Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

ha-HA-hha̤? Intensity and voice quality characteristics of laughter

Bogdan Ludusan, Petra Wagner

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

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

Abstract

Laughter is one of the most widely-encountered paralinguistic phenomena in human interaction and it has been studied from different perspectives throughout the years, including its acoustic-prosodic realization. However, previous studies have mostly focused on fundamental frequency and duration measures, with other prosodic features being less studied. We examine here the acoustic marking of laughter in terms of intensity and voice quality characteristics, by using a corpus of spontaneous conversations. We operationalized the two cues by means of the root-mean-square energy and the cepstral peak prominence, respectively. Examining laughs, speech-laughs and speech instances at two different levels (that of the entire event and at the syllable nucleus level) we observed the least regular phonation for laughs and the most regular one for speech, while intensity was the highest for speech-laughs, followed by laughter and the lowest for speech. Using mixed effect models we determined that all three vocalization classes differ significantly from one another, in terms of both acoustic cues. Moreover, an interesting effect of syllable position was seen for laughter, with phonation becoming more regular for later syllables.

Index Terms: laughter, speech intensity, voice quality, cepstral peak prominence, laugh, speech-laugh, speech

1. Introduction

Laughter has been studied in various fields and from different perspectives over the years (for an overview of laughter research, see [1]). One of the main aspects investigated is its acoustic characterization. A number of studies have focused on the acoustic features of laughter, looking at prosodic characteristics, such as fundamental frequency, duration or intensity.

The fundamental frequency (f0) of the voice has been extensively examined, at both the level of the entire event and at the individual syllable level. It has been shown that, compared to speech, laughter exhibits higher average f0 values [2, 3], as well as considerably higher f0 range and maximum values [2, 3, 4]. Furthermore, while f0 declination is a common phenomenon in speech, it does not seem to occur for laughter [5, 4]. Nevertheless, significant f0 variation between the syllables belonging to the same laughter event has been observed, due to very high intra-individual variation [6].

Studies that examined the temporal and distributional aspects of laughter have also looked at the duration of laughter syllables and the timing between them. For instance, it was shown that the overall length of a laughter event is positively correlated to the number of syllables within it [2] and that a high variability of the inter-syllabic intervals exists [6]. Some investigations suggested a more dense production of syllables (as opposed to silences) at the laughter onset [4], while others offered different explanations (such as a shorter syllables towards the end of an event [7]). The differences found between the duration of the silence intervals and in the number of syllables per event may be caused by the type of elicited laughter (mirthful vs. non-mirthful) [6], the same study showing that, even within the same type of laughter, these measures may vary based on the conversational context.

As for intensity, while it has been suggested it plays an important role in the perception of laughter (e.g. [8]), only a few studies have actually investigated it. Among them, some limited themselves to a qualitative description of laughter intensity characteristics [9, 10]. Those reporting quantitative intensity measures looked only at laughter [7], at specific laughter dimensions (e.g. arousal, valence) [11, 12], or compared laughter produced by normal-hearing and deaf populations [13]. The few works that compared speech and laughter produced by the same speakers, from a signal intensity point of view, concluded that there is no difference between the two types of vocalizations [5, 14]. This is in contrast to evidence coming from speech technology, where automated systems for laughter detection employ intensity as one of the discriminating features (e.g. [15, 16]). Thus, there seems to be limited knowledge about the role intensity plays in the production of laughter compared to other prosodic features. Moreover, as the previous two studies comparing speech and laughter intensity profiles employed a small number of samples, taken from a few speakers only, we believe an analysis on a larger dataset is warranted.

Other prosodic characteristics investigated in relation to laughter include rhythm and voice quality. Based on existing investigations looking at air flow, pressure and muscle activity during the production of laughter, it has been suggested that laughter has its own distinct rhythm [10]. [17] found evidence for the role of rhythm in the perception of laughter quality, leading the authors to consider rhythmic structure as a suitable knowledge source for making a decision whether a succession of vocal elements represents laughter or not. This hypothesis has been confirmed by [18], in which a speech-based rhythm representation has been successful in differentiating between laughter and speech.

