



Investigating prosodic relations between initiating and responding laughs

Khiet P. Truong¹, Jürgen Trouvain²

¹Human Media Interaction, University of Twente, Enschede, The Netherlands

²Phonetics, Saarland University, Saarbrücken, Germany

k.p.truong@utwente.nl, trouvain@coli.uni-saarland.de

Abstract

In dialogue, it is not uncommon for people to laugh together. This joint laughter often results in overlapping laughter, consisting of an initiating laugh (the first one), and a responding laugh (the second one). In previous studies, we found that overlapping laughs are acoustically different from non-overlapping ones. So far, we have considered overlapping laughs as one category. Consequently, it is unknown whether there are also acoustic differences between initiating laughs and responding laughs. In this paper, we make a distinction between initiating, responding, and non-overlapping laughs and compare their acoustic characteristics. In particular, we will investigate the prosodic relations between initiating and responding laughs. Do these relations point to a form of accommodation and mimicry? To what extent are initiating and responding laughs paired to each other? The analyses were performed on two speech corpora containing spontaneous conversations between two speakers. Results show indications that initiating and responding laughs share several similar acoustic features that point towards accommodation and mimicry mechanisms.

Index Terms: laughter, overlap, dialogue, joint action, mimicry

1. Introduction

“Not all laughs are alike” is the title of a paper by Bachorowski and Owren [1] in which they uncover the fact that there are substantial differences between laughs, for instance between voiced (‘song-like’) and unvoiced forms of laughter (including ‘grunt-like’ and ‘snort-like’ forms). These different forms of laughs are being evaluated by humans in different ways: voiced laughs were perceived as more positive than unvoiced laughs [1]. In their follow-up study [2] they also investigated laughs with respect to duration. Substantial differences were found for laughs produced by males compared to those produced by females. The results of these studies point towards the existence of a complex relation between the phonetic form of laughter and its evaluation by humans.

Although Bachorowski et al. [2] elicited laughs in social interaction (friends watching funny video clips together), the laughing there was not part of a conversation. Conversational speech does not exclusively consist of articulated words but also non-verbal vocalisations such as laughter. The analysis of various conversational corpora reveals that laughter is one of the most frequent non-verbal vocalisations in dialogue [3]. More importantly, laughter in spontaneous dialogue can also have a social function: the act of laughing together often creates a bond between interactants. As a result of laughing together, laughter of one speaker often overlaps with laughter of the other speaker.

Interestingly, we have shown in our previous studies [4, 5] that laughs that overlap (OL) show acoustic differences to non-overlapping laughs (NOL). OLs are in general longer, more

intense, higher pitched and have a higher portion of voiced frames than NOL. Although the differences between OL and NOL are known it is yet unclear whether within the OLs the different laugh productions of the overlapping speakers acoustically differ or not. So far, OLs have been considered as *one* category. However, the laugh which comes first in an OL (*initiating* laugh) might be acoustically different to the laugh of the speaker who joins in (*responding* laugh).

There are several reasons to make a distinction between initiating and responding laughs and to consider their acoustics separately. First, initiating laughs could be produced to invite people (see [6, 7]) to laugh as a way to bond [8]. If there is no difference between initiating and responding laughs (except the starting time in the interaction) we can assume that there is a very high degree of phonetic adaptation (imitation/mimicry) supporting this bond. Second, in case we do find an acoustic difference, in particular in increased feature values in the responding laugh, this could indicate a mechanism similar to the Lombard effect [9] where people tend to increase their vocal effort in order to be audible in the presence of a ‘noise’ source, which, for example, could be another speaker talking, or in this case, another speaker laughing.

In this paper, we present a study into the acoustic differences (or similarities) between initiating and responding laughs. To what extent are initiating and responding laughs paired to each other, and how are they related to each other in terms of several acoustic features? Pairs of initiating and responding laughs, and non-overlapping laughs were automatically extracted from speech corpora containing conversational speech between two interactants. Various acoustic features such as fundamental frequency, intensity, amount of voicedness and duration were extracted automatically for all laughs. First, comparisons among the three ‘types’ of laughs were made. Subsequently, we looked more into detail into the prosodic relations between the initiating and responding laughs.

