Laughter Annotations in Conversational Speech Corpora – Possibilities and Limitations for Phonetic Analysis

Khiet P. Truong[1] & Jürgen Trouvain[2]

[1] University of Twente, The Netherlands & [2] Saarland University, Germany

E-mail: [1]k.p.truong [at] utwente.nl & [2]trouvain [at] coli.uni-saarland.de

Abstract

Existing laughter annotations provided with several publicly available conversational speech corpora (both multiparty and dyadic conversations) were investigated and compared. We discuss the possibilities and limitations of these rather coarse and shallow laughter annotations. There are definition issues to be considered with respect to speech-laughs and the segmentation of laughs: what constitutes one laugh, and when does a laugh start and end? Despite these issues, some durational and voicing analyses can be performed. We found for all corpora considered that overlapping laughs are longer in duration and are generally more voiced than non-overlapping laughs. For a finer-grained acoustic analysis, we find that a manual re-labeling of the laughs adhering to a more standardized laughter annotations protocol would be optimal.

1. Introduction

Laughter is a non-verbal phonetic activity that usually occurs in conversational interaction with an interlocutor. In contrast to this we can state that most studies on the acoustics of laughter were *not* based on conversational settings but settings in which actors produce pre-selected laughter categories (Habermann 1955; Szameitat et al. 2009) or in which subjects watch funny video clips, either alone (Urbain et al. 2010) or with another person (Bachorowksi et al. 2001).

One important social feature of laughter *in conversations* is that it frequently is a joint action of two persons. Subsequently, laughs of interlocutors often overlap with laughs of the other. Since we are interested in studying phonetic and social aspects of laughter in conversation, of which overlapping laughter represents an important aspect, the first step to be taken is to look for laughter in conversational speech corpora.

Most studies focusing on laughter in conversations are based on rather restricted amounts of data either investigating actors in movies (Pompino-Marschall et al. 2007), focusing on interviews in mass media (O'Connol & Kowal 2004), eliciting experimental data, e.g. on male-female encounters (Grammer & Eibl-Eibesfeldt 1990) or on mother-child interaction (Nwokah et al. 1999), analysing a small corpus of acted dialogues recorded in a push-to-talk mode (Trouvain 2000), or performing qualitative studies of conversational analysis with only a few examples (e.g. Jefferson 1985).

Studies with larger data sets are often not publicly available, such as the natural dyadic conversations used in Vettin & Todt (2004). And sometimes, the conversations are recorded in a language unknown to the researchers that can be rather inconvenient, such as the recordings in Japanese used in Campbell (2007) where strangers have repeated telephone calls with each other.

There are a number of large conversational speech corpora publicly available containing laughter but usually, the developers of these databases did not record these with the aim to study laughter or other paralinguistic phenomena. Therefore, often only coarse and shallow annotation of laughter is available because only little attention was given for how to label laughter. Consequently, we cannot expect to find a standard labelling of laughter across multiple corpora.

In this study, we explore laughter annotations in different speech corpora and show how these can be used for phonetic analysis. The aims of this study are three-fold: 1) to compare and select different corpora suitable for phonetic laughter analysis, 2) to identify difficulties in laughter labelling, 3) to show how shallow laughter annotations can be used to explore durational and voicing aspects of overlapping laughter in conversation.

2. Conversational speech corpora

Prerequisites of conversational speech corpora ideally comprise: 1) separated channels for each speaker, 2) searchability of annotated laugh events in the transcription document, 3) time alignment of transcription and audio file with time stamps for the beginning and the end of the laugh event, 4) publicly available.

Not all corpora meet the mentioned criteria such as the separation of the recording channels. An example for a corpus with one channel for all speakers is the Santa Barbara Corpus of Spoken American English (SBC). Another example is the Buckeye corpus (Pitt et al. 2007)

for which only the data of the interviewed person is available but not the data of the interviewer as the interlocutor. The disadvantage of having only one channel is that during overlapping signals like cross-talk or overlapping laughs it is not clear which part of the signal stems from which speaker. However, for a fine-grained acoustic and temporal analysis this intertwining of both speakers can be very important as illustrated in Fig. 1 (taken from the Diapix Lucid corpus (Baker & Hazan 2011)).

Corpora can differ very much with respect to the annotation of laughter. For two larger Dutch conversational speech corpora, CGN (Oostdijk 2000) and IFADV (van Son et al. 2008) laughter was annotated with a label that also comprised other types of non-verbal vocalizations, e.g. coughs.

