On the acoustics of overlapping laughter in conversational speech

Khiet P. Truong[1]*, Jürgen Trouvain*[2]

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

Abstract

The social nature of laughter invites people to laugh together. This joint vocal action often results in overlapping laughter. In this paper, we show that the acoustics of overlapping laughs are different from non-overlapping laughs. We found that overlapping laughs are stronger prosodically marked than non-overlapping ones, in terms of higher values for duration, mean F_0, mean and maximum intensity, and the amount of voicing. This effect is intensified by the number of people joining in the laughter event, which suggests that entrainment is at work. We also found that group size affects the number of overlapping laughs which illustrates the contagious nature of laughter. Finally, people appear to join laughter simultaneously at a delay of approximately 500 ms; a delay that must be considered when developing spoken dialogue systems that are able to respond to users' laughs.

Index Terms: laughter, conversation, overlap, entrainment

1. Introduction

An important social feature of laughter in conversation is that it frequently is a joint action of two or more persons. This joint action often results in overlapping laughter. Although there is a substantial amount of speech overlap present in conversational speech [1, 2], laughter gives rise to more overlap than speech does [3]. This is partly due to the social nature of laughter which is often described in the literature as having the function of 'social bonding' or 'creating affiliation' [4]. Presumably, the decision of one to join the laughing event or not can serve various social functions in addition to transporting affective information. In this paper we focus on the acoustic realization of laughter as a joint vocal activity in conversations.

Overlapping laughter has been qualitatively described with several examples in studies of conversational analysis, e.g. [4]. However, not much phonetic work on overlapping laughter has been reported in previous literature, in contrast to overlapping speech. In this first exploration we address the main research question whether laughing (partially) simultaneously has an effect on the laughs' *acoustic* realizations. We suspect that a laugher acknowledges a joining-in laugher by acoustic means (among others). More specifically, we hypothesize that overlapping laughs are stronger prosodically marked. A form of entrainment (also known in the literature as alignment or adaptation) could be at work, both laughers intensifying each others' laughs, as soon as one hears that the other one is laughing along.

In addition, we investigate whether the *group size* of a conversation has an effect on the acoustics and occurrence of overlapping laughter. We suspect that in multiparty conversations, there will be more overlapping laughter than in dyadic conversations because the 'infection risk' of laughing is higher in multiparty conversations due to the larger number of participants, and more social conventions will come into play in small group behaviors. Thus, we ask how the participants' laughing behaviors depend on the size of the conversational group.

Finally, we address *interactive* aspects of laughter that will give insights into how laughter interaction can be modelled in spoken dialogue systems. Analogously to analyses of speech overlap and vocal responses that address the matter of suitable response times for spoken dialogue systems to initiate a vocal response, e.g., [1], we will look at the range of reponse times with respect to laughter. Within what time range should a spoken dialogue system react upon a user's laugh, and should this response time have an effect on laughter acoustics?

We first describe the corpora used in our study in Section 2. In Section 3, we go into more detail describing our research questions and how we address these. The results are presented in Section 4 and our conclusions are discussed in Section 5.

2. Data

For our analysis, we used 4 different corpora: the ICSI Meeting Recorder corpus (real-life meetings of researchers) [5], the AMI corpus (role-playing and real-life meetings) [6], the HCRC Map Task corpus (task: finding the route on a map) [7], and the Diapix Lucid corpus (task: spot-the-difference in a picture) [8]. The first two corpora contain multiparty meeting recordings and the latter two consist of task-based dyadic conversations. The main reason for considering 4 different corpora is to increase the level of generalization. In addition, the multiparty meeting recordings allow us to address the social aspect of laughter given that these meetings have various sizes in terms of number of participants. For more detailed descriptions of the corpora, readers are referred to [5, 6, 7, 8]. We used the laughter annotations provided with the corpora and did not perform any annotations or re-labeling ourselves. Also, meta-information such as meeting formality or participants' cultural backgrounds were not considered in the current study. Only laughs from close-talk microphone recordings with valid timestamps indicating the start and ending of laughs were included. Speech-laughs were excluded from the current analysis. Only conversations (whole AMI and ICSI meetings are regarded as 'conversations') in which laughter was present were considered. This amounts to 96% of the time of all corpora being considered. Furthermore, some conversations in the corpora were excluded due to either missing transcripts or audio problems. For a discussion of other drawbacks of the laughter annotations see [9].