Voice quality is used to convey different types of information in human communication [19], but its role is less studied than that of other prosodic characteristics. A study comparing laughter vowels to speech vowels [20] reported an increase in the F1 value of the former, the authors arguing that laughter vowels were produced with a "pressed" voice quality, stemming from physiological constraints in the pharyngeal region. In a subsequent study, investigating the coarticulation effects of laughter on preceding speech, a similar effect (an increased F1) was observed also for the speech vowel preceding laughter [21]. As evidence exists that laughter may be associated with a distinctive voice quality, we would like to explore further the connection between voice quality and laughter and how it may be employed to discriminating laughter from speech.

10.21437/SpeechProsody.2022-114

2. Methods and materials

2.1. Materials

We investigated laughter by means of the DUEL corpus [22], which consists of spontaneous dyadic interactions. It includes recordings in three languages: French, German and Mandarin Chinese, and we chose here materials from the German subpart. As each sub-part contains three different interaction scenarios (Dream Apartment, Film Script and Border Control), we selected for our analyses the scenario having the most laughter. Thus, the recordings corresponding to the Film Script scenario were employed.

The materials included more than two hours of data from 10 dyads, composed of mainly friends and acquaintances. In the Film Script scenario, the participants were asked to come up with the script for a film, based on an embarrassing event, which may be inspired from personal experience. This scenario contains many instances of mirthful laughter, but also social laughter, in particular for signalling embarrassment. The corpus was fully transcribed and segmented at the speaker-turn level, including annotations for laughter and other conversational phenomena, such as disfluencies and hesitations. Moreover, two classes of laughter were annotated: laughs and speech-laughs, with the latter class representing simultaneous productions of speech and laughter, in which neither of the two components is dominant. Therefore, we compared in this study the acoustic marking of laughs (LG), speech-laughs (SL) and speech (SP).

2.2. Analysis

We considered two analysis levels: the interval level and the syllable nucleus level. The former level was composed of the entire laugh or speech-laugh event, while for speech it was represented by the speaker turn. Non-verbal vocalizations, such as coughs or in-breaths, and the parts of the recordings marked as unclear by the annotators were not included in the analysis. When these or laughter interrupted a speaker's turn, two interval units were formed, one composed of the turn until the phenomenon and the other one consisting of the remaining of the turn, after the phenomenon. The syllable nuclei were determined by applying a script for the automatic detection of nuclei [23]. The obtained nuclei were then numbered according to their position in the corresponding interval. Table 1 shows the total number of intervals and syllable nuclei considered for each class in this analysis, as well as the mean and maximum number of syllable nuclei found in an interval.

Table 1: *Statistics about the number of analyzed units (intervals/nuclei) corresponding to each of the three studied vocalization classes (laugh - LG, speech-laugh - SL, speech - SP).*

Type	**Intervals**	**Nuclei**	**Nuclei/Interval**	
			mean	**max**
LG	607	1740	3.8	27
SL	257	1593	6.8	51
SP	2830	24894	10.7	118

We operationalized speech intensity and voice quality by means of the root-mean-square energy and the cepstral peak prominence, respectively. Cepstral peak prominence (cpp) represents the amplitude of the cepstral peak relative to the regression line over the entire cepstrum [24], with noisier (less periodic) signals exhibiting lower cpp values. Previous work has shown cpp to be the best signal-based correlate for perceived voice quality in continuous speech [25], being able to capture also changes in speech characteristics due to linguistic processes (e.g. prosodic prominence [26]).

The values of the root-mean-square energy and of the cepstral peak prominence were extracted using the VoiceSauce software [27], using a variable window length equal to five pitch periods and a frame shift of 1 ms. We then computed the mean (*meanInt*) and the maximum (*maxInt*) of the root-mean-square energy as well as the mean of the cepstral peak prominence (*cpp*) within each analysis unit. The analysis unit at the interval level was the length of the entire interval, while at the nucleus level it was the analysis frame containing the time instant corresponding to the nucleus marker given by the automatic detection script, together with the preceding and the following 12 frames. Figure 1 illustrates the employed analysis levels on a laugh produced by one of the speakers in the corpus. As silence intervals might have an effect on the values of two of the investigated acoustic measures (*intMean* and *cpp*), we used the voice activity detection given by the automatic script and we removed the chunks marked as silence from the analysis units.

