INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Towards an annotation scheme for complex laughter in speech corpora

Khiet P. Truong[1], *Jürgen Trouvain*[2], *Michel-Pierre Jansen*[1]

[1]Human Media Interaction, University of Twente, Enschede, The Netherlands
[2]Language Science and Technology, Saarland University, Saarbrücken, Germany
k.p.truong@utwente.nl, trouvain@coli.uni-saarland.de, m.jansen-1@utwente.nl

Abstract

Although laughter research has gained quite some interest over the past few years, a shared description of how to annotate laughter and its sub-units is still missing. We present a first attempt towards an annotation scheme that contributes to improving the homogeneity and transparency with which laughter is annotated. This includes the integration of respiratory noises as well as stretches of speech-laughs, and to a limited extend to smiled speech and short silent intervals. Inter-annotator agreement is assessed while applying the scheme to different corpora where laughter is evoked through different methods and varying settings. Annotating laughter becomes more complex when the situation in which laughter occurs becomes more spontaneous and social. There is a substantial disagreement among the annotators with respect to temporal alignment (when does a unit start and when does it end) and unit classification, particularly the determination of starts/ends of laughter episodes. In summary, this detailed laughter annotation study reflects the need for better investigations of the various components of laughter.
Index Terms: speech recognition, human-computer interaction, computational paralinguistics

1. Introduction

For quite some time now, laughter has been a topic of research addressed from different angles - for example, automatic modeling of laughter, e.g., [1, 2, 3], an acoustic-phonetic analysis of laughter, e.g., [4, 5], or more from a conversational analytic point of view, e.g., [6, 7, 8, 9]. One of the prerequisites for carrying out that kind of research is that there are annotations available of laughter that, among others, indicate where laughter starts and ends. However, it appears that there is no shared annotation standard available for laughter, nor are there often clear descriptions available of how laughter is annotated in corpora [10]. This is rather undesirable as the way laughter is annotated and/or the lack of clear guidelines has an effect on the results of the analysis, and it affects the reliability and reproducibility of studies. Hence, our aim is to work towards an annotation scheme of laughter.

When annotating laughter, questions such as the following are expected to arise: "When does it start and when does it end?", "What are the units and sub-units?", "What is the role of speech-laughs, smiled speech, and voicedness in laughter and should we annotate that?". Other factors that play a role in the annotation are differences in the annotators' level of expertise for example, or differences in the situations in which laughter occurs. Posed laughter and laughter that is evoked by watching video stimuli alone in a lab room can have different characteristics than laughter in social conversation where it can be more often interspersed with speech and where smiled speech can also occur more often. In this context the following question is also relevant: "To what extend does the situation in which laughter occurs affect the quality of the annotations?"

In this paper, we will present our first attempt towards an annotation scheme of laughter, a set of guidelines that aims to achieve laughter annotations that are obtained in a more homogeneous way. We will apply the scheme to several corpora that contain laughter evoked in different ways to see how that affects the annotations. In addition, we analyze how often and where two annotators show (dis)agreement.

2. Related work

Trouvain [12] points out that there is an extreme heterogeneous terminology to denote units and sub-units of laughter. At the lowest level, the vowel-like segment is named "vocal peak" [13], "laugh pulse" [14], "note" [15], "laugh burst" [16], "call" [5], "syllabic vocalization" [17], "syllable" [15], "plosive" [18] or "vowel" [5, 19, 20]. The span from one vowel-like element to the next is referred to as "interpulse interval" [14], "laugh event" [13], "call" [5] or a "laugh syllable" [21].

At the next higher level the term "bout" is frequently used [5, 14, 22] defined as a sequence of laughter elements in one exhalation phase. Thus, an entire laugh can consist of several "bouts" separated by inhalation. The whole laugh is then called an "episode" [22, 15, 14], "laugh response" [16] or "laughing sound" [18]. The role of the inhalation phases often remains unclear, for instance whether this inhalation is part of the entire laugh. For Chafe [23] and Urbain et al. [24] the respiratory signals form the offset and possibly the onset of a laugh. But apart from the breath noises, there is no agreement how to define the start and the end of a laugh. An important distinction on the bout-level seems to be whether it contains (predominantly) voiced or unvoiced elements [5, 20]. In addition to distinguish voiced from voiceless bouts there are various attribute names to characterise the quality of laugh such as chuckle or giggle.