The structure of the paper is as follows. We describe the speech material used in Section 2 and present the analyses and results in Section 3. Finally, we discuss our conclusions in Section 4.

2. Data

We selected two spontaneous speech corpora that have laughter annotations available and that contain dyadic interactions. We briefly describe the two corpora with (British) English speakers; for more details, readers are referred to the references.

The Diapix Lucid corpus [10] contains audio recordings from spontaneous task-based dialogues. The task concerns a ‘spot-the-difference’ task in which each participant is given a different version of a cartoon picture. The goal for the participants is to locate the twelve differences between the two pic-

	Diapix corpus						HCRC MapTask corpus					
	A	B	C	A·B	B·C	A·C	A	B	C	A·B	B·C	A·C
intens	-0.20760	-0.20299	-0.22327				0.02397	0.07879	-0.05831			
max intens	1.82166	1.60877	1.60325			*	1.91735	1.84759	1.46519		*	*
pitch	1.05389	1.07910	0.71089		*		0.91137	0.77979	0.00000		*	*
voiced	-0.23879	-0.00724	-0.27859				-0.21589	-0.09735	-0.30456			
dur	0.17313	-0.315997	-0.38712	*		*	0.29800	-0.22998	-0.44639	*		*

Table 1: Medians of initiating, responding, and non-overlapping (NOL) laughs. * indicates significance of post-hoc Mann Whitney U tests at $p < .003$

tures. The participants are seated in separate rooms and communicate via headsets with each other.

The HCRC Map Task corpus [11] contains audio recordings from spontaneous task-based dialogues. Each participant, who has the role either of ‘instruction giver’ or ‘instruction follower’, is given a different version of a map. The map of the instruction giver contains a route that should be reproduced on the instruction follower’s map with as few deviations as possible.

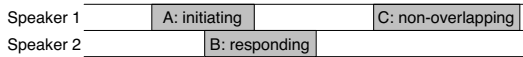


Figure 1: Illustrations of initiating (A), responding (B) and non-overlapping (C) laughs.

From the laughter annotations that were included with the corpora, we automatically derived three types of laughter: the initiating (A), responding (B) and non-overlapping laugh (C), see Fig. 1. The initiating and responding laughs are always overlapping each other and constitute a sub-group of the overlapping laughs that we have previously considered in [4]. The number of laughs found in each corpus are shown in Table 2. There were a few responding laughs that overlapped with the same initiating laughs, hence the different numbers for A and B.

	A	B	C	#sp	#conv
Diapix	111	119	345	37	52
HCRC	169	177	620	62	96

Table 2: Number of initiating (A), responding (B), non-overlapping (C) laughs found in both corpora, and the number of unique speakers (#sp) and conversations (#conv).

3. Analysis

As in our previous study [4] we extracted several acoustic features from all laughs: duration, percentage of voiced frames, mean F_0 (henceforth pitch), mean intensity and the maximum intensity. The measurements were transformed to z-scores (for each speaker separately) by using $z = (x - \mu) / \sigma$ where μ and σ were derived from the speaker’s vocalised utterances. Based on these features, we first determined whether initiating, responding, and non-overlapping laughs differ from each other. Second, we looked more closely at the pairing of initiating and responding laughs. How strongly are initiating and responding laughs paired to each other, and can we say that there is a similarity relation between these two?

3.1. Comparing among initiating, responding and non-overlapping laughs

We assessed differences in the aforementioned acoustic features among the initiating, responding, and non-overlapping laughs. The three groups formed the independent variable *group* while the dependent variables consisted of the five acoustic features. Several Kruskal Wallis tests were performed, followed by post-hoc Mann Whitney U tests. Since multiple Kruskal Wallis tests were performed, the significance level was adjusted to .01. If this test was significant, we followed up with post-hoc tests that were considered significant if $p < .003$ (adjusted for multiple comparisons). We discuss the results per corpus.