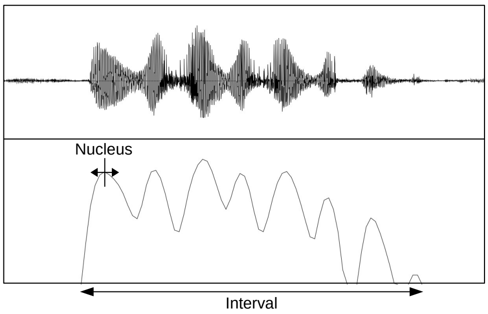

Figure 1: *The analysis units at which the acoustic cues are computed. The upper panel shows the waveform and the lower panel the intensity profile of a laugh from the corpus. The interval unit includes the entire laughter event, while the nucleus unit consists of the point where the automatic script determined the position of a nucleus, along with the previous and the following 12 frames.*

After computing the three acoustic measures for the laugh, speech-laugh and speech classes, at the two analysis levels, we applied linear mixed effects models to test if significant differences exist in how the three classes are marked by these acoustic cues. Separate models were fitted for each of the three dependent variables, with the vocalization class as fixed factor, and the speaker as random factor (random slope). At the nucleus level, we included an additional random factor, the position of the nucleus inside the event/turn. We also considered a laughter (LT) class, by merging the laugh and speech-laugh classes, and we evaluated whether it differed from the speech class.

Finally, we tested whether the position of the nucleus within each interval has an effect on any of the acoustic cues. For each of the three vocalization classes, we fitted linear regression models with the values of the acoustic cues as dependent variable and the position of the nucleus (numeric) as predictor. For the classes and acoustic cues for which a significant effect was observed we investigated this in more detail, by transforming the position predictor to a categorical variable and refitting the linear models. For all the statistical analysis we used the functions offered by the R software [28].

Table 2: *Estimates (β), t-values and significance level (*: $p < .05$, **: $p < .01$, ***: $p < .001$) of the difference between each pair of classes considered in the analysis, for the three analyzed cues. These were obtained by means of linear mixed effects models with the cues as dependent variables and the vocalization class (laugh LG, speech-laugh SL, or speech SP) as predictor. The random effect for the interval-based models was the speaker, while the nucleus-level analysis considered speaker and position as random effects. We also investigated the differences between two classes, laughter LT (laughs and speech-laughs together) and speech.*

Analysis level	**# classes**	**Difference**	**intMean**		**intMax**		**cpp**	
			β	t-value	β	t-value	β	t-value
Interval	2	SP-LT	-0.361	-12.09 (***)	-4.209	-16.18 (***)	0.753	45.47 (***)
	3	SL-LG	0.389	6.904 (***)	4.393	8.971 (***)	0.358	11.62 (***)
		SP-SL	-0.634	-12.82 (***)	-7.293	-16.98 (***)	0.501	18.52 (***)
		SP-LG	-0.245	-7.186 (***)	-2.900	-9.803 (***)	0.859	46.03 (***)
Nucleus	2	SP-LT	-2.208	-42.09 (***)	-2.729	-45.25 (***)	0.972	52.12 (***)
	3	SL-LG	0.232	2.452 (*)	0.193	1.770	0.519	15.44 (***)
		SP-SL	-2.329	-32.32 (***)	-2.829	-34.15 (***)	0.700	27.35 (***)
		SP-LG	-2.097	-30.25 (***)	-2.636	-33.09 (***)	1.219	49.71 (***)

3. Results

The values of *intMean*, *intMax* and *cpp*, for the three vocalization classes, at the two different analysis levels are illustrated in Figure 2. It shows that, at both analysis levels, a similar trend can be observed: laughter has the lowest *cpp* values (least regular phonation), followed by speech-laughs, with the most regular phonation (highest *cpp* values) being obtained for speech. The speech intensity measures show a similar trend not only between the two analysis levels, but also among themselves. In both cases, the class having the highest values was the speech-laugh class, followed by laugh and, lastly, by speech.

