Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

An acoustic-prosodic analysis of laughter types

Bogdan Ludusan, Marin Schröer, Petra Wagner

Phonetics Workgroup, Faculty of Linguistics and Literary Studies & CITEC, Bielefeld University, Germany

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

Abstract

Laughter is a non-verbal phenomenon, widely used in human interaction, which has been shown to differ from speech along various acoustic-prosodic dimensions. Previous work has also revealed that the production of laughter is subject to a high degree of variation, with speakers normally having several types of laughter in their repertoire. Despite this, relatively little is known about how different types of laughter are marked prosodically and whether prosodic features may be used to discriminate between laughter types. We investigated here two types of laughter events produced in spontaneous interaction, with respect to five prosodic characteristics: duration, pitch, intensity, rhythm and voice quality. Our results showed that each of these characteristics, except for rhythm, differ between laughter types. We then employed prosodic features in a machine learning system trained to discriminate between three classes: speech and the two types of laughter. The proposed system obtained a similar speech/laughter classification performance to that of a system that considers only two classes, speech and laughter, while also having the advantage that a finer distinction, i.e., between laughter types, may be achieved.

Index Terms: laughter, prosodic features, acoustic analysis, laughter type, automatic detection

1. Introduction

Laughter, as one of the most often encountered non-verbal vocalizations in human interaction, has been extensively studied, including its acoustic realization. Despite the general belief that laughter is a stereotypical vocalization, including by earlier acoustic studies on laughter [1], it has been shown that the realization of most laughter instances in conversation is actually much more complex, differing from this stereotypical definition [2]. This view was corroborated by [3], which found that laughter instances corresponding more closely to (stereo)typical laughter (called song-like laughter, in the study), represented only 30% of total laughs produced, with the rest being snort-like, grunt-like or mixed laughs. Moreover, the same study showed that the laughter repertoire of the speakers is quite rich, with over 80% of the speakers using two or more types of laughter. These laughter characteristics have been found not only quantitatively, but also in qualitative studies, such as in [4].

In spite of the evidence showing that several types of laughter, with respect to phonetic form, are employed, most previous work treated laughter as one homogeneous class and did not make any distinction between different laughter types. The realization of various laughter types has been investigated only in a few studies [5, 6, 3, 7, 8] and not all of them looked at its acoustic marking. Moreover, those that did look at acoustic features of laughter, did so for a limited set of prosodic parameters, composed of duration and/or fundamental frequency (f0) measurements. [3] and [8] looked at song-like, grunt-like and snort-like laughter (with speech-laughs being an additional type in [8]) in elicited and conversational laughter, respectively. [6] examined four laughter types (comment, chuckle, rhythmic and squeal) produced by children, while [7] considered voiced, ingressive, chuckle and nasal in their analysis of conversational laughter. However, laughter is known to be marked, besides these aforementioned cues also by other prosodic characteristics, such as intensity [9], voice quality [10, 9] or rhythm [11, 12].

Another typology under which laughter has been studied, pertains to the functions that laughter plays. Vettin and Todt [13], analyzing laughter produced in spontaneous interactions, including many instances of social laughter, showed that these instances differed from laughter employed in previous studies (humorous laughter) along different dimensions: laugh composition, duration, and f0, suggesting that laughter realization varies also with the function of laughter. Similar conclusions, that laughter function types may be differently marked acoustically, have also been drawn in more recent studies looking at volitional/spontaneous [14, 15] or voluntary/involuntary laughter [16].

The view that laughter represents a uniform phenomenon is seen not only in speech sciences, but also in speech technology, especially in one of the laughter-related applications that witnessed considerable interest over the years – automatic laughter detection [17]. A wide variety of system implementations were proposed, differing in the employed decision level (frame [18] or segment [19]), the types of features considered (e.g., spectral and prosodic [20] or mostly prosodic [21]), or the type of learning (shallow [22] or deep-learning [23]). Despite the breadth of approaches, all these systems focused on laughter/speech discrimination. There have been few proposals that tried to discriminate speech from more than one laughter class [24, 25, 26, 27]. These systems classified either speech, laughs and speech-laughs [27], speech, voiced laughs and unvoiced laughs [24, 25] or all of the above [26]. Comparing these results with those obtained for only laughter/speech discrimination, [28, 26] showed that increasing the number of classes to discriminate between normally decreases the overall system performance.

