Towards unravelling prosodic characteristics of speaker-overlapping laughing in conversational speech corpora

Jürgen Trouvain and Khiet P. Truong

Abstract
It is often reported that in spontaneous discourse the laughing of interlocutors overlaps with each other. Although transcripts of interactional linguistics do consider features of overlap, the main prosodic information regarding pitch, intensity, duration and interactional timing of laughter so far remained untracked. This paper aims to show that acoustic-phonetic studies can help to unravel the prosodic characteristics of speaker-overlapping laughs. For this purpose the annotated laughs of four corpora of conversational speech were analysed. The inspection of several thousands of laughs reveals that overlapping laughs are significantly longer, with a higher intensity, a higher fundamental frequency and are more voiced than non-overlapping laughs. Moreover, the frequency of overlapping laughs seems to be determined by the number of participants. Finally, a closer analysis of the interactional timing in one dialogue corpus revealed that the scheme "speaker invites recipient to join in to laugh" occurs more often than "recipient laughs before speaker".

laughter – corpora – conversational speech – overlapping vocalisation - prosody

Es wird häufig berichtet, dass sich im spontanen Diskurs die Gesprächspartner beim Lachen überlappen. Obwohl Transkripte in der Gesprächsforschung Merkmale der Überlappung beachten, blieb bisher die primäre prosodische Information zu Tonhöhe, Intensität, Dauer und Timing in der Interaktion unbeachtet. Akustisch-phonetische Studien können dabei helfen, die prosodischen Charakteristika sprecher-überlappenden Lachens aufzudecken. Zu diesem Zwecke wurden mehrere Tausend annotierter Lachereignisse in vier Gesprächskorpora analysiert. Es zeigte sich, dass überlappende Lachereignisse signifikant länger, mit höherer Intensität, einer höheren Grundfrequenz und mehr stimmhaften Anteile produziert wurden als nicht-überlappende Lacher. Die meisten Lacher in Dialog-Korpora waren allerdings nicht-überlappend, im Gegensatz zu einem Korpus mit Mehr-Parteien-Gesprächen. Zudem zeigte eine genauere Analyse des interaktionellen Timings in einem Dialog-Korpus, dass das Schema "Sprecher lädt Rezipient zum Lachen ein" häufiger vorkommt als "Rezipient lacht vor dem Sprecher".

Lachen – Korpora – Gesprächsdaten – überlappende Vokalisierungen – Prosodie

1. Introduction

In contrast to most read speech we can frequently find in conversational speech nonverbal vocalisations such as laughter, coughing or clearing the throat. When we exclude so-called fillers like *uh* and *uhm* as well as feedback expressions such as *hm*, *u-huh* and *yeah* from the nonverbal vocalisations, laughter seems to be the

most frequent vocalisation after audible breathing sounds (Trouvain/ Truong 2012). Although laughing is so immanently present in every-day speech, research of laughter in conversations can so far be considered as rather limited from a phonetic/ prosodic as well as from a perspective of discourse and conversation analysis (exceptions are the collections in Trouvain/ Campbell 2007 and Wagner/ Vöge 2010). The aim of this paper is to present some findings of the prosodic characteristics of laughter in interaction which can be of interest for discourse and conversation analysis.

1.1. Transcribing laughter

In discourse and conversation analytical transcription systems such as the "Gesprächsanalytische Transkriptionssystem 2" (GAT 2) (Selting et al. 2009: 367), it is recommended to transcribe laughter either as *hahaha, hehe, hihi* when it is a shorter or "syllabic" laugh corresponding to the number of pulses or laugh syllables, or as "((laughs))" between double rounded brackets as a description of laughter. Speech-laughs or laughed speech should be transcribed as "<<laughing> word >" whereas the convention for smile voice is "<< :-) > word>".

In their example transcript taken from a real conversation, Selting et al. (2009: 394) depict laughter at various places. The transcriptions of various laughing events differ from event to event: In line 28 the laughs start with *phhh*, continued by *hohoho*, in lines 38 and 39 *hm*, in lines 40 and 40 and 41 it is *(lacht verstohlen, ca. 1.2 Sek lang)*, in lines 62, 63 and 65 it is *hehehe* or *he he*, and in line 66 *he ho ho*. (Interlinear translations of the text were inserted below the respective lines in the transcript.)