We tested the significance of the differences seen between classes by means of linear mixed effects models, for which the dependent variables were z-normalized on a per-speaker basis, using successive differences contrasts. The results of each pairwise comparison are presented in Table 2. We notice that the investigated acoustic cues mark differently the three vocalization classes, with all but one of the pairwise differences being significant. The only difference that did not reach significance was for *intMax* between speech-laugh and laugh, at the nucleus level. Also when grouping laughs and speech-laughs into one class, laughter, the differences between this class and the speech class were always significant.

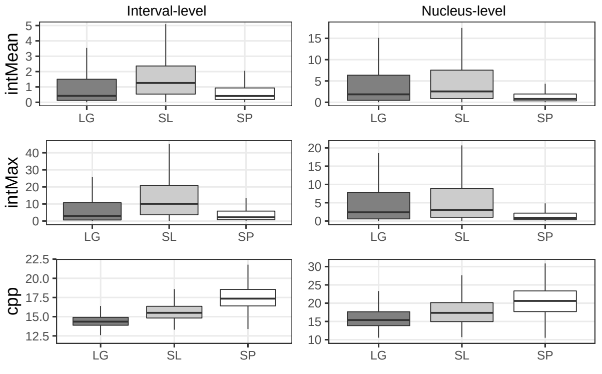

Figure 2: *Boxplots of the values (median and first and third quartiles) of the three investigated acoustic measures (intMean, intMax and cpp) across the two analysis levels (interval- and nucleus-level). The plots do not show the outlier values.*

To determine any position effects on the investigated cues, we analyzed the linear models fitted for this purpose. For laughs, only for *cpp* did the position of the nucleus within the event have a significant effect ($t = 3.253, p = 0.001$). The model revealed that a unit increase in the position increases *cpp* by 0.017 standard deviations. For the other two acoustic measures, the difference did not reach significance, $p = 0.054$ for *intMean* and $p = 0.098$ for *intMax*. For speech-laughs instead, no position effect on *cpp* values was observed ($p = 0.343$), but significant effects were seen for *intMean* ($t = -2.891, p = 0.004$) and *intMax* ($t = -2.743, p = 0.006$). A similar outcome was obtained also for speech: no effect on *cpp* ($p = 0.181$), only on the intensity measures: *intMean* ($t = -10.24, p < 2e^{-16}$) and *intMax* ($t = -10.32, p < 2e^{-16}$).

We then fitted the models that showed a significant effect of position using the position as categorical predictor. In Figure 3 we display the estimates, for the first 10 positions, of two of these models, considering *cpp* as dependent variable for laughs (cpp_LG) and the maximum intensity for speech (intMax_SP), respectively. A low cpp value can be seen for the first nucleus in an interval and higher values for the subsequent position. For speech intMax, instead, a clearer decreasing trend may be observed.

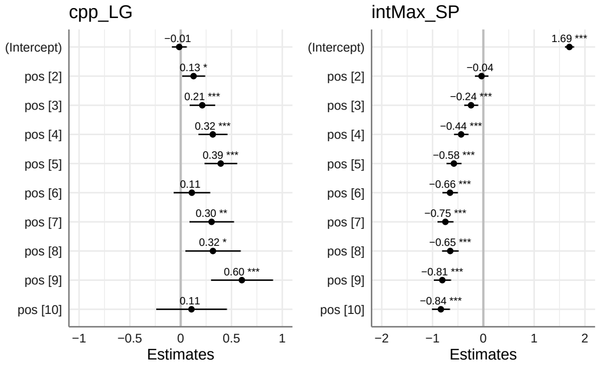

Figure 3: *The estimates of the first ten positions, for the models fitted with laugh cpp values (cpp_LG) and with speech intMax values (intMax_SP). The Intercept represents the value of the first position, with the other values being the difference between the first position and that position. The estimates value and the significance of the difference to the intercept are also displayed.*

4. Discussion and conclusions

The findings of our study show that laughs, speech-laughs and speech are marked differently in terms of intensity and voice quality characteristics, both at the level of the entire interval (laughter event/speaker turn) and at the level of individual syllable nuclei. These differences seem to persist also when one collapses laughs and speech-laughs into one class, laughter. Moreover, our results provide evidence that speech-laughs are indeed a combination of laughs and speech, based on the two acoustic-prosodic characteristics studied here, as well: Speech-laughs exhibit an intermediate level of cepstral peak prominence, between that of laughs and speech, and a higher intensity level than both speech or laughs, as to be expected from a phenomenon seen as the superposition of these two vocalization classes.