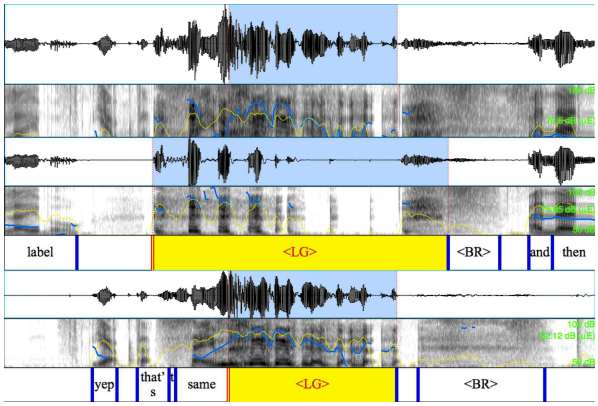

Figure 1: Example of an overlapping laugh (waveform and spectrogram. Top: mixed signal with masked information of speaker identity. Middle: signal of speaker A. Bottom: signal of speaker B.

Even if laughter was somehow annotated in the transcription files, the laughter annotations sometimes cannot readily be used for signal analysis because of missing ending times of laugh events (e.g. Lindenstraße corpus IPDS 2006).

In selecting suitable speech corpora, we restricted ourselves to the English language. However, the considered corpora do not represent an exhausted list because availability of data depends e.g. on financial aspects. We selected 4 corpora that met our prerequisites: the AMI meeting corpus (Carletta et al. 2007), the ICSI meeting corpus (Janin et al. 2003), the HCRC Map Task Corpus (Anderson et al. 1991), and the Diapix Lucid corpus (Baker & Hazan 2011), see also Table 1. The first two corpora contain multi-party meeting recordings and the latter two consist of task-based dyadic conversations. The main reason for considering 4 different corpora that we wanted to test how general our findings are.

3. Laughter annotations

We manually inspected some of the laughter annotations in the four mentioned corpora and encountered a number of problems in the annotations.

3.1 Definition problems

1. Are speech-laughs considered as a sub-type of laughs?
Sometimes speech-laughs are ignored and sometimes they are inconsistently labeled.

2. What is the definition of one laugh?
Sometimes the annotated laugh is in reality composed of two or more laughs, and vice versa, two annotated laughs are in reality one laugh. It also happens that the annotated laugh is only partially a laugh or sometimes it is unclear whether it was a laugh or not.

3. When does the laughter event start and end?
Sometimes the annotated laughs show incorrect time stamps for beginning and/or end.

3.2 Other problems

1. Are all audible laughs annotated?
Sometimes laughs in the audio file were missing in the annotation.

2. Are there any technical errors?
Sometimes there were annotated laughs with negative durations, or no timestamps at all.

Exploiting the information about laughter needs clear labelling criteria and a consistent application of these criteria. It seems to be that human annotation is better than annotations obtained by a machine (i.e. automatic forced alignment). In any of the corpora inspected we would consider a re-annotation as necessary to obtain more homogeneous laughter annotations across corpora that in turn will lead to more consistent and reliable research results.

4. Laughter analysis

Despite the listed drawbacks the existing corpora can be used as they are – but always with the restriction that we are not considering completely correct data.

4.1 Data used

The laughs used in the analysis were automatically extracted based on the transcriptions available from the four corpora under inspection. Speech-laughs were sometimes annotated in the corpora, e.g., ICSI meeting corpus (Janin et al. 2003) and Diapix Lucid corpus (Baker & Hazan 2011), but these were discarded in our analysis to make the data comparable to the HCRC Map Task corpus (Anderson et al. 1991) and the AMI meeting corpus (Carletta 2007). The transcribed laughs were most of the times treated as words with starting and ending times. However, a subpart of the annotated laughs was discarded due to missing time stamps, missing transcriptions or other technical issues. Since we are investigating overlapping laughs, only those laughs that have a start and end time were included in our analysis. Table 1 gives a short description of the corpora and laugh data used.

Table 1: Descriptive features of inspected corpora.

	no. of annotated laughs	no. of used laughs	no of speakers	no. of convers.	no. of speakers per convers.	mean duration of convers. (in mins)	task	visual contact	relationship between speakers
AMI	16477	8803	679	171	3-4	35.1 (13.5)	acted meeting	yes	mostly strangers
ICSI	12574	8388	494	75	3-11	55.0 (15.9)	real meeting	yes	colleagues
HCRC	1002	966	250	125	2	6.8 (3.1)	giving route on a map	yes/no	friends + strangers
DiaPix	582	575	114	57	2	7.7 (2.3)	spot-the-difference	no	friends

4.2 Frequency of occurrence

Fig. 2 reveals that overlapping laughs represent a substantial part of all laughs in all corpora ranging from 35% to 63% of all annotated laughs. Only the ICSI corpus shows more overlapping than non-overlapping laughs. This can be easily explained by the fact that in the ICSI corpus there are many more persons present and thus increasing the probability that two speakers will overlap with their laughs. Additionally a 'contagious effect' could be at work for laughter as was already shown by Laskowski & Burger (2007b).

