have an effect also on the types of head gestures produced. We employed multinomial logistic regression models with the gesture type as dependent variable and either speech style or interlocutor ID as predictors. The *none* was considered the baseline level of the models. Both models showed a significant effect of their respective predictor on the gesture type production (speech style: $\chi^2 = 44.17, p = 2.1e^{-8}$; interlocutor: $\chi^2 = 239.4, p = 4.7e^{-5}$). However, comparing the models by means of their AIC value revealed a better fit for the model having speech style as predictor (AIC = 960.8) than the one considering the interlocutor (AIC = 1055.5). The former model also showed that the odds of producing head turns vs. no gesture increase by 212.6% (95% CI [1.112, 3.608]) in the narratives,
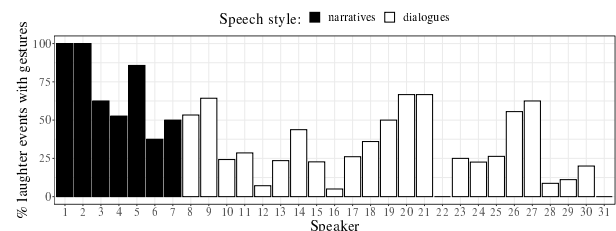

Figure 2: *Percentage of laughter events preceded by head gestures, out of the total number of laughter events, for each of the 31 speakers considered here. Speakers part of the narratives data are illustrated by black bars, while those in the dialogue data by white bars.*

compared to the dialogue data ($p = 8.7e^{-9}$), while the odds of head nods vs. no head gesture increase by 96.6% (95% CI [0.390, 1.781]) in the same data ($p = 1.3e^{-4}$).

Finally, we determined if the presence and the type of head gestures may have an effect on the form or the function of laughter. Looking at the form of the produced laughter event, the logistic regression models fitted on the overall data, with laughter type (laugh/speech-laugh) as dependent variable and preceding gesture presence (no/yes) as predictor, showed a significant effect of the latter (ANOVA type II: $\chi^2 = 12.09, p = 5.1e^{-4}$). The odds of producing speech-laughs, compared to laughs, increased by 44.0% (95% CI [0.173, 0.770]) when a gesture preceded the laughter event. We then built a logistic model on the subset of the data that was preceded by gestures, in order to determine whether there is an effect of gesture type on laughter form. The subsequent ANOVA (type II) analysis showed no overall gesture type effect ($\chi^2 = 6.17, p = 0.187$), although a significant effect for head turns was observed in the model ($p = 0.039$). The use of a head turns before a laughter event increased the odds of it being a speech-laugh by 103.1% (95% CI [0.045, 3.040]).

For investigating any possible role of gesture type on the function of following laughter (mirthful/non-mirthful), we limited the analysis to the dialogue data, since it was the part for which a higher inter-rater agreement was observed. While the agreement on the overall data (two recordings from each dataset) was only fair ($\kappa = 0.36$), it reached a κ value of 0.59 on the dialogue recordings, representing a moderate to substantial agreement between annotators. A similar logistic model to the one employed for determining the effect of presence of co-gestures on the laughter type was used also for laughter function. An ANOVA analysis of the model revealed no effect of co-gesture presence ($\chi^2 = 3.59, p = 0.058$). A logistic model, built on the subset of the data that was preceded by gestures, showed that also gesture type had no effect on laughter function (ANOVA of the model: $\chi^2 = 2.81, p = 0.590$).

4. Discussion and conclusions

Our findings point towards a specific role for the head gestures preceding laughter, seeing how their distribution differs from the overall gesture distribution. This role seems to vary with the speech style, with significantly more shakes occurring in narratives and more protrusions in dialogues. Moreover, by employing two different speech types we were able to determine a significant effect of this factor on the production of gestures, in line with the results found in the gesture literature [14]. These findings may have implications also for the use of gestures by virtual agents, with a higher amount of gestures being expected before laughter in case of a narrative than during a dialogue.