In terms of voice quality, we have seen that laughs show the least regular phonation, and speech the most regular one. Lower cpp values characterize a breathy voice quality and this breathy characteristic of laughter may be explained by the laryngeal setting that characterizes laughter. A breathy voice quality was similarly observed in a laryngoscopic investigation of laughter [29], explained by a wider opened glottis, and a higher volume airflow than for modal voice speech, in the abduction phase. Our findings seem to be in contrast to those of [20], where the authors reported a raised F1 in laughter as evidence of a pressed voice quality. It may be that we use here a narrower definition of voice quality (referring only to phonatory aspects), while the previous work considered this, along with supra-laryngeal or filter-related characteristics. Another study that found the same characteristic (a raised F1) [21], also found an increased center of gravity in the spectrum (usually associated with arousal), which may partly explain the raised F1 values. Further investigations would be necessary to understand these aspects, since raised F1 may be due to several voice quality settings.

An interesting positional effect was noticed for laughs, characterized by an apparent increase in cpp throughout the event, corresponding to a more regular phonation for syllables found later within the laughter event. Additionally, we observed some regular changes in the trend seen for the cpp values, when looking at the first ten syllable nucleus positions. We consider these as indications of newly beginning exhalation phases that often occur in (long) laughter events. Nevertheless, the cpp values of the positions where these changes occur, do not fall below that of the first nucleus. Thus, we do not have an indication of a full "phonation reset" within single laughter events.

The class exhibiting the highest intensity at the two analysis levels, both in term of average and maximum value, was the class of speech-laughs. Laughs exhibited a lower value, and speech showed the lowest intensity overall. Although these results stand in contrast to those of previous studies comparing laughter and speech intensity (no difference between laughter and speech [5, 14]), one must note that these studies used a low number of laughter samples (5-10 [5], or a number of laughter events having certain acoustic characteristics [14]), uttered by a few speakers (2 and 3, respectively). Our analysis, considering a set of 20 speakers and including all laughs and speech-laughs produced, showed that laughter is marked differently than speech from an intensity point of view, as likewise indicated by automatic systems employing this cue for discrimination (e.g. [15]).

We have also found a significant position effect, showing a decreasing trend on intensity for both speech-laughs and speech. We interpret this trend in speech-laughs as the similar energy declination effect that has been established for speech [30]. Taking into account that both intensity and f0 are subject to declination effects throughout the utterance [30] and that laughs do not appear to exhibit f0 declination effects [4], our results, showing no significant effect of position on intensity in the case of laughs, may seem coherent. These findings do not align with those of a previous study on laughter that reported a negative correlation between syllable position and syllable intensity [7]. The different results may be due to the somewhat different mix of laughter types included in our data (a mix of social and mirthful laughter here vs. mirthful laughter in [7]). This interpretation is supported by previous studies which found that acoustic-phonetic differences in laughter characteristics could be explained by the type of the elicited laughter (e.g. [6] for duration). However, as we did not control for laughter function in our data, this interpretation is tentative and in need of further empirical analysis.

We acknowledge the fact that the automated procedure for nuclei detection may introduce errors (by not detecting, or by falsely detecting nuclei), but we believe that our conclusions are not affected by them, since the errors affect all three vocalization classes. The inclusion of a relatively high number of instances, from each class, in the analysis would counteract any noise introduced in the analysis by errors in the nuclei detection procedure. Moreover, similar results were obtained at both analysis levels (including the interval level, which does not use information related to the automatically obtained syllable nuclei), suggesting that the findings are robust.

Summing up, our study provides further evidence on the role that acoustic-prosodic cues play in the marking and characterization of laughter, and shows that the investigation of two lesser studied prosodic characteristics, such as intensity and voice quality, is fruitful for getting a better understanding of the fine-grained differences between three qualitatively different, but relevant modes of human vocalization. It is highly likely that their use may be beneficial not just for the automatic detection of laughter, but also for a differentiated characterization of laughter types and, potentially, functions.