The present study focuses on two aspects. First, we investigate how five prosodic characteristics: duration, fundamental frequency, speech intensity, voice quality and rhythm change with respect to phonetic laughter type. We considered for this analysis two types: typical and non-typical laughter. We will also explore, within each laughter type class, whether function (social or mirthful) has an effect on the acoustic marking of laughter. This analysis will help us better understand the acoustic-prosodic realization of laughter in conversation. Second, we explore how the findings of the previous analysis may inform speech technology, by testing if a laughter detection trained on these prosodic cues system may profit from considering more than one laughter class in its discrimination process.

 10.21437/SpeechProsody.2024-122

2. Methods and Materials

2.1. Data

We employed recordings from the German part of the DUEL corpus [29] for the analyses. It contains conversations between pairs of speakers involved in three different tasks. We considered here the tasks that elicited the most laughter, the Dream Apartment and the Film Script. For the first task, the interlocutors had to discuss the planning and the furnishing of a large apartment they were supposed to share, taking into account they had a large sum of money at their disposal for this. The second task involved coming up with the script for a movie containing an embarrassing scene, that could be based on own experience. The data included recordings from 19 pairs of speakers (9 Dream Apartment and 10 Film Script), for a total of 260 minutes. The corpus was orthographically transcribed and manually annotated for speaker turns and for several conversational phenomena, including laughter.

2.2. Analyses

We considered all laughs produced by the interlocutors during conversation. We then manually annotated laughter type, differentiating between typical and non-typical laughter. First, we looked for laughs that were at least two syllables long, having a clear rhythmical structure, being egressive, non-nasal, non-fricated and being voiced, with f0 modulation (variability between laughter syllables). If the entire laughter event satisfied all these conditions, it was labelled as typical and, if only a part of it satisfied all the conditions, it was annotated as part-typical. Otherwise, it was considered to be non-typical. For the purpose of this analysis, we considered both typical and part-typical laughs to belong to the typical laughter class. Thus, our definition of typical laughter is similar to that of song-like [3, 8], rhythmic [6] and voiced laughter [7], respectively, although being slightly broader.

Our annotations yielded a data set containing 285 typical and 727 non-typical laughter events, with 31 of the 38 speakers (a similar proportion to [3]) having produced both types of laughs. In order to also compare the acoustic realization of these laughter types with that of speech, we included speech intervals in our analysis, as well. For this purpose, we considered speaker turns at least 1.5 seconds long, not containing any laughter event or any annotated breathing sounds (but they could contain other conversational phenomena, such as silent or filled pauses). We limited the number of speech intervals to a maximum of 25 per speaker, resulting in a set of 929 items.

For these items, we extracted the following acoustic features: the duration (*dur*), the fundamental frequency (*f0*), the signal intensity (*en*), the cepstral peak prominence (*cpp*) and the modulation spectrum (*modSp*). The duration was determined based on the corpus annotation, the *f0*, *en* and *cpp* were extracted by means of VoiceSauce [30], and *modSp* was extracted using the AM_FM_ Spectra toolbox [31]. Cepstral peak prominence represents the amplitude of the cepstral peak relative to the regression line over the entire cepstrum [32] and is a measure of voice quality, with lower values corresponding to more breathy phonation. The *modSp* measure used here represents the mean value of the modulation rate/audio frequency channels of the modulation index spectrum [33] that maximize the discrimination between laughter and speech classes, as given by [34]. The default parameters of VoiceSauce were employed to extract the features, with a frame shift of 1 ms. The modulation index spectrum was computed with default parameters and an analysis window of 1.5 seconds. In case the analyzed item was shorter than 1.5 seconds, we centered the analysis window on the middle of the item (and, thus, the window contained also some speech material on both sides of it). For items longer than 1.5 seconds, we considered a number of analysis frames equal to: (duration div 1.5) + 1, and we distributed the analyses windows equally within the item.