So far the laughs were isolated or standalone laughs, in contrast to speech-laughs, where laughter occurs simultaneously with articulation [25, 26]. This can be reflected by breathiness in the voiced portions and stronger aspiration in the unvoiced portions. Occasionally, a vibrato-like voice quality can be observed. Another important speech-synchronous form relevant for laughter is smiled speech. Audible smiling while speaking [27] can be distinguished from speech-laughs, as laughter and smiling should be treated as distinct categories.

In summary, previous research illustrates the variety and complexity in defining and labeling units and sub-units of laughter. As a result, many of the laughter annotations available in various databases have been annotated in different ways and are therefore difficult to compare. Hence, having shared guidelines of how to annotate laughter could be helpful.

http://dx.doi.org/10.21437/Interspeech.2019-1557

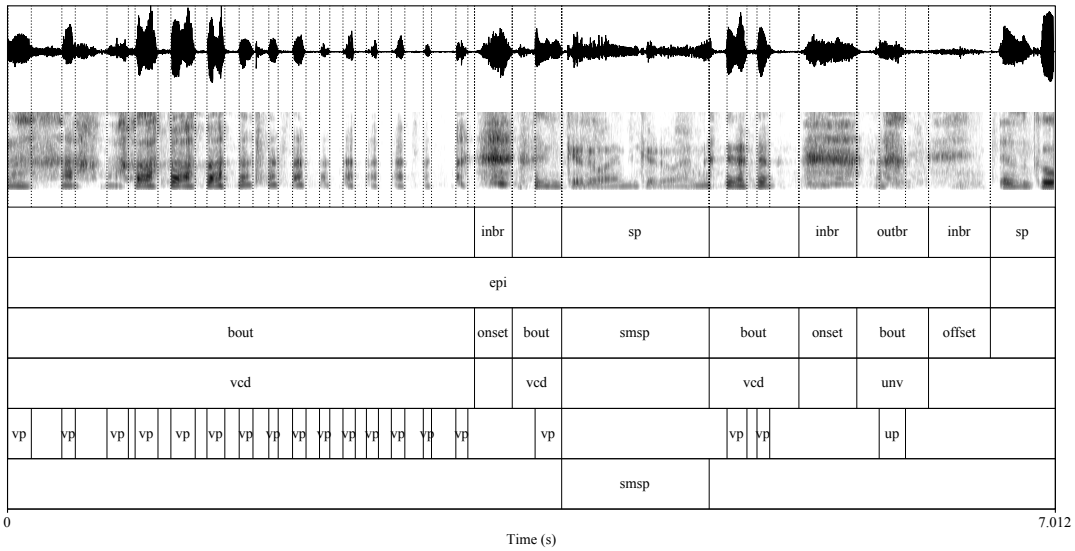

Figure 1: *Example of a 7 sec-laughter episode taken from [11] with the six levels of laughter annotation (see text for description).*

3. Approach

The annotation scheme proposed here is developed for audio recordings; annotations can be made based on auditory perception and visual inspection of the acoustic signal. However, it can easily be extended to audio-visual recordings as well.

There are six levels of annotation. On each level there are several categories specific for each level. Fig. 1 gives an illustration of the annotation scheme of a laughter episode.

Level 1 refers to any kind of speech and respiration, called SPEECH-RESP. Speech (SP) is necessary to find so-called speech-laughs and also smiled speech. On the respiratory level, we can differentiate between inbreath noises (INBR) and outbreath noises (OUTBR) which can both serve as an onset or offset of a laugh but they can also be part of a bout.