The frequency of co-use of laughter and head gestures across the two datasets (34.5%) is similar to the occurrence rate of head gestures overlapping with laughter reported in [3] – 40%. However, significant differences between co-gestures occurrence rate were observed between the two speaking styles, the narrative materials seeing a higher overall use of gestures, with more turns and less protrusions than in dialogues. Previous work on gestures reported more gestures in dialogues (more interactive) than in monologues (less interactive) [14], which may appear contradictory to our findings. Yet, neither of the employed styles here is similar to monologues, since the narratives data is actually a dialogue in which the listener intervenes with feedback and clarification questions.

How can the higher use of co-gestures in the narratives be explained? One the one hand, this might be due to the nature of the particular interaction: for narratives, the speaker is mainly responsible for keeping the engagement of the listener and, thus, may employ a wider range of communicative elements to achieve their goal, including gestures. For dialogues, this task is split between the interlocutors and, thus, might require less effort. On the other hand, the differences seen between the two styles may be due to the different functions that laughter plays in those conversations: narratives contain more instances of social laughter, while the majority of laughter events in the dialogue involve mirthful laughter, a consequence of the scenarios used for the recordings. Unfortunately, due to the low inter-rater agreement in our laughter function annotation, we could not assess the second possible explanation. Further analyses are needed to disentangle these two possible rationales.

Comparing our results on the types of gestures preceding laughter with those of [5], for gestures overlapping with laughter, we notice some differences, for instance a lower proportion of nods (here 43%, in their data between 40-65%, depending on the function of laughter) and a higher amount of laughter events without gestures (here 65%, compared to 5-20% in [5]). These differences might be due, not only to the different analysed gesture positions with respect to laughter, but also because some of the movements reported in [5] may be involuntary movements, caused by laughing. Our results also differ from those of Griffin et al. [25], who observed that the amount of body movement during laughing correlated with the type of laughter (mirthful > social > non-laughter). Here, we saw no difference in preceding head gestures between mirthful and non-mirthful laughter. Taking into account the less than optimal inter-rater agreement on laughter function in our data, we would like to verify this result on a dataset either with annotations having a higher level of agreement, or by employing a larger number of annotators and considering the majority label class.

We also found that the presence of head gestures may discriminate between types of laughter (laughs vs. speech-laughs), despite the fact that head movement features were found to be uninformative for laughter vs. speech discrimination [6]. Finally, in addition to examining a previously unstudied locus of gestures (with respect to laughter), our study reveals further information about the multimodal expression of laughter, showing that laughter co-gestures are produced with important individual variation, the percentage of laughter events accompanied by gestures varying from 0 to 100% across speakers.

To summarize, we have seen that the distribution of head gestures preceding laughter differs from the overall distribution and also from the distribution of movements during laughter, when comparing our findings with those in the literature. Despite important individual variation in the production of laughter co-gestures, we saw that the gesture frequency is modulated by dialogic (speech style) and social (addressee familiarity) factors, in line with previous work on gestures. Moreover, we found that the presence of co-gestures, as well as their type, plays a role in the form of the following laughter event. These results provide new insights into the interplay between gestures and the voice (here, paralinguistic) component, allowing also the modelling of more realistic behaviours in human-machine interaction.

5. Acknowledgements

6. References

[1] J. Trouvain and K. Truong, "Comparing non-verbal vocalisations in conversational speech corpora," in *Proc. of the LREC Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, 2012, pp. 36–39.

[2] P. Glenn, *Towards a social interactional approach to laughter*, ser. Studies in Interactional Sociolinguistics. Cambridge University Press, 2003, pp. 7–34.

[3] S. Kousidis, J. Hough, and D. Schlangen, "Exploring the body and head kinematics of laughter, filled pauses and breaths," in *Proc. of the 4th Interdisciplinary Workshop on Laughter and Other Nonverbal Vocalisations in Speech*, 2015, pp. 23–25.

[4] H. Griffin, G. Varni, G. Volpe, G. T. Lourido, M. Mancini, and N. Bianchi-Berthouze, "Gesture mimicry in expression of laughter," in *Proc. of ACII*, 2015, pp. 677–683.