After feature extraction, we computed the mean value over each item, for all features (except duration). As in some cases, there might be cross-talk occurring within the analyzed laughter unit, we also considered the maximum values for f0 and signal energy, as being more robust measures: f0 values of laughter are higher than those of speech, thus this measure should be less affected by cross-talk. Similarly, the maximum of the intensity should be less perturbed by speech from the other channel. All extracted values were then employed in a statistical analysis to determine whether any differences exist between typical and non-typical laughter, as well as between the different laughter types and speech (except for duration, since the duration of the speech intervals was arbitrarily chosen). We also performed an additional exploratory investigation looking at laughter function, considering that differences along several acoustic dimensions have been found between various laughter functions [14, 15, 16]. However, since laughter function annotation has very low levels of inter-rater agreement [35], we operationalized laughter function based on its duration, relying on evidence that social laughter is generally shorter than mirthful laughter [36, 35]. Thus, we compared how the results obtained with this simplifying assumption hold with respect to other studies that used different methods for ascertaining the functions of the analyzed laughs. As the distribution of the data did not allow for the use of linear regression models or of parametrical tests, we employed Mann-Whitney U tests (Wilcoxon rank sum tests), when comparing two classes, or Kruskal-Wallis tests, in the case of three classes. All analyses were conducted using the R software [37].

Finally, we examined whether the four prosodic features analyzed here (all, but duration), can help the automatic classification of laughter vs. speech. We built two Naive Bayes (NB) classifiers, one classifying two classes: speech and laughter, and the other three classes: speech, typical laughter and non-typical laughter. NB is a supervised learning algorithm that fits a number of Gaussian distributions equal to the number of classes given, considering each feature as being independent. We made use here of the implementation given by the scikit-learn library [38]. We compared the performance of the two systems in terms of precision (how many items, out of the total classified as laughter, were correct), recall (how many laughs, out of the total number of laughs, were found), F-score (the harmonic mean of precision and recall) and accuracy (the ratio of correctly classified items, speech or laughter, out of the total). In the case of the latter system, for evaluation purposes, we also folded the two laughter types into one class, to determine whether trying to discriminate two instead of only one laughter class would affect the laughter/speech discrimination. We used a 10-fold cross-validation paradigm, such that the data from each speaker would appear only in one fold at a time. We report the average values for precision, recall, F-score and accuracy over the 10 folds.

3. Results

The mean values of the five investigated prosodic cues and the maximum values for *f0* and *en*, for the two laughter type classes,

Table 1: *Summary of the values of the prosodic features examined in this study, for each laughter type (typical/non-typical). The mean values of the fundamental frequency (f0), root-mean-square energy (en), cepstral peak prominence (cpp) and modulation spectrum (modSp), as well as the maximum values for f0 and en, averaged over all the events corresponding to each class are illustrated. The values of the same features for the speech utterances considered in the analysis are also displayed, for comparison. The duration of the speech intervals is not listed, since it was arbitrarily chosen.*

Event	**Type**	**dur** [s]	**f0** [Hz]	**en** [Pa]	**cpp** [dB]	**modSp** [dB]	**max_f0** [Hz]	**max_en** [Pa]
Laughter	typical	1.231	263	1.648	14.72	0.701	402	18.80
	non-typical	0.919	226	0.523	14.31	0.710	331	4.94
Speech	-	-	181	1.119	17.33	0.192	286	8.46

are illustrated in the first two rows of Table 1. We also presented the values of the same cues for the speech class, for comparison purposes. We did not report a value for the duration of the speech intervals, as this would not have any meaning, since we chose arbitrarily long speech intervals (at least 1.5 seconds long). We observed higher values for all cues, but *modSp*, for typical laughter than for non-typical laughter. Also, laughter seems to have higher values than speech for all cues, except for *cpp*.