Level 2 refers to the EPISODE level, i.e. a laugh event at a higher level. The only category here is episode (EPI) that consists of one or several bouts with possible onsets and/or offsets and one or several speech-laughs. Where appropriate the boundaries from INBR and OUTBR (level 1) should be inherited.

Level 3 refers to the bout as a laugh event a lower level, the BOUT level. Every episode consists of one or more elements of the categories bout (BOUT), onset (ONSET), offset (OFFSET), speech-laugh (SPL) and possibly smiled speech (SMSP). A bout is a sequence of laugh-like syllables in one exhalation phase, sometimes a bout can consist of just one laugh-like syllable. A bout can be separated by inhalation noises that are identical with the inbreath noises from level 1, here serving as either onset or offset. In some cases the breath noise can be both and will be treated as interchangeable (for the analysis, we will map these two to ONOFFSET). A speech-laugh is a stretch of articulated speech that shows characteristics of a laugh. Smiled speech is a stretch of articulated speech that contains audible characteristics of smiling. Speech-laughs are always part of a laugh whereas smiled speech to be annotated may only occur in complex laughs, where it serves as a link between elements on the bout level. Where appropriate the boundaries from INBR and OUTBR (level 1) and EPI (level 2) should be inherited.

Level 4 refers to the sub-level of a bout that we call SUB-BOUT level. Each bout consists of one or several sub-bouts that can be voiced, unvoiced or silence. The voiced sub-bouts (VCD) mainly contain voiced pulses (or syllables) whereas the unvoiced sub-bouts (UVD) mostly contain unvoiced material, possibly also unvoiced pulses. There are sometimes silent phases within a bout (SIL) when there is no clearly audible signal.

Level 5 is the SEGMENTAL level. Here the pulses of bouts are categorised as voiced (VP) when there is a clear periodical structure present, or as unvoiced (UP). These pulses represent (rhythmical) beats in a bout, similar to syllables in speech.

Level 6 is optional and refers to the information of audible smiling. This information is always bound to speech (at least in acoustic-only corpora). In complex laugh events, a stretch of smiled speech (SMSP) can act as a link between elements on the bout level. It should be distinguished from a SPL (see level 3).

4. Material and Method

4.1. Material

Laughter can occur in various communicative situations and can also be evoked on purpose. Thus, the acoustic recordings containing laughter can be rather diverse. To represent this diversity we decided to study not only one database (with ideally rather "clean" data) but to consider laughter samples from several corpora to test our annotation scheme. In total we selected the following 6 corpora that differ in research purpose, communicative situation, and acoustic quality of which we randomly selected 10 items per corpus, see Table 1. For all databases selected, close-talk recordings were used in our annotations. However, when multiple speakers were involved, clean separation of the audio channels was not always achieved. In the ICSI and Mulai corpora for example, crosstalk was present in the recordings (but only the vocalizations of the speaker in question was annotated, even if this person was rather quiet in the selected sample). Such acoustic overlap heavily influences the annotation of laughter since most laughter is produced in the group simultaneously which leads to possible masking effects. In addition, some intense laughter vocalisations were blasted. The databases selected also cover a diversity of communicative situations that may influence the production and annotation of laughter.

Table 1: *Short descriptions of databases used (#spk=number of speakers selected – f=female, m=male, #spkconv=number of speakers in conversation, dur=average duration of samples selected in seconds, lang=language spoken)*

Database	Communicative situation	lang	dur	#spk	#spkconv
Montreal Affective Voices [28]	actors, posed ("happiness")	None	6s	5f, 5m	1
AVLaughterCycle [29]	watching funny video clips, evoked	None	15s	5f, 5m	1
GECO [30]	free conversation	German	10s	10f	2
Lindenstrasse [31]	task-based dialogs (talk about TV-soap related task)	German	10s	8f, 2m	2
ICSI [32]	authentic multi-party meetings	English	12s	5f, 5m	3–10
Mulai [11]	task-based dialogs (tell jokes and make the other laugh)	English	19s	5f, 5m	2