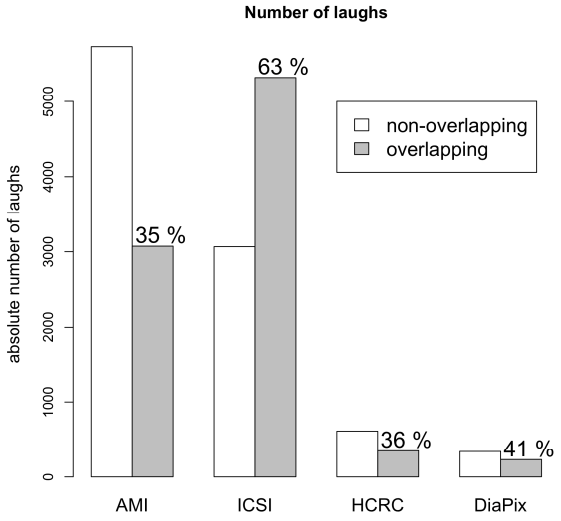

Figure 2: Frequency of occurrence of non-overlapping and overlapping laughs for each corpus. Percentages indicate the relative number of overlapping laughs.

4.3 Duration

The descriptive statistics illustrated in Table 2 and Fig.3 clearly show that overlapping laughs are longer than non-overlapping laughs. T-tests reveal that for each corpus these durational differences reach statistical significance at $p<0.01$. Interestingly, the multi-party meetings show higher durations in average, at least for overlapping laughs. The ICSI corpus differs again compared to the others by showing longer mean durations for overlapping as well as for non-overlapping laughs.

Table 2: Mean duration and standard deviation in seconds of all laughs (left), non-overlapping laughs (NO) and overlapping laughs (OL) pooled over the inspected corpora.

	all		NO		OL	
	mean	sd	mean	sd	mean	sd
AMI	1.042	1.184	0.775	0.842	1.541	1.521
ICSI	1.661	1.298	1.195	0.753	1.929	1.460
HCRC	0.838	0.652	0.715	0.524	1.052	0.784
DiaPix	0.899	0.689	0.755	0.495	1.107	0.860

4.4 Voiced vs. unvoiced laughter

Laughter is sometimes classified in voiced vs. unvoiced forms (e.g. Grammer & Eibl-Eibesfeldt 1990, or Bachorowski et al. 2001). For our analysis we define those laughs as unvoiced that show no voiced frame at all (as obtained from a pitch analysis with a window length of 40 ms and time step of 20 ms). The rest of the laughs are defined as "voiced" even if the number of voiced frames can be relatively low (in contrast to Laskowski & Burger (2007a) who did a manual classification of voicedness leading to a higher number of unvoiced laughs for the ICSI corpus).

In Fig. 4, we can observe a positive correlation between the level of voicing and duration (similar to Laskowski & Burger 2007a). There are hardly any unvoiced laughs longer than 1.6 sec and most unvoiced laughs are shorter than 800 ms.

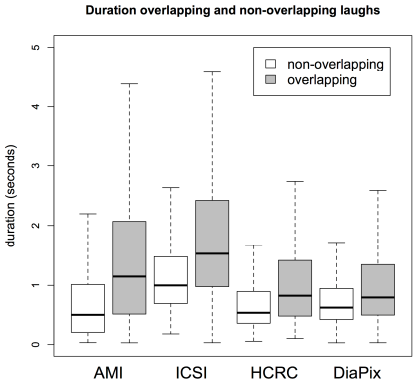

Figure 3: Boxplots of the duration in seconds of non-overlapping and overlapping laughs in the four inspected corpora. Outliers were computed but not shown for illustrative reasons. Whiskers indicate 1.5*inter-quartile range of the data.