We compared the differences between typical and non-typical laughter, in terms of how they are marked, by means of Wilcoxon tests. They showed significant differences for *dur* ($p < 2.2e^{-16}$), *en* ($p < 2.2e^{-16}$), *f0* ($p = 4.6e^{-12}$) and *cpp* ($p = 7.3e^{-13}$). Also the differences between laughter types in terms of maximum f0 ($p = 5.0e^{-15}$) and maximum intensity ($p < 2.2e^{-16}$) were found to be significant. Only the value of *modSp* did not differ between the two laughter types ($p = 0.227$). Moreover, Kruskal-Wallis tests revealed that the type of vocalization (typical laughter / non-typical laughter / speech) had a significant effect on the values of each of the four acoustic cues (*f0*, *en*, *cpp*, *modSp*; $p < 2.2e^{-16}$ in all cases).

We then performed a more detailed investigation of laughs having different duration, within each of the two types of laughter. We considered two classes, shorter vs. longer, defined with respect to the median of all laughter instances included in our analysis (see Table 2). It showed that shorter laughs have higher mean intensity ($p = 0.04$) and lower *modSp* values ($p = 0.002$) than longer laughs, in the typical laughter case, as well as a lower mean f0 ($p = 4.1e^{-4}$) and a lower *modSp* value ($p = 2.0e^{-4}$), in the non-typical laughter condition. Correcting these results for multiple comparisons, by applying Bonferroni correction, revealed that the modulation spectrum (for both types of laughter) and f0 (for non-typical laughter) are significantly different between shorter and longer laughs.

Lastly, we examined whether the four acoustic-prosodic cues: *f0*, *en*, *cpp*, *modSp*, are also useful in automatic classification systems. The performance of the NB classifiers trained to discriminate between laughter and speech (first line) and between typical laughter, non-typical laughter and speech (second line) is presented in Table 3. The evaluation was performed based on two classes (laughter, speech), having merged the two laughter classes of the second system into one class (laughter), and we report the obtained precision, recall, F-score and accuracy. We can see a similar performance between the two classifiers in terms of laughter/speech discrimination. The performance of the 3-class classifier for distinguishing between all three classes, stood at 81.0% accuracy. Since this classifier is equally good at discriminating between laughter and speech as the 2-class system, the lower overall accuracy is due to a higher confusion between laughter classes (the accuracy for the two laughter types being 76.6%).

Table 2: *The values of the prosodic features examined in this study, for each laughter type (typical/non-typical) and each duration class (shorter/longer), as defined based on the median duration of all laughs.*

Type	**Duration**	**f0** [Hz]	**en** [Pa]	**cpp** [dB]	**modSp** [dB]
typical	shorter	255	1.955	14.80	0.617
	longer	266	1.507	14.70	0.739
non-typical	shorter	218	0.595	14.32	0.676
	longer	237	0.425	14.30	0.754

4. Discussion

We investigated here the acoustic-prosodic marking of two types of laughter produced in conversation, comparing it also to the acoustic characteristics of speech. We identified in our data the typical and non-typical laughs, showing that a proportion of 28.2% of the laughter events are of the former type. These values are similar to those reported in previous studies (30% in [3] and 34% in [8], at the laughter event level, and around 30% in [7], at the laughter syllable level), thus providing further evidence that laughs are less often produced in a (stereo)typical manner in conversation, and that other laughter types occur more often.

Our analysis showed that the two laughter types differ from each other with respect to four of the five examined characteristics: duration, fundamental frequency, speech intensity and voice quality (represented by the *cpp*), but not with regards to

Table 3: *The performance of a classifier trained on the four acoustic features (f0, en, cpp, modSp) to distinguish either two classes (laughter vs. speech) or three classes (typical laughter vs. non-typical laughter vs. speech). We illustrate the average precision, recall, F-score and accuracy obtained in a 10-fold cross-validation experiment, for the discrimination laughter/speech (i.e., the two laughter classes in the 3-class classifier were merged for the evaluation).*

#classes	**Precision**	**Recall**	**F-score**	**Accuracy**
2	0.957	0.947	0.951	0.950
3	0.958	0.938	0.947	0.946

rhythm. We observed longer events, with higher f0, higher intensity and increased breathiness for typical laughs, compared to non-typical ones. Our results on duration are in line with previous findings on both conversational [8] and emotion-induced laughter [3], including for laughter produced by children [6]. The differences between laughter types seen for f0 mirror those of [6] in which rhythmic laughs had a higher f0 than all other types of laughter, but squeal. While the results of [7] differ from ours, one needs to take into account the small sample size of that study (one order of magnitude smaller than here and coming from one speaker only), and also the fact that it reported unreliable f0 estimation for one type of laughter.