3. Analysis

We explain what aspects of overlapping laughter we are addressing in the current study. First, a description of how we define and count overlapping laughs is given.

Table 1: *The four inspected corpora (ICSI also shown as sub-corpora by group size). In column 5 the absolute number of laughs is followed by the percentage of* OL *related to* OL + NOL*. Average percentages of overlapping laugh and overlapping speech time (cross-talk) relative to the total laughing time (column 6) and the total speaking time (column 7), and relative to the total meeting time (column 8 and 9) are shown. Standard deviations are shown in italics. These numbers include speakers with no close-talk microphone.*

Corpus	Group size	# Conversations	Dur. (hrs)	#Laughs all	#Laughs %OL	$\frac{\text{duration OL}}{\text{duration (OL+NOL)}}$ (%)	$\frac{\text{duration cross-talk}}{\text{duration all talk}}$ (%)	$\frac{\text{duration OL}}{\text{total duration}}$ (%)	$\frac{\text{duration cross-talk}}{\text{total duration}}$ (%)
HCRC	2 spks	96	11.4	966	36.4	9.72 *13.22*	4.91 *2.25*	0.23 *0.37*	3.21 *1.80*
Diapix	2	52	6.8	575	40.5	12.64 *13.54*	7.95 *4.87*	0.27 *0.36*	6.03 *3.92*
AMI	4	160	91.3	8531	35.2	12.37 *8.29*	13.24 *6.27*	0.42 *0.52*	10.73 *5.58*
ICSI	3	1	0.6	43	41.9	16.79	11.13	0.43	8.58
ICSI	4	3	2.7	172	45.3	18.31 *6.31*	17.55 *2.57*	0.41 *0.26*	13.96 *2.46*
ICSI	5	14	12.9	1099	53.2	24.36 *10.26*	13.65 *5.39*	0.72 *0.65*	11.04 *4.52*
ICSI	6	21	19.8	1645	55.9	27.03 *11.37*	17.58 *5.24*	0.75 *0.47*	15.05 *4.66*
ICSI	7	16	16.0	2114	67.8	33.17 *13.66*	19.54 *7.10*	1.46 *1.50*	17.30 *6.93*
ICSI	8	11	12.0	1314	67.8	36.20 *12.01*	17.13 *8.69*	1.34 *0.94*	14.95 *8.19*
ICSI	9	5	4.6	1432	74.4	41.05 *2.30*	29.77 *9.83*	3.39 *1.08*	27.77 *10.2*
ICSI	10	2	1.5	244	61.9	33.89 *9.11*	26.76 *9.06*	1.99 *0.57*	21.79 *9.28*
ICSI	11	1	1.1	234	67.1	30.13	21.49	1.60	19.32
ICSI	3–11	74	71.4	8297	63.8	29.90 *12.17*	18.23 *7.53*	1.19 *1.13*	15.73 *7.25*

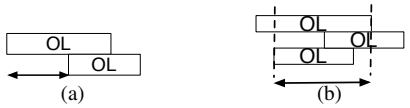

Figure 1: *Measurements of (a) 'delay time' (for two speakers) and (b) overlapping laughing time (for two or more speakers).*

3.1. Defining and counting overlapping laughs

We consider annotated laughs as non-overlapping (NOL) or overlapping (OL). An overlapping laugh is defined as an annotated laugh token of a speaker that occurs partially (or fully) at the same time of another annotated laugh token of another speaker – both these tokens are then marked and counted as 2 OLs, see Fig. 1 The overlap assessment is based on the existing laugh annotations as provided with the corpora under study. No threshold of overlap duration is applied; each laugh token that shows the slightest overlap is considered as OL (depending on the corpus, the resolution is in centiseconds or milliseconds).

3.2. Acoustic measurements

We hypothesize that people realize their laughs in a different way when their laughs are being overlapped by the interlocutors' laughs due to a form of entrainment caused by signalling affiliation or by social pressure. To be more precise, we hypothesize that overlapping laughs are more prosodically marked in terms of their duration, intensity, F_0 and the number of voiced frames. We have already found evidence for duration and the number of voiced frames [9] but not yet for intensity and F_0. In addition, the number of persons overlapping each other's laugh might influence the strength of prosodic markedness.

For each laugh token, we measure duration, number of voiced frames (time step of 0.02 s), mean F_0 (logarithmic, time step 0.01 s, the mean was only taken over the voiced frames), mean intensity (in dB, time step 0.01 s), and the maximum intensity (in dB, time step 0.01 s). In order to make the acoustic measurements comparable across speakers, we transformed all values to z-scores (for each speaker separately) by using $z = (x - \mu)/\sigma$ where μ and σ are derived from the individual speakers' speaking time. All acoustic measurements were performed using Praat [10].