Diapix corpus: The Kruskal Wallis tests were not significant for mean intensity ($\chi^2(2) = .453, p = .797$) and percentage of voiced frames ($\chi^2(2) = 6.144, p = .046$). The tests did show a significant effect for the group variable for mean pitch ($\chi^2(2) = 11.7, p = .003$), max intensity ($\chi^2 = 13.478, p = .001$), and duration ($\chi^2 = 39.823, p = .000$). For these variables, we performed post-hoc tests that consisted of (two-tailed) Mann Whitney U tests for each of the three comparisons.

For mean pitch, we find that responding laughs have a higher mean pitch than non-overlapping laughs ($U = 12892, p = .003$). When we compare initiating laughs to responding laughs ($U = 5310, p = .733$) or to non-overlapping laughs ($U = 12892, p = .014$), we do not find significant differences.

For max intensity, tests show that initiating laughs are significantly different from non-overlapping laughs ($U = 14726, p = .000$) with initiating laughs having higher max intensity. Comparing initiating laughs to responding laughs yield a nearly significant difference ($U = 5267, p = .008$) while comparing responding laughs to non-overlapping laughs does not yield a significant difference ($U = 20276, p = .842$).

For duration, significant differences are found when we compare initiating laughs to non-overlapping laughs ($U = 11491, p = .000$) and when we compare initiating laughs to responding laughs ($U = 4894.5, p = .001$); in both comparisons, the initiating laughs are longer. No significant difference is found between responding laughs and non-overlapping laughs ($U = 17966, p = .042$).

A summary of all the comparisons made in the post-hoc analyses is shown in Table 1. In conclusion, the results are not conclusive. For most of the features, it seems that non-overlapping laughs are the most distinctive group. We can also observe that initiating and responding laughs are not very different from each other, with the exception of duration. We will apply similar analyses to the HCRC MapTask corpus and compare the results of both corpora.

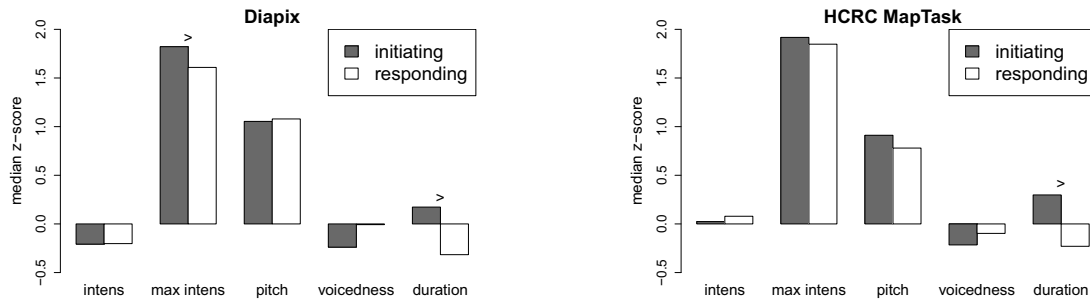


Figure 2: Barplots showing medians of initiating and responding laughs. > indicates statistical significance ($p < .01$) of Wilcoxon signed-rank tests assuming paired samples of laughs.

HCRC MapTask corpus: According to Kruskal Wallis tests and in correspondence with the Diapix corpus, mean intensity ($\chi^2(2) = 3.614, p = .164$) and percentage voiced frames ($\chi^2(2) = 4.561, p = .102$) do not show significant results. The tests for mean pitch ($\chi^2(2) = 26.587, p = .000$), max intensity ($\chi^2(2) = 35.382, p = .000$), and duration ($\chi^2(2) = 83.268, p = .000$) do show significant results. For the variables that showed significant results, post-hoc (two-tailed) Mann Whitney U tests were performed for each of the three comparisons.

For mean pitch, post-hoc analyses show that both initiating ($U = 19493, Z = -4.491, p = .000$) and responding laughs ($U = 21758, Z = -3.492, p = .000$) are significantly different from non-overlapping laughs. But initiating and responding laughs are not significantly different from each other ($U = 7750, Z = -1.151, p = .250$).