The fact that the last examined feature, *modSp* – a representation of rhythm, did not differ between laughter types, confirms the hypothesis of Kipper and Todt [11] about a specific rhythm for laughter. This shows that, from a rhythm point of view, different laughter types represent one unique class, at least for the two types we analyzed here. This finding is in contrast to that on another type of laughter considered in many recent studies (e.g., [12, 8]), speech-laughs (the concurrent production of laughter and speech), which has been shown to differ from both speech and laughter in terms of its rhythm [12]. Regarding the findings concerning signal intensity, voice quality and rhythm, to our knowledge, this is the first study examining laughter types in conjunction with these characteristics, thus helping us better understand the acoustic realization of laughter.

As the function of laughter may also have an effect on its acoustic realization, we performed an additional analysis, looking within each of the two laughter types at shorter vs. longer laughs, assuming that these have a rather social and a rather mirthful function, respectively [36, 35]. It showed that the employed rhythm representation, *modSp*, had a lower value for shorter laughs, for both laughter types. This is not surprising, considering that this feature is computed using a large analysis window (1.5 seconds) and that, although the laughter event is at the center of this analysis window, shorter instances will have also some speech materials included to their left and to their right. The lower value makes sense, considering that speech has a substantially lower *modSp* value than laughter (see Table 1). This also explains why an automatic detection system using modulation spectrum information as feature fails to identify short laughter instances [34]. We also found lower f0 values for shorter laughs (albeit the difference was significant only for non-typical laughter), which aligns well with the finding of [13], that conversational laughter containing many instances having social functions exhibited lower f0 than laughter induced by means of humorous materials.

Although differences between laughter types with respect to their function (spontaneous vs. volitional) have been previously found for part of the features employed here also in other studies (e.g., for f0, but not for intensity [14]; for f0, duration and intensity, but not for voice quality [15]), we must note that social laughter is not the same as the volitional laughter. In these studies, volitional laughter was prompted from the speakers in isolation, by asking them to laugh without an external stimulus, while social laughter occurs in interaction. In a study employing more similar materials to here [16], laughs produced in conversation were rated by listeners as either voluntary or involuntary (which could correspond to social vs. mirthful). It found involuntary laughter to be longer, to have a higher f0 and a lower intensity, but no difference in any voice quality measure they investigated, all these being in line with the trends seen in our data for longer laughs. Thus, it seems that the operationalization we used here for laughter function (based on the length of the laughs), while imperfect, still captures acoustic differences reported in more controlled studies, while allowing us to make predictions on prosodic characteristics which have not been investigated before in this context (rhythm).

Based on the findings of the acoustic analysis, at least part of the examined features should be able to discriminate between the two types of laughter investigated here. We tested this, by using these four features in a more ecological setting, that takes into account also speech as one of the predicted classes (as it is normally done in speech technology applications, discriminating between laughter and speech). We have seen that the system distinguishing between speech and the two laughter types obtained an overall lower performance, when considering one additional class (similar to [28, 26]; although the most confused classes differed: the laughter ones, here, and voiced laughter/speech, in [28]). Nevertheless, it returned similar results to the two-class system, for laughter vs. speech classification, while having the advantage of offering a more detailed classification of laughter, into typical and non-typical types, with an adequate accuracy.