3.3. Group size effects

It is often said that laughter is 'contagious', 'infecting' other participants to laugh along. We would like to see how this 'infection', which presumably leads to more laugh overlap, develops in multiparty meetings with a varying number of participants, ranging from 3–11. Or to put it differently, whether OL is more a phenomenon of 1-speaker-to-just-1-listener or 1-speaker-to-entire-audience interaction. We address these aspects in two ways. Firstly, by comparing the amount of overlapping laugh time (relative to the total laughing and meeting time) and the amount of overlapping speech time (relative to the total speaking and meeting time) for various group sizes of multiparty meetings. Per conversation, the overlapping laugh time is measured as the duration in which at least 2 people are simultaneously laughing (see Fig. 1b). And secondly, by looking at the number of persons involved in overlapping laughs for different group sizes of meetings. We also investigated whether group size would affect the acoustics of OL.

3.4. Interactive aspects

The decision of an interlocutor to respond to a laugh from another person requires some planning time. This so-called 'delay time' of the interlocutor joining in the other one's laughter gives us information about how fast humans respond to laughs and how well humans can anticipate laughs. Furthermore, the average 'delay time' will give us insights into what range of response times are suitable for spoken dialogue systems that aim to detect and react to laughter. Since these systems mostly target dyadic conversations, we restrict ourselves in the analysis of the delay time to the HCRC and Diapix Lucid corpus, both of which contain dyadic conversations. For each pair of overlapping laughs, we calculate the difference in time, the 'delay time', between the first laugher and the laugher who joins in, see Fig. 1a. When one laugh token of the first speaker is overlapped with multiple laugh tokens of the other speaker the 'delay time' of the first overlapping token is used.

Table 2: *Averages of acoustic measurements for each corpus considered, divided into overlapping* OL *and non-overlapping laughs* NOL*. When applicable, standard deviations are shown in italics. These numbers consider close-talk microphone recordings only.*

	AMI		ICSI		HCRC		Diapix	
# Conversations used/hrs. used	165/94.5		67.9/74		11.4/96		6.8/52	
	OL	NOL	OL	NOL	OL	NOL	OL	NOL
Number of laughs	3004	5583	5281	2987	352	614	233	342
Duration (s)	1.555	0.782	1.931	1.200	1.052	0.715	1.107	0.754
	1.531	*0.846*	*1.464*	*0.758*	*0.784*	*0.524*	*0.860*	*0.496*
Mean log F_0 (z-score)	0.576	0.260	0.717	0.05	1.646	0.776	1.228	0.919
	1.510	*1.666*	*1.482*	*1.497*	*1.995*	*2.350*	*1.056*	*1.355*
Mean intensity (z-score)	-0.355	-0.492	-0.1637	-0.309	0.071	-0.003	-0.212	-0.169
	0.590	*0.643*	*0.531*	*0.568*	*0.832*	*0.806*	*0.853*	*0.896*
Maximum intensity (z-score)	1.085	0.554	1.609	1.403	1.804	1.442	1.464	1.364
	0.968	*1.018*	*0.730*	*0.807*	*0.849*	*0.859*	*0.702*	*0.748*
Number of voiced frames	23.8	12.7	23.8	12.4	11.7	7.2	16.1	9.4
	28.5	*18.0*	*27.4*	*14.3*	*12.5*	*7.7*	*14.9*	*9.4*
Completely unvoiced (%)	9.5	19.3	2.8	6.4	6.8	11.2	4.7	9.6

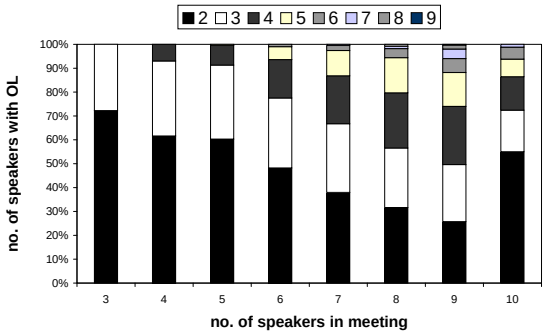

Figure 2: *For the ICSI corpus: number of speakers present in the meeting with a close-talk microphone (x-axis) and stacked bars with number of speakers involved in* OL *(y-axis).*

4. Results

4.1. Occurrence of overlapping laugh

We first looked at the occurrence of overlapping laugh in comparison to overlapping speech. In Table 1, we can observe that it is more likely that laughter overlaps with laughter than speech overlaps with speech (in line with [3]): the proportion of overlapping laughter relative to the total laughing time (column 6) is higher than the proportion of overlapping speech relative to the total speaking time (column 7). This observation is valid for *all* corpora inspected, irrespectively of the number of participants present.