4.2. Annotation procedure and analysis

The first two authors who have some experience in laughter acoustics served as annotators of the above described material. Each annotator labelled and segmented independently of each other with the speech editor Praat [33] according to the agreed annotations guidelines. The 120 annotation files (2 annotators x 6 corpora x 10 audio samples) represent the base for the inter-rater agreement analysis. Finding an appropriate inter-rater agreement for our laughter annotation task that takes into account the joint tasks of unit *locating* (unitizing - when does the laugh start and end?) and unit *labeling* (categorization - is this a laugh or something else?) was not easy [34, 35]. In order to express (dis)agreement in unit labeling, we measure the number of inserted/deleted, substituted, and matched units of all annotators. In the current study, we limit ourselves to an analysis of the matched labels only. Each unit of each annotator is checked if it is 'matched' - a unit is counted as a match if there is a unit from another annotator that overlaps with the current one and that has the same label. The % of matched units is calculated by dividing the sum of all matched units by the sum of all units. In order to express (dis)agreement in unit locating, we measure the misaligned time (in ms) of each matched unit.

5. Results

5.1. Corpora

As expected there are substantial differences but also many parallels between the corpora. It is very evident from the absolute number of labels in Table 2 that the Montreal database behaves quite differently on **all** levels compared to the other corpora. The Montreal samples basically consist of **voiced bouts** whereas the conversational corpora and AVLC show a much more diverse picture. On the bout level, they often show offsets and onsets. On the sub-bout and the segmental level their amount of unvoiced elements is by far higher, sometimes the unvoicedness represent the half of all items on these two levels.

A remarkable difference can be seen between the corpora with and without conversational data. The conversational data show laughed speech (SPL) and smiled speech as integral part of an episode (SMSP) which can reach nearly the half of all elements on the bout level (Lindenstrasse). There is, however, one exception, namely ICSI, where no SPL and SMSP was found because our randomly selected samples were mainly taken from participants who were in the moment of laughing in a listener role in the multi-party meetings.

5.2. Label matching

The label matching that is illustrated in the middle part of Table 2 shows that there is a quite good match between the annotators regarding the episodes they labelled in all corpora. However, there is again a large difference between the Montreal database and the other corpora on the remaining four levels. With respect to the BOUT and SUB-BOUT level, agreement seems to be substantially lower, again, with the exception of the Montreal corpus. In general, ONOFFSET seem to cause some confusion among all corpora - not all inbreaths or outbreaths were equally well perceived. We note the low agreement of ONOFFSET in the Montreal corpus where 2 out of 3 in/outbreaths were disagreed on. The concepts of SPL and SMSP, and on a lower level those of UVD and SIL seem to cause some disagreement between the two annotators.

5.3. Misalign time

For the episodes the annotators are very close together for the Montreal database (29 ms), a bit further away for ICSI (291 ms) but very far away for the other four corpora (779-1531 ms) as can be seen in the right-hand part of Table 2. Speech-laughs and smiled speech differ between c. 35 ms and 529 ms which can roughly correspond to one and three syllables in conversational speech. In contrast, the misalignment at the segmental level is comparably small (between 10 and 44 ms) with unvoiced pulses showing the larger misalignment between the annotators than the voiced pulses.

6. Discussion

The results clearly show substantial differences between the Montreal dataset with posed laughter and the data with non-posed laughter where often social interaction is involved. Posed laughter as a sort of idealized and "prototypical" form of laughter shows a comparably high degree of homogeneity as laughs that consist of voiced bouts and sometimes an inbreath noise as offset. The annotation of these laughter samples was rather straightforward regarding labelling and alignment in time. An advantage of this extremely reduced variance is that it guarantees a high recognizability of laughter as an affective vocalization when there is no further context. All other samples were produced in mainly a social and interactive context. This increased the variance of phonetic forms in a tremendous way leading to rather complex laugh episodes with manifold challenges for annotation.