The laughter detection performance obtained here is high, compared to other proposed systems. However, in our experiment we only had to classify already cut segments, while in a real-use scenario one needs to perform a prior segmentation into some sort of classification unit (e.g., syllables) or use a frame-based decision system. One must note though, that in order to be able to take advantage of these findings in an automated laughter detection system, a larger analysis unit than the frame may need to to be considered, as individual frame values of the employed prosodic features might differ considerably from the average value across longer segments. There is also evidence supporting this claim, with prosodic representations which have been shown to discriminate between laughter and speech, such as those for rhythm [12], not improving the classification performance when used in a frame-based detection system [19].

5. Conclusions

Investigating the acoustic-prosodic marking of two types of laughter, typical and non-typical, we have seen that features that represent duration, fundamental frequency, speech intensity and voice quality differ between the two classes, while a measure characterizing rhythm does not. However, this latter feature has shown to reliably discriminate between shorter and longer laughs, within each of the two laughter classes, suggesting a possible role for it for the categorization of laughter based not on phonetic form, but on function (social vs. mirthful). Using the prosodic features in an automatic classification system, we observed that a classifier considering three classes (typical laughter, non-typical laughter, speech) has a similar laughter/speech discrimination performance to that of a two-class, system, with the added benefit of a more detailed classification of laughs. The findings of this study help increase our understanding of how laughter is realized in conversation, including how this new information may be exploited in automatic systems.

6. Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 461442180.

7. References

[1] R. Provine and Y. Yong, "Laughter: A stereotyped human vocalization," *Ethology*, vol. 89, no. 2, pp. 115–124, 1991.

[2] J. Trouvain, "Segmenting phonetic units in laughter," in *Proc. of ICPhS*, 2003, pp. 2793–2796.

[3] J.-A. Bachorowski, M. J. Smoski, and M. J. Owren, "The acoustic features of human laughter," *The Journal of the Acoustical Society of America*, vol. 110, no. 3, pp. 1581–1597, 2001.

[4] K. Wang, C. Ishi, and R. Hayashi, "Laughter patterns in multi-speaker conversation data: comparison between spontaneous laughter and intentional laughter," in *Proc. of ICPhS*, 2023, pp. 1771–1775.

[5] K. Grammer and I. Eibl-Eibesfeldt, "The ritualisation of laughter," *Natürlichkeit der Sprache und der Kultur*, vol. 18, pp. 192–214, 1990.

[6] E. E. Nwokah, P. Davies, A. Islam, H.-C. Hsu, and A. Fogel, "Vocal affect in three-year-olds: A quantitative acoustic analysis of child laughter," *The Journal of the Acoustical Society of America*, vol. 94, no. 6, pp. 3076–3090, 1993.

[7] H. Tanaka and N. Campbell, "Acoustic features of four types of laughter in natural conversational speech," in *Proc. of ICPhS*, 2011, pp. 1958–1961.

[8] T. Rathcke and S. Fuchs, "Laugh is in the air: An exploratory analysis of laughter during speed dating," *Frontiers in Communication*, vol. 7, p. 170, 2022.

[9] B. Ludusan and P. Wagner, "ha-HA-hhä? Intensity and voice quality characteristics of laughter," in *Proc. of Speech Prosody*, 2022, pp. 560–564.

[10] J. Esling, "States of the larynx in laughter," in *Proc. of Interdisciplinary Workshop on the Phonetics of Laughter*, 2007, pp. 15–20.

[11] S. Kipper and D. Todt, "The role of rhythm and pitch in the evaluation of human laughter," *Journal of Nonverbal Behavior*, vol. 27, no. 4, pp. 255–272, 2003.

[12] B. Ludusan and P. Wagner, "Speech, laughter and everything in between: A modulation spectrum-based analysis," in *Proc. of Speech Prosody*, 2020, pp. 995–999.

[13] J. Vettin and D. Todt, "Laughter in conversation: Features of occurrence and acoustic structure," *Journal of Nonverbal Behavior*, vol. 28, no. 2, pp. 93–115, 2004.

[14] G. A. Bryant and C. A. Aktipis, "The animal nature of spontaneous human laughter," *Evolution and Human Behavior*, vol. 35, no. 4, pp. 327–335, 2014.