Table 1 (columns 5 and 6) also shows that there is an effect of group size on the amount of overlapping laugh in the ICSI corpus: the larger the group size, the more overlapping laughs. This effect can be partly attributed to the mere fact that there are more participants present which increases the chance of overlap, and partly to laughter being 'infectuous' or 'contagious'. As can be observed in Fig. 2, this 'infection' effect affects the proportion of 1-speaker-to-1-listener laugh interaction that decreases with the number of participants (but it remains a relatively high proportion), and we can see that the 'infection' rarely spreads out to all participants present in the group.

4.2. Laughter acoustics

4.2.1. General findings

For each laugh token, acoustic measurements were extracted to see whether OL are more prosodically marked than NOL. Table 2 presents these acoustic measurements (averaged over all tokens), grouped by OL and NOL. We observe that OLs are indeed stronger prosodically marked than NOLs: over all corpora, on average, OLs have a longer duration, a higher number of voiced frames, a higher mean intensity, a higher maximum intensity, and a higher mean F_0 than NOLs. T-tests (separately for each measurement and each corpus) revealed that these acoustic differences between OL and NOL are statistically significant at a level of $p < 0.005$, except for mean intensity in the HCRC corpus and the mean and maximum intensity in the Diapix corpus.

Laughs, in general, have a higher mean F_0 and maximum intensity than average speech. The negative z-scores for mean intensity suggest that laughter is softer than average speech which is contrary to belief, but this can be explained by our measuring method that measures intensity over all frames, including the unvoiced ones, of the laugh token. Depending on the number of unvoiced frames, the mean intensity will decrease – that is the reason we also measured maximum intensity which is independent from the number of voiced frames.

4.2.2. Group size effects

We also looked at whether the number of persons laughing (partially) at the same time affected the acoustic realization of laughs – this was possible for the multiparty ICSI and AMI corpora. In Table 3, we can observe that the more people are laughing simultaneously (partially), the more prosodically marked the laughs are (increased duration, pitch, intensity, and the number of voiced frames) – the group size thus plays a role in the acoustic realization of overlapping laughs.

4.2.3. Interactive aspects

We calculated the 'delay time' for each pair of overlapping laughs in the HCRC and Diapix Lucid corpus and found mean average 'delay times' of 0.489 s (0.456 standard deviation, based on 181 pairs) and 0.532 s (0.573 standard deviation, based on 124 pairs) respectively. This means that spoken dialogue systems have a 'delay time' of approximately 500 ms

Table 3: *Averaged acoustic measurements for laugh tokens shown by number of people simultaneously laughing (standard deviation in brackets). '1' person laughing means* NOL*, '2' and more persons laughing means* OL

	# Persons laughing – ICSI							# Persons laughing – AMI			
	1	2	3	4	5	6	>= 7	1	2	3	4
Duration (s)	1.197	1.520	1.820	2.278	2.758	2.917	3.038	0.782	1.374	2.114	2.994
	0.758	*0.998*	*1.244*	*1.545*	*2.211*	*2.179*	*1.872*	*0.846*	*1.352*	*1.894*	*1.851*
Mean log F_0 (z)	0.054	0.427	0.712	0.907	1.100	1.244	1.520	0.260	0.534	0.693	0.836
	1.499	*1.485*	*1.479*	*1.443*	*1.332*	*1.456*	*1.404*	*1.667*	*1.517*	*1.503*	*1.242*
Mean intensity (z)	-0.309	-0.209	-0.152	-0.129	-0.111	-0.131	-0.040	-0.492	-0.361	-0.323	-0.438
	0.568	*0.549*	*0.515*	*0.516*	*0.508*	*0.536*	*0.544*	*0.643*	*0.597*	*0.568*	*0.528*
Max intensity (z)	1.403	1.545	1.593	1.650	1.714	1.805	1.967	0.554	1.018	1.310	1.420
	0.807	*0.768*	*0.704*	*0.694*	*0.664*	*0.769*	*0.683*	*1.018*	*0.976*	*0.897*	*0.965*
# Voiced frames	12.4	17.0	23.0	29.3	38.1	35.5	34.5	12.7	21.6	31.0	39.1
	14.3	*18.5*	*24.6*	*30.7*	*41.7*	*37.1*	*28.9*	*18.0*	*25.7*	*34.3*	*48.3*

available upon detection of a laugh before deciding how to respond to a laugh. A histogram of all the 'delay times' pooled together is shown in Fig. 3.