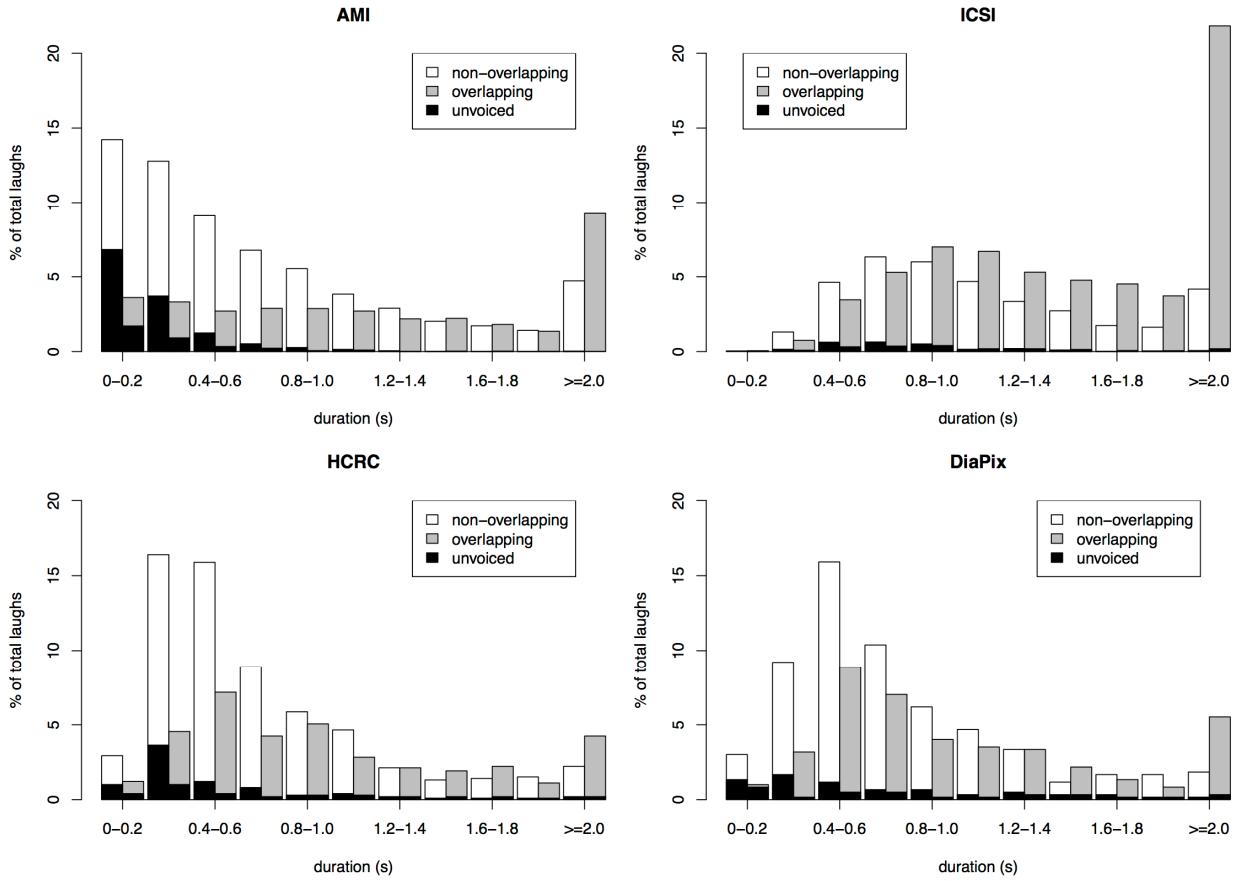

Figure 4: Histograms (for each corpus) of non-overlapping vs. overlapping laughs distinguishing unvoiced and "voiced" laughs in bins of 200 ms.

Fig. 4 also shows for all four corpora that the longer the laugh the higher the probability that the interlocutor joins in, resulting in an overlapping laugh. This effect is clearest for the ICSI meeting corpus where up to 11 conversational partners were present. For this corpus there are also the fewest unvoiced laughs counted in relation to the total number of laughs.

5. Concluding Remarks

In comparing conversational speech corpora we have found differences in the duration and numbers of overlapping laughs between corpora, particularly between multi-party conversations and dialogues. In general we could observe the tendency that overlapping laughs are more likely to be longer than non-overlapping ones; we hypothesize that this has to do with the social function of laughing together. In addition we saw that among the shorter laughs there was a relatively high proportion of unvoiced laughs.

The "noise" of the laughter annotations could have influenced results but the observations are made in multiple corpora giving strong evidence for our conclusions. However, we still consider a manual re-labelling of the laughter annotations as optimal for further more fine-grained acoustic analyses. Future research should include looking at acoustic characteristics of various kinds of laughter (overlapping vs. non-overlapping, voiced vs. unvoiced, speech-laughs), in addition to duration.