For max intensity, we find similar results: both initiating ($U = 39944, Z = -4.739, p = .000$) and responding laughs ($U = 42595, Z = -4.544, p = .000$) are different from non-overlapping laughs showing a higher max intensity. There is no significant difference between initiating and responding laughs with respect to max intensity ($U = 14881, Z = -.081, p = .935$).

For duration, we find that initiating laughs are significantly longer than both responding ($U = 10040.5, Z = -5.286, p = .000$) and non-overlapping laughs ($U = 28589, Z = -9.062, p = .000$). But responding laughs are not significantly longer than non-overlapping laughs ($U = 47490, Z = -2.732, p = .006$).

In conclusion, the observations for the HCRC MapTask corpus are in concordance with those obtained for the Diapix corpus, and are more pronounced than the Diapix results. Non-overlapping laughs show distinctive characteristics from initiating and responding laughs, while initiating and responding laughs seem to be relatively similar to each other (except for duration). Responding laughs seem to hold an ‘intermediate’ position between initiating and non-overlapping laughs with respect to the features studied; responding laughs seem to be somewhat distinct from the non-overlapping laughs, but initiating laughs appear to be more distinctive from non-overlapping laughs.

3.2. Comparing between paired initiating and responding laughs

In the previous subsection, we compared the medians and spread of several features taken from three groups of different types of laughs. We found that initiating and responding laughs in general share similar feature characteristics. Here,

we will take a closer look at the similarities and differences between pairs of initiating and responding laughs. First, we compared medians from the initiating and responding laughs, taking the pairing attribute into account. Assuming paired samples of laughter, Wilcoxon signed-rank tests, that compares two related samples, were carried out for each feature (while adopting a corrected p -level of .01). Second, because initial analyses showed that initiating and responding laughs showed several shared similar acoustic characteristics between the two, we were interested in seeing whether there was a ‘similarity relation’ between an initiation and responding laugh. We assessed this ‘similarity relation’ through a method based on comparing so-called ‘pseudo-paired’ laughs (see [12]) to genuinely-paired laughs.

3.2.1. Compare medians between paired initiated and responding laughs

For each corpus, we performed Wilcoxon signed-rank tests (assuming paired samples) and considered the tests significant if $p < .01$.

Diapix corpus: The Wilcoxon signed-rank test showed that intensity ($Z = -.623, p = .533$), pitch ($Z = -.873, p = .383$) and percentage voiced frames ($Z = -1.658, p = .097$) are not significant. The features max intensity ($Z = -3.434, p = .001$) and duration ($Z = -3.714, p = .000$) do show significant differences. It seems that an initiating laugh has higher max intensity and longer duration than its responding counterpart.

HCRC MapTask corpus: For the HCRC MapTask corpus, we only find one significant difference: an initiating laugh appears to be longer in duration than its paired responding laugh ($Z = -5.422, p = .000$). No significant results were found for intensity ($Z = -1.662, p = .096$), pitch ($Z = -.378, p = .705$), max intensity ($Z = -.273, p = .785$), and percentage of voiced frames ($Z = -2.456, p = .014$).

A summary of the results is shown in Fig. 2. It is no surprise that the results are very similar to the ones shown in Table 1. Initiating laughs seem to have a longer duration than responding laughs, and, at least in the Diapix corpus, initiating laughs have a higher max intensity than responding laughs.

3.2.2. Is there a similarity relation between the initiating and responding laugh?

We wanted to gain more insights into the relation between a paired initiating and responding laugh. To that end, we car-

ried out an analysis that builds on distances between genuinely paired and ‘pseudo’ paired features, e.g., [12, 13]. We created two conditions. In the *genuine* condition, we calculated Euclidean distances between features of the original, genuine initiating and responding laugh pairs. In the *pseudo* condition, we created pairs of laughs that were not genuinely paired to each other. Each initiating laugh was paired with every other responding laugh while excluding the originally paired responding laugh. For each initiating laugh, the minimum and average distance over all possible pairings were calculated and used as a ‘pseudo’-paired distance. If the distances calculated in the genuine condition are significantly smaller than the distances calculated in the pseudo condition (using the minimum or average distance), then we have reason to assume that there is a similarity relation between the initiating and responding laugh.