The complex laughs show many more onsets and offsets than the idealized posed laughs. This finding puts the respiratory noises into a more central role of laughing which is not necessarily common practice in laughter annotation so far. Inbreath noises often mark the offset of a laughter episode (sometimes with voiced inhalation) and outbreath noises can represent the onset of a bout but can simultaneously act as an unvoiced pulse in a series of pulses of the bout. Unvoiced sections in bouts are numerous in general, though not the majority, and unvoiced pulses show more temporal misalignment among both annotators than voiced pulses. This reflects a larger insecurity about the acoustic details of this kind of pulses in addition to the

Table 2: *Results of annotations of 2 annotators for all six corpora.*

	Montr	AVLC	GECO	Linden	ICSI	Mulai	Montr	AVLC	GECO	Linden	ICSI	Mulai	Montr	AVLC	GECO	Linden	ICSI	Mulai
	absolute no. of units						% of matched units						averaged misaligned time (in ms)					
speech-resp	4	81	109	115	102	162	100.0	80.2	85.3	80.9	69.6	78.4	8	72	117	65	143	192
inbr	4	74	35	53	47	60	100.0	87.8	77.1	77.4	80.9	75.0	8	72	68	47	106	194
outbr	0	6	8	6	45	20	na	0	50.0	33.3	51.1	20.0	na	na	240	185	172	352
sp	0	1	66	56	10	82	na	0	93.9	89.3	100.0	95.1	na	na	130	74	221	183
episode	22	35	26	23	36	46	100.0	100.0	100.0	95.7	88.9	100.0	29	1522	683	779	291	1139
bout	23	114	42	94	71	122	91.3	67.5	71.4	62.8	80.3	69.7	31	481	300	140	199	251
onoffset	3	50	8	38	27	45	33.3	68.0	50.0	71.1	92.6	71.1	11	115	406	52	100	191
bout	20	64	19	34	43	51	100.0	67.2	84.2	70.6	74.4	72.5	32	770	374	158	276	349
spl	0	0	10	12	0	11	na	na	80.0	41.7	na	72.7	na	na	168	529	na	108
smsp	0	0	5	10	1	15	na	na	40.0	30.0	0	53.3	na	na	35	145	na	179
sub-bout	24	109	24	53	56	62	91.7	61.5	66.7	64.2	73.2	67.7	36	385	168	86	138	239
vcd	22	53	12	31	27	30	100.0	75.5	75.0	83.9	77.8	83.3	36	202	78	98	114	163
uvd	1	28	10	11	20	20	0.0	53.6	70.0	27.3	80.0	75.0	0	1111	284	40	190	389
sil	1	28	2	11	9	12	0.0	42.9	0.0	45.5	44.4	16.7	0	90	0	55	60	72
segmental	148	375	43	100	196	172	98.6	77.6	88.4	68.0	90.3	72.7	12	17	16	37	19	24
vp	139	253	28	75	128	86	99.3	85.4	96.4	84.0	97.7	81.4	11	11	10	36	19	26
up	9	122	15	25	68	86	88.9	61.5	73.3	20.0	76.5	64.0	34	35	32	44	22	22

strong outbreath noises. Moreover, the consideration of silent or near-silent sections of a bout seems to be a rather new descriptive unit that could be found in all six corpora. The results seem to legitimate the three-way distinction of bouts with voiced, unvoiced, and silent attributes.

Another challenge for the annotation of laughs in speech corpora (that excludes the Montreal and the AVLC databases) are speech-laughs and smiled speech. Although the latter should not be part of a laugh by strict definition it sometimes had a linking character in a middle part of a longer episode. However, the agreement on smiled speech links is rather low which exhibits the unclear situation of how to handle those pieces. In addition, there were cases of complex laughs in which it was not clear at all whether some syllables should be accounted as smiled speech or as laughed speech. These unsatisfying circumstances call for more detailed research.