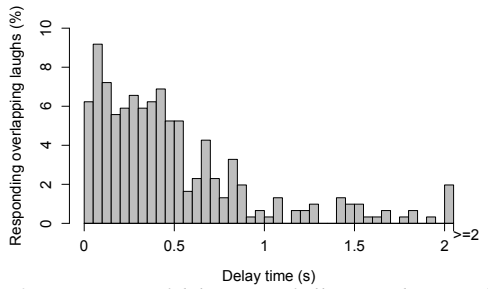

Figure 3: *Histogram of delay time of all responding overlapping laughs in HCRC and Diapix (binsize is* 0.05 *s,* $N = 305$*).*

We also looked at whether the 'delay time' has an effect on the phonetic properties of a laugh, but did not find clear effects. However, this aspect should be investigated in more detail in future research.

5. Discussion and conclusions

Our analyses show that laughter is a highly social and joint vocal action whose acoustic realization depends on several measurable factors. We have seen that the proportion of overlapping laugh increases with the meetings' group sizes. This increase can be partly attributed to chance probability but it is boosted by the 'infectuous' nature of laughing which was also illustrated in Fig. 2. Future work will address how to quantify this 'infectuous' effect in order to show that this effect is indeed induced by 'contagion' rather than the increased chance of overlap (at random) caused by a larger number of participants.

The joint and 'infectuous' nature of laughter also has an effect on laughter's phonetic properties. Overlapping laughs are consistently more prosodically marked than non-overlapping ones, showing higher values of duration, mean F_0 and mean and maximum intensity, and this effect becomes stronger when more persons join in the simultaneous laughing event. This suggests that persons joining in laughing re-enforce each other's laugh realizations which could be seen as a form of entrainment. However, more research is needed to show that these higher values, especially with respect to intensity, are a result of entrainment rather than Lombard effects [11].

Finally, we have seen that people join in overlapping laughter events with a delay of approximately 500 ms. This delay must be taken into account when developing laughter-reacting spoken dialogue systems that need to decide when and how to initiate an overlapping laugh response. However, it is not clear yet how variations in 'delay times' and variations in phonetic properties of laughter affect different interpretations. This is one of the challenges that should be investigated in more detail in future research.

6. Acknowledgements

We would like to thank Eva Lasarcyk and three anonymous reviewers for their helpful comments. This work was partly supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet) and the UT Aspasia Fund.

7. References

[1] M. Heldner and J. Edlund, "Pauses, gaps and overlaps in conversations," *Journal of Phonetics*, vol. 38, pp. 555–568, 2010.

[2] E. Shriberg, A. Stolcke, and D. Baron, "Observations on overlap: findings and implications for automatic processing of multi-party conversation," in *Proc. Eurospeech*, 2001, pp. 1359–1362.

[3] K. Laskowski and S. Burger, "Analysis of the occurrence of laughter in meetings," in *Proc. Interspeech*, 2007, pp. 1258–1261.

[4] G. Jefferson, H. Sacks, and E. 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.

[5] A. Janin, D. Baron, D. Edwards, D. Ellis, D. Gelbart, and N. Morgan, "The ICSI meeting corpus," in *Proc. ICASSP*, 2003, pp. 364–367.

[6] J. Carletta, "Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus," *Language Resources and Evaluation*, vol. 41, pp. 181–190, 2007.

[7] A. H. Anderson, M. Bader, E. Gurman Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weintert, "The HCRC Map Task Corpus," *Language and Speech*, vol. 34, pp. 351–366, 1991.

[8] R. Baker and V. Hazan, "DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs," *Behavior Research Methods*, vol. 43, pp. 761–770, 2011.

[9] K. P. Truong and J. Trouvain, "Laughter annotations in conversational speech corpora - possibilities and limitations for phonetic analysis," in *Proc. of Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, 2012, pp. 20–24.

[10] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[11] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," *J Speech Hear Res*, vol. 14, no. 4, pp. 677–709, 1971.