[5] C. Ishi, H. Hatano, and H. Ishiguro, "Audiovisual analysis of relations between laughter types and laughter motions," in *Proc. of Speech Prosody*, 2016, pp. 806–810.

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

[7] R. Niewiadomski, M. Mancini, G. Varni, G. Volpe, and A. Camurri, "Automated laughter detection from full-body movements," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 113–123, 2015.

[8] B. Türker, Y. Yemez, M. Sezgin, and E. Erzin, "Audio-facial laughter detection in naturalistic dyadic conversations," *IEEE Transactions on Affective Computing*, vol. 8, no. 4, pp. 534–545, 2017.

[9] A. Kendon, "Gesticulation and speech: Two aspects of the process of utterance," in *The relationship of verbal and nonverbal communication*, M. Key, Ed. De Gruyter Mouton, 1980, pp. 207–228.

[10] E. A. Schegloff, "On some gesture's relation to talk," in *Structures of social action: Studies in conversation analysis*, J. M. Atkinson, Ed. Cambridge University Press, 1984, pp. 266–296.

[11] D. McNeill, *Gesture and Thought*. University of Chicago Press, 2005.

[12] P. Morrel-Samuels and R. M. Krauss, "Word familiarity predicts temporal asynchrony of hand gestures and speech." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 18, no. 3, pp. 615–622, 1992.

[13] E. McClave, "Linguistic functions of head movements in the context of speech," *Journal of Pragmatics*, vol. 32, no. 7, pp. 855–878, 2000.

[14] J. Holler and J. Bavelas, "Multi-modal communication of common ground: A review of social functions," in *Why gesture? How the hands function in speaking, thinking and communicating*, M. W. A. R. Breckinridge Church and S. D. Kelly, Eds. Benjamins, 2017, pp. 213–240.

[15] J. B. Bavelas, N. Chovil, D. A. Lawrie, and A. Wade, "Interactive gestures," *Discourse Processes*, vol. 15, no. 4, pp. 469–489, 1992.

[16] B. Roustan and M. Dohen, "Gesture and speech coordination: The influence of the relationship between manual gesture and speech," in *Proc. of INTERSPEECH*, 2010, pp. 498–501.

[17] E. De Stefani and D. De Marco, "Language, gesture, and emotional communication: An embodied view of social interaction," *Frontiers in Psychology*, vol. 10, p. 2063, 2019.

[18] Z. Malisz, M. Wlodarczak, H. Buschmeier, J. Skubisz, S. Kopp, and P. Wagner, "The ALICO corpus: Analysing the active listener," *Language Resources and Evaluation*, vol. 50, no. 2, pp. 411–442, 2016.

[19] J. Hough, Y. Tian, L. de Ruiter, S. Betz, S. Kousidis, D. Schlangen, and J. Ginzburg, "DUEL: A multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter," in *Proc. of LREC*, 2016, pp. 1784–1788.

[20] P. L. Rohrer, I. Vila-Giménez, J. Florit-Pons, G. Gurrado, N. Esteve Gibert, A. Ren, S. Shattuck-Hufnagel, and P. Prieto, "The multimodal multidimensional (M3D) labelling scheme for the annotation of audiovisual corpora," in *Proc. of GESPIN*, 2020.

[21] M. Rossi, M. Schröer, B. Ludusan, and M. Zellers, "A multimodal account of listener feedback in face-to-face interactions," in *Proc. of ICPhS*, 2023.

[22] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org/

[23] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer, 2002, iSBN 0-387-95457-0. [Online]. Available: http://www.stats.ox.ac.uk/pub/MASS4

[24] A. Kassambara, *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*, 2023, R package version 0.7.2. [Online]. Available: https://CRAN.R-project.org/package=rstatix

[25] H. J. Griffin, M. Aung, B. Romera-Paredes, C. McLoughlin, G. McKeown, W. Curran, and N. Bianchi-Berthouze, "Perception and automatic recognition of laughter from whole-body motion: Continuous and categorical perspectives," *IEEE Transactions on Affective Computing*, vol. 6, no. 2, pp. 165–178, 2015.