We intend to further extend this investigation, by looking at the acoustic marking of different types of laughter and by analyzing their perceived quality. The majority of laughter studies considered only "typical" laughter (consisting of rhythmic, voiced, alternations), or made no distinction between the different types of laughs appearing in their data. Yet, more than half of the laughter produced by speakers is of the non-typical variety [4] being acoustically produced in a wide variety of manners [31]. Furthermore, this latter type of laughter seems to be worse recognized than typical laughter by automatic detection systems [32], presumably due to the increased variation present in its acoustic marking. The fact that the voice quality feature investigated here was highly discriminative between laughter and speech and that non-typical laughter types may be associated with specific voice quality traits (e.g. nasal, chuckle, ingressive), suggests that this feature may be useful for discriminating between non-typical laughter and speech or even between non-typical and typical laughter.

5. Acknowledgements

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

6. References

[1] J. Trouvain and K. Truong, "Laughter," in *The Routledge Handbook of Language and Humor*. Routledge, 2017, pp. 340–355.

[2] D. Mowrer, L. LaPointe, and J. Case, "Analysis of five acoustic correlates of laughter," *Journal of Nonverbal Behavior*, vol. 11, no. 3, pp. 191–199, 1987.

[3] H. Rothgänger, G. Hauser, A.-C. Cappellini, and A. Guidotti, "Analysis of laughter and speech sounds in Italian and German students," *Naturwissenschaften*, vol. 85, no. 8, pp. 394–402, 1998.

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

[5] C. Bickley and S. Hunnicutt, "Acoustic analysis of laughter," in *Proc. of ICSLP*, 1992, pp. 927–930.

[6] 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.

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

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

[9] M. Edmonson, "Notes on laughter," *Anthropological Linguistics*, vol. 29, no. 1, pp. 23–34, 1987.

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

[11] D. Szameitat, K. Alter, A. Szameitat, D. Wildgruber, A. Sterr, and C. Darwin, "Acoustic profiles of distinct emotional expressions in laughter," *The Journal of the Acoustical Society of America*, vol. 126, no. 1, pp. 354–366, 2009.

[12] N. Lavan, S. Scott, and C. McGettigan, "Laugh like you mean it: Authenticity modulates acoustic, physiological and perceptual properties of laughter," *Journal of Nonverbal Behavior*, vol. 40, no. 2, pp. 133–149, 2016.

[13] M. Makagon, S. Funayama, and M. Owren, "An acoustic analysis of laughter produced by congenitally deaf and normally hearing college students," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 472–483, 2008.

[14] E. Nwokah, H.-C. Hsu, P. Davies, and A. Fogel, "The integration of laughter and speech in vocal communication: A dynamic systems perspective," *Journal of Speech, Language, and Hearing Research*, vol. 42, no. 4, pp. 880–894, 1999.

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

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

[17] 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.

[18] 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.

[19] C. Gobl and A. Ní Chasaide, "The role of voice quality in communicating emotion, mood and attitude," *Speech Communication*, vol. 40, no. 1-2, pp. 189–212, 2003.

[20] D. Szameitat, C. Darwin, A. Szameitat, D. Wildgruber, A. Sterr, S. Dietrich, and K. Alter, "Formant characteristics of laughter," in *Proc. of the Interdisciplinary Workshop on the Phonetics of Laughter*, 2007, pp. 9–13.

[21] B. Ludusan and P. Wagner, "No laughing matter: An investigation into the acoustic cues marking the use of laughter," in *Proc. of ICPhS*, 2019, pp. 2179–2182.

[22] 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.

[23] N. H. De Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior Research Methods*, vol. 41, no. 2, pp. 385–390, 2009.

[24] 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.

[25] Y. Maryn, N. Roy, M. De Bodt, P. Van Cauwenberge, and P. Corthals, "Acoustic measurement of overall voice quality: A meta-analysis," *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2619–2634, 2009.

[26] B. Ludusan, P. Wagner, and M. Włodarczak, "Cue interaction in the perception of prosodic prominence: The role of voice quality," in *Proc. of INTERSPEECH*, 2021, pp. 1006–1010.

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

[28] 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/

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

[30] J. Pierrehumbert, "The perception of fundamental frequency declination," *The Journal of the Acoustical Society of America*, vol. 66, no. 2, pp. 363–369, 1979.

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

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