Acknowledgements

This research has been supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet) and the UT Aspasia Fund.

6. References

Anderson, A.H., Bader, M., Gurman Bard, E., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S., Weintert, R. (1991). The HCRC Map Task Corpus. *Language and Speech* 34(4), pp. 351-366.

Bachorowski, J.-A., Smoski, M.J., Owren, M.J. (2001). The acoustic features of human laughter. *Journal of the Acoustical Society of America* 111(3), pp. 1582-1597.

Baker, R, Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43(3), pp. 761-770.

Campbell, N. (2007). Whom we laugh affects how we laugh. *Proc. Workshop on "The Phonetics of Laughter"*, Saarbrücken, pp. 61-65.

Carletta, J.C. (2007). Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. *Language Resources and*

Evaluation 41(2), pp. 181-190.

Grammer, K., Eibl-Eibesfeldt, I. (1990). The ritualisation of laughter. In: Koch, W.A. (Hrsg.) *Natürlichkeit der Sprache und der Kultur: acta colloquii*. (Bochumer Beiträge zur Semiotik; 18) Bochum: Brockmeyer, pp. 192-214.

Habermann, G. (1955). *Physiologie und Phonetik des lauthaften Lachens*. Leipzig: J. A. Barth.

IPDS (2006). *Video Task Scenario: Lindenstraße* – The Kiel Corpus of Spontaneous Speech, Volume 4, DVD, Institut für Phonetik und Digitale Sprachsignalverarbeitung Universität Kiel.

Janin, A., Baron, D., Edwards, D., Ellis, D., Gelbart, D., Morgan, N. (2003). The ICSI meeting corpus. *Proceedings of ICASSP*, pp. 364-367.

Jefferson, G. (1985). An exercise in the transcription and analysis of laughter. In T. Van Dijk (Ed.) *Handbook of discourse analysis*, Vol. 3: *Discourse and dialogue* (pp.25-34).London, UK: Academ. Pr.

Laskowski, K., Burger, S. (2007a). On the correlation between perceptual and contextual aspects of laughter in meetings. *Proc. Workshop on "The Phonetics of Laughter"*, Saarbrücken, pp. 55-60.

Laskowski, K., Burger, S. (2007b). Analysis of the occurrence of laughter in meetings. Proceedings of Interspeech, Antwerp, pp. 1258-1261.

Nwokah, E.E., Hsu, H.-C., Davies, P. & Fogel, A. (1999). The integration of laughter and speech in vocal communication: a dynamic systems perspective. *Journal of Speech, Language and Hearing Research* 42, pp. 880-894.

O'Connell, D. C., Kowal, S. (2004). Hillary Clinton's laughter in media interviews. *Pragmatics* 14(4), pp. 463-478.

Oostdijk, N. (2000). The Spoken Dutch Corpus. Overview and first evaluation. *Proc. LREC*, pp. 887-894.

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Pompino-Marschall, B., Kowal, S., O'Connell, D. (2007). Some Phonetic Notes on Emotion: Laughter, Interjections, and Weeping. *Proc. Workshop on "The Phonetics of Laughter"*, Saarbrücken, pp. 41-42.

Santa Barbara Corpus. http://www.linguistics.ucsb.edu/research/sbcorpus.html retrieved 15 Feb 2012

Szameitat, D.P., Alter. K., Szameitat. A.J., Dietrich. S., Wildgruber. D., Sterr. A., Darwin C.J. (2009). Acoustic profiles of distinct emotional expression in laughter, *Journal of the Acoustical Society of America* 126, pp. 354-366.

Trouvain, J. (2001). Phonetic aspects of 'speech-laughs'. Proceedings of the *2nd Conference on Orality & Gestuality (ORAGE)*, Aix-en-Provence, pp. 634-639.

Urbain, J., Bevacqua, E., Dutoit, T., Moinet, A., Niewiadomski, R., Pelachaud, C., Picart, B., Tilmanne J., Wagner, J. (2010). The AVLaughter-Cycle Database. *Proc. LREC*, Malta, pp. 2996-3001.

Van Son, R., Wesseling, W., Sanders, E., van den Heuvel, H. (2008). The IFADV corpus: A free dialog video corpus. *Proc. LREC*, pp. 501-508.

Vettin, J., Todt, D. (2004). Laughter in conversation: features of occurrence and acoustic structure. *Journal of Nonverbal Behaviour* 28(2), pp. 93-115.