(1) Excerpt from Selting et al. (2009)

```
27   S2:   und sich mit  den NACHbarn   ange[legt,=ne,  ](1.2)
           and PRON with the neighbours argued     PTCL
           and argued with the neighbours

28   S1:                                     [phhh hohoho]

((...))

36   S2:   und ham wir immer  gesagt HIT[ler;=ne,]
           and have we always said   Hitler   PTCL
           and we always said Hitler

37   S1:                                  [HITler; ]

38   S1:   h[m, ]

39   S2:    [hm,]
40   S1:   [((lacht verstohlen,        ca. 1.2 Sek. lang))]
           ((laughs in a stealthy way, ca. 1.2 sec.))

41   S2:   [((lacht verstohlen,        ca. 1.2 Sek. lang))]

((...))
```

61 S2: jetzt wohnt er nämlich mit seiner SCHW:ESter
now lives he actually with his sister

zusa[mmen.]
together
now he actually lives together with his sister

62 S1: [hehehe]

63 S1: he [he]

64 S2: [und] die LAUfen: (.) RUM wie n URaltes
and they run around like an old-age

ehe[paar.]=ne,
couple PTCL
and they behave like an old-age couple

65 S1: [he he]

66 S1: he ho [ho]

67 S2: [OUH] mann. (.) heh
PTCL man
Oh boy!

The diversity of these transcriptions show various interpretations of various laughter events, be it a laugh with a "syllabic" structure ("hohoho"), laughs with different vowel qualities ("he" vs. "ho"), laughing without any vowel articulation ("hm"), with a more literal description ("verstohlen" in German or "stealthy" in English) or with temporal information ("ca. 1.2 Sek. lang"). Especially with the interpretative comments, it is unclear whether a second transcriber would come up with the same or similar transcription of the laughing events. While it is not the aim of this paper to criticise this kind of diversity in transcription, it shows two things: 1) each transcription is always an interpretation. This holds in particular for non-words without standard spelling, and 2) the diversity of the ways to transcribe laughter mirrors the diversity of laughter in general. The current paper would like to take up this issue by comparing the prosodic features of speaker-overlapping laughter with those of non-overlapping laughter.

1.2. Overlapping laughter

The transcripts above show a further important feature of laughter in interaction: all laughs in this example transcript were overlapping, either with speech of the co-participant (e.g. lines 28-65) or with laughter of the co-participant (e.g. lines 38-41). It is not unusual that speakers overlap with their vocalisations. The conversational principle "one speaker at a time" (Sacks et al. 1974, Stivers et al. 2009) does not always hold (see e.g. Liddicoat 2007). This restriction is supported by corpora of conversational speech where a considerable amount of "cross-talk" was observed (e.g. Campbell 2007b or Heldner/ Edlund 2010). Laughter in particular has a tendency to overlap with laughter as could be shown by Laskowsi/ Burger (2007), Truong/ Trouvain (2012b) and also Smoski/ Bachorowski (2003).

Laughter seems to serve as the optimal opportunity for a joint vocalisation. In lines 40 and 41 of the transcript above we see synchronous laughter of both interlocutors whereas the laughter in lines 28, 62, 63 and 65 overlaps with speech of the interlocutor. Interestingly, the overlaps in the latter cases are transcribed as aligned on a syllabic basis, e.g. "he [he]" in line 63 aligns with "[und]" in line 64.

We define overlapping laughter (henceforth OL) here as laughter of speaker A that overlaps with laughter of speaker B. We consider all other instances of laughter as non-overlapping laughter (henceforth NOL), i.e. the laughter in lines 28, 62, 63 and 65 would be NOL. So far it seems unclear whether OL can be seen as a frequent phenomenon in conversational speech. But in any case it would be interesting to find out more about the mechanisms of how interlocutors manage to produce