[15] N. Lavan, S. K. Scott, and C. McGettigan, "Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter," *Journal of Nonverbal Behavior*, vol. 40, pp. 133–149, 2016.

[16] V. Krepsz, V. Horváth, A. Huszár, T. Neuberger, and D. Gyarmathy, ""Should we laugh?" Acoustic features of (in)voluntary laughters in spontaneous conversations," *Cognitive Processing*, pp. 1–18, 2023.

[17] S. Cosentino, S. Sessa, and A. Takanishi, "Quantitative laughter detection, measurement, and classification – A critical survey," *IEEE Reviews in Biomedical Engineering*, vol. 9, pp. 148–162, 2016.

[18] M. T. Knox and N. Mirghafori, "Automatic laughter detection using neural networks." in *Proc. of INTERSPEECH*, 2007, pp. 2973–2976.

[19] L. S. Kennedy and D. P. Ellis, "Laughter detection in meetings," in *Proc. of NIST ICASSP 2004 Meeting Recognition Workshop*, 2004.

[20] T. F. Krikke and K. P. Truong, "Detection of nonverbal vocalizations using gaussian mixture models: looking for fillers and laughter in conversational speech," in *Proc. of INTERSPEECH*, 2013, pp. 163–167.

[21] B. Ludusan, "The usefulness of phonetically-motivated features for automatic laughter detection," in *Proc. of DiSS Workshop*, 2023, pp. 33–37.

[22] J. Oh, E. Cho, and M. Slaney, "Characteristic contours of syllabic-level units in laughter," in *Proc. of INTERSPEECH*, 2013, pp. 158–162.

[23] G. Rennie, O. Perepelkina, and A. Vinciarelli, "Which model is best: Comparing methods and metrics for automatic laughter detection in a naturalistic conversational dataset," in *Proc. of INTERSPEECH*, 2022, pp. 4008–4012.

[24] K. Laskowski, "Contrasting emotion-bearing laughter types in multiparticipant vocal activity detection for meetings," in *Proc. of ICASSP*, 2009, pp. 4765–4768.

[25] S. Petridis, B. Martinez, and M. Pantic, "The MAHNOB laughter database," *Image and Vision Computing*, vol. 31, no. 2, pp. 186–202, 2013.

[26] R. B. Kantharaju, F. Ringeval, and L. Besacier, "Automatic recognition of affective laughter in spontaneous dyadic interactions from audiovisual signals," in *Proc. of ICMI*, 2018, pp. 220–228.

[27] G. Gosztolya, A. Beke, and T. Neuberger, "Differentiating laughter types via hmm/dnn and probabilistic sampling," in *Proc. of SPECOM*, 2019, pp. 122–132.

[28] S. Petridis and M. Pantic, "Audiovisual discrimination between laughter and speech," in *Proc. of ICASSP*, 2008, pp. 5117–5120.

[29] 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.

[30] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, "VoiceSauce: A program for voice analysis," in *Proc. of ICPhS*, 2011, pp. 1846–1849.

[31] L. Varnet, "Matlab toolbox for the computation of amplitude- and frequency- modulation spectra," https://github.com/LeoVarnet/AM_FM_Spectra, 2018.

[32] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 4, pp. 769–778, 1994.

[33] L. Varnet, M. C. Ortiz-Barajas, R. G. Erra, J. Gervain, and C. Lorenzi, "A cross-linguistic study of speech modulation spectra," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 1976–1989, 2017.

[34] B. Ludusan and P. Wagner, "An evaluation of manual and semi-automatic laughter annotation." in *Proc. of INTERSPEECH*, 2020, pp. 621–625.

[35] M. Koutsombogera and C. Vogel, "Understanding laughter in dialog," *Cognitive Computation*, vol. 14, no. 4, pp. 1405–1420, 2022.

[36] T. Shochi, M. Guerry, J.-l. Rouas, M. Chaumont, T. Nishida, and Y. Ohmoto, "Perceptual and acoustic correlates of spontaneous vs. social laughter," in *Proc. of International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots*, 2017, pp. 37–41.

[37] 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/

[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.