The most obvious difficulty of this annotation study of complex laughs is manifested in the question whether we face one or two (or more) laughter episodes. Though there is a very high number of events where the labels of both annotators overlap in time, the exact placement of boundaries of episodes shows tremendous differences. This contradictory situation can be explained by the fact that rather often one annotator considered a given laugh event as one episode and the other annotator as two episodes which resulted in large divergences in time. Probably this problem cannot be satisfyingly solved, it resembles the phenomenon in speech where two syntactically main clauses, maybe separated by a short pause, can be considered as two autonomous sentences or as a coordinated paratactic construction.

7. Conclusions

Many descriptions of the phonetic make-up of laughs and their sub-units were based on rather simplistic views, probably because only "simple" laughs were under consideration. The resulting possibilities for annotation is obviously not sufficient for an adequate treatment of complex laughs. Thus, we observe differences in annotation quality between corpora in which laughter is evoked in different ways. Misalignment in time and errors in identifying the label are much lower when laughter is posed evoked by stimuli than in spontaneous laughter in interaction.

We provide here a refined annotation scheme that attempts to also take into account the huge variance found in complex laughs as revealed in conversational speech. The resulting annotation scheme could help increase transparency and agreement on for example when a laugh starts and ends that is relevant for research where laughter needs to be annotated. Although there were very intensive discussion needed before the annotation guidelines were set up in the current version, the proposed scheme can only be viewed as a draft and hopefully as the beginning of an annotation standard. The challenges were found on different levels, for instance while we observe that some of the units are relatively easy to identify, annotators disagree on when that unit starts and ends, and vice versa whereas some units are relatively hard to identify but agreement on the start and end of that unit appears to be better.

The task of providing rather detailed annotation guidelines forced us to touch upon challenges not addressed so far, including the handling of short silent phases but also the possible integration of smiled speech into the entire laugh event. The latter evoked considerable disagreement in determining whether speech has a smiling quality or not and whether it has a linking character between bouts of a laughter episode.

This annotation scheme concentrated so far only on acoustic data and deliberately excluded visual data. One reason is that most speech corpora are available as audio-only data. It is the hope that the parallel analysis and annotation of audio and visual data with laughter, e.g. the Mulai corpus [11], can reveal a better handling and understanding of the special challenges of complex laughs such as silent intervals, speech-laughs, smiled speech, and ends and starts of episodes.

8. Acknowledgements

We thank Emer Gilmartin and Kevin El Haddad for intensive discussions on similar topics a longer time ago.

9. References

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

[2] S. Petridis, M. Leveque, and M. Pantic, "Audiovisual detection of laughter in human-machine interaction," in *Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*, 2013, pp. 129–134.

[3] J. Urbain, H. Çakmak, A. Charlier, M. Denti, T. Dutoit, and S. Dupont, "Arousal-driven synthesis of laughter," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 2, pp. 273–284, 2014.

[4] K. P. Truong and J. Trouvain, "On the acoustics of overlapping laughter in conversational speech," in *Proceedings of Interspeech*, 2012.

[5] 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.

[6] P. J. Glenn, "Initiating shared laughter in multi-party conversations," *Western Journal of Communication (includes Communication Reports)*, vol. 53, no. 2, pp. 127–149, 1989.

[7] E. Holt, "The last laugh: Shared laughter and topic termination," *Journal of Pragmatics*, vol. 42, no. 6, pp. 1513–1525, 2010.

[8] G. Jefferson, H. Sacks, and E. A. Schegloff, "Notes on laughter in the pursuit of intimacy." in *Talk and Social Organisation*, G. Button and J. Lee, Eds. Clevedon: Multilingual Matters, 1987, pp. 152–205.

[9] Y. Tian, C. Mazzocconi, and J. Ginzburg, "When do we laugh?" in *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 2016, pp. 360–369.

[10] K. P. Truong and J. Trouvain, "Laughter annotations in conversational speech corpora-possibilities and limitations for phonetic analysis," *Proceedings of the 4th International Worskhop on Corpora for Research on Emotion Sentiment and Social Signals*, pp. 20–24, 2012.

[11] M.-P. Jansen, D. K. Heylen, K. P. Truong, D. S. Nazareth, and G. Englebienne, "The MULAI Corpus: Multimodal Recordings of Spontaneous Laughter in Dyadic Interaction," in *Proceedings of the 2018 Laughter Workshop*, 2018, pp. 58–63.