In order to assess whether distances in the genuine condition are indeed smaller than in the pseudo condition, multiple one-tailed Wilcoxon signed-rank tests were performed. Differences were considered significant if $p < .0083$ (adjusted for multiple comparisons). In the pseudo condition, we calculated for each initiating laugh a ‘pseudo-paired’ distance based on the minimum and average distance of all possibly paired responding laughs (excluding the originally paired one). We found that using the ‘pseudo-paired’ minimum distance was too strict; for each feature, this always yielded significantly smaller distances for the pseudo condition in comparison with the genuine condition. In other words, in the pool of responding laughs, there was always a laugh to be found that yielded a smaller distance than the original paired responding laugh; obviously, the larger the pool, the larger the chance that such a laugh could be found. Therefore, we decided to be less strict and mitigated this factor by carrying out the analyses using the average distance.

Diapix corpus: We found for the Diapix corpus that only pitch and max intensity seem to be significant, see Table 3. For pitch, there seems to be a similarity relation between an initiating and responding laugh. For max intensity, the relation is unclear since previously, we found that max intensity appeared to be higher for the initiating laugh than the responding laugh, see Fig. 2.

HCRC MapTask corpus: For the HCRC MapTask corpus, all features were significant except max intensity, see Table 4. The results shown in Fig. 2 indicate that the intensity, pitch and voicedness for paired initiating and responding laughs have similar values, while the duration of an initiating laugh is longer than that of a responding laugh. It seems that we can say that intensity, pitch, and voicedness show a similarity relation while for max intensity and duration, this is not clear.

In conclusion, although initiating and responding laughs seem very similar acoustically, only for a few features, we could show a similarity relation according to this ‘pseudo-pairing’ paradigm method. For the Diapix corpus, only pitch shows a similarity relation. For the MapTask corpus, the results are more pronounced: intensity, pitch, and voicedness appear to show similarity relations.

4. Conclusion and discussion

We have compared prosodic characteristics among initiating, responding, and non-overlapping laughs. Based on the results obtained, we can come to two main conclusions. First, we can confirm that non-overlapping laughs are very distinctive from initiating laughs, and that responding laughs seem to have intermediate-level features, showing feature values that lie between initiating and non-overlapping laughs. Second, we found

Feature	dist(A·B) vs. dist(A·B')
Intens	$Z = -2.0498, p = .0202$
Pitch	$Z = -3.2418, p = .0000$
Percvoiced	$Z = -1.2743, p = .1013$
Maxintens	$Z = -2.4438, p = .0073$
Duration	$Z = .83289, p = .2025$
All (combined)	$Z = -1.9924, p = .0232$

Table 3: *Diapix corpus: results of (one-tailed) Wilcoxon signed-rank tests, testing the difference between the genuine (dist(A·B)) and ‘pseudo’ distance (dist(A·B')), bold means significant at $p < .0083$.*

Feature	dist(A·B) vs. dist(A·B')
Intens	$Z = -3.0650, p = .0011$
Pitch	$Z = -4.5271, p = .0000$
Percvoiced	$Z = -4.8954, p = .0000$
Maxintens	$Z = -2.1130, p = .0173$
Duration	$Z = -2.4743, p = .0067$
All (combined)	$Z = -4.7523, p = .0000$

Table 4: *MapTask corpus: results of (one-tailed) Wilcoxon signed-rank tests, testing the difference between the genuine (dist(A·B)) and ‘pseudo’ distance (dist(A·B')), bold means significant at $p < .0083$.*

that, although initiating and responding laughs show similar acoustic characteristics, only for a few features (particularly pitch), the initiating and responding laughs seem to be paired systematically and show a relation that could indicate an accommodation/mimicry phenomenon. Additional observations for duration show that the initiating laugh seems to be longer than the responding laugh (although the just significant similarity relation found in the HCRC MapTask corpus suggests otherwise). Lastly, we did not find clues that point towards a Lombard effect in joint laughter.