[12] J. Trouvain, "Segmenting phonetic units in laughter," in *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 2003, pp. 2793–2796.

[13] 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.

[14] W. Ruch and P. Ekman, "The expressive pattern of laughter," in *Emotions, qualia, and consciousness*, A. W. Kaszniak, Ed. Tokyo: World Scientific, 2001, pp. 426–443.

[15] R. R. Provine, "Laughter punctuates speech: Linguistic, social and gender contexts of laughter," *Ethology*, vol. 95, no. 4, pp. 291–298, 1993.

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

[17] D. E. Mowrer, "A case study of perceptual and acoustic features of an infant's first laugh utterances." *Humor: International Journal of Humor Research*, vol. 7, no. 2, pp. 139–155, 1994.

[18] 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.

[19] E. Yanagisawa, M. J. Citardi, and J. Estill, "Videoendoscopic analysis of laryngeal function during laughter," *Annals of Otology, Rhinology & Laryngology*, vol. 105, no. 7, pp. 545–549, 1996.

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

[21] C. A. Bickley and S. Hunnicutt, "Acoustic analysis of laughter," in *Second International Conference on Spoken Language Processing (ICSLP))*, 1992, pp. 927–930.

[22] G. Habermann, *Physiologie und Phonetik des lauthaften Lachens: Untersuchungen zum Ausdruck im Stimmklang und zur Bildung der Stimmlaute*. Leipzig: Verlag J.A. Barth, 1955.

[23] W. L. Chafe, *The importance of not being earnest: The feeling behind laughter and humor*. John Benjamins Publishing, 2007, vol. 3.

[24] J. Urbain and T. Dutoit, "A phonetic analysis of natural laughter, for use in automatic laughter processing systems," in *Proceedings of International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2011, pp. 397–406.

[25] E. 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.

[26] J. Trouvain, "Phonetic aspects of speech-laughs," in *Oralité et Gestualité: Actes du colloque ORAGE, Aix-en-Provence. Paris: LHarmattan*, 2001, pp. 634–639.

[27] V. C. Tartter, "Happy talk: Perceptual and acoustic effects of smiling on speech," *Perception & psychophysics*, vol. 27, no. 1, pp. 24–27, 1980.

[28] P. Belin, S. Fillion-Bilodeau, and F. Gosselin, "The montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing," *Behavior research methods*, vol. 40, no. 2, pp. 531–539, 2008.

[29] J. Urbain, E. Bevacqua, T. Dutoit, A. Moinet, R. Niewiadomski, C. Pelachaud, B. Picart, J. Tilmanne, and J. Wagner, "The avlaughtercycle database." in *Proceedings of LREC*, 2010, pp. 2996–3001.

[30] A. Schweitzer and N. Lewandowski, "Convergence of articulation rate in spontaneous speech." in *Proceedings of Interspeech*, 2013, pp. 525–529.

[31] IPDS, "Video Task Scenario: Lindenstrasse - The Kiel Corpus of Spontaneous Speech," *Institut für Phonetik und Digitale Sprachsignal-verarbeitung Universität Kiel*, vol. 4, 2006, DVD.

[32] A. Janin, D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke *et al.*, "The icsi meeting corpus," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP03)*, 2003, pp. 364–367.

[33] Boersma, Paul and Weenink, David, "Praat: doing phonetics by computer," Version 6.0.45, retrieved from http://www.praat.org/, 2018.

[34] Y. Mathet, A. Widlöcher, and J.-P. Métivier, "The unified and holistic method gamma (γ) for inter-annotator agreement measure and alignment," *Computational Linguistics*, vol. 41, no. 3, pp. 437–479, 2015.

[35] J. Makhoul, F. Kubala, R. Schwartz, R. Weischedel *et al.*, "Performance measures for information extraction," in *Proceedings of DARPA broadcast news workshop*, 1999, pp. 249–252.