We can see the finding of some similarity links as possible evidence that the paired laughs are linked to each other in the way that the speakers imitate (or entrain to) each other to a certain level. This kind of mimicry has parallels to speech shadowing where subjects have to repeat words as they hear it. Research, e.g. [14, 15] shows that shadowed words were perceptually more similar to the speech to be shadowed than words spoken without shadowing. Moreover, words of immediate shadowing show a greater perceptual similarity than those in delayed shadowing [14]. Our findings regarding the acoustic similarity of paired laughs in a social situation can be interpreted that there might be a general tendency for vocal imitation that goes beyond *verbal* vocalisations in a *non-social* setting.

Since our first results are not very conclusive, there is much room for future work. In future studies for example, the variability of the speakers can be better controlled for. Since we only looked at a few prosodic parameters, we also suggest to employ spectral and cepstral features as well in combination with techniques such as the ones described in [16]. Additionally, in order to see whether this possible similarity/mimicry relation between initiating and responding laughs also exists perceptually, we suggest to follow-up with a perception experiment in which subjects judge the genuine and ‘pseudo’ pairs of laughs to find out whether this relation also exists perceptually.

5. References

- [1] J.-A. Bachorowski and M. J. Owren, "Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect," *Psychological Science*, vol. 12, pp. 252–257, 2001.
- [2] J.-A. Bachorowski, M. J. Smoski, and M. J. Owren, "The acoustic features of human laughter," *Journal of the Acoustical Society of America*, vol. 110, pp. 1581–1597, 2001.
- [3] J. Trouvain and K. P. Truong, "Comparing non-verbal vocalisations in conversational speech corpora," in *Proceedings of the LREC Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, 2012, pp. 36–39.
- [4] K. P. Truong and J. Trouvain, "On the acoustics of overlapping laughter in conversational speech," in *Proceedings of Interspeech*, 2012, pp. 851–854.
- [5] —, "Laughter annotations in conversational speech corpora - possibilities and limitations for phonetic analysis," in *Proceedings of the Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, 2012, pp. 20–24.
- [6] G. Jefferson, "A technique for inviting laughter and its subsequent acceptance/declination," in *Everyday language: studies in ethnomethodology*, G. Psathas, Ed. New York: Irvington, 1979, pp. 79–96.
- [7] J. Trouvain and K. P. Truong, "Exploring sequences of speech and laughter activity using visualisations of conversations," in *Proceedings of the Workshop on Affective Social Speech Signals*, 2013.
- [8] M. Smoski and J. Bachorowski, "Antiphonal laughter in developing friendships," *Annals of the New York Academy of Sciences*, vol. 1000, pp. 300–303, 2003.
- [9] H. Lane and B. Tranel, "The Lombard sign and the role of hearing in speech," *The Journal of Speech and Hearing Research*, vol. 14, no. 4, pp. 677–709, 1971.
- [10] 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.
- [11] 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.
- [12] F. J. Bernieri and R. Rosenthal, "Interpersonal coordination: Behavior matching and interactional synchrony," in *Fundamentals of nonverbal behavior*, R. S. Feldman and B. Rime, Eds. New York: Cambridge University Press, 1991, pp. 401–432.
- [13] R. Levitan, A. Gravano, and J. Hirschberg, "Entrainment in speech preceding backchannels," in *Proceedings of the ACL (Short Papers)*, 2011, pp. 113–117.
- [14] S. Goldinger, "Echoes of echoes? An episodic theory of lexical access," *Psychological Review*, vol. 105, no. 2, pp. 251–279, 1998.
- [15] K. Shockley, L. Sabadini, and C. Fowler, "Imitation in shadowing words," *Perception and Psychophysics*, vol. 66, pp. 422–429, 2004.
- [16] C.-C. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions," *Computer Speech and Language*, vol. 28, pp. 518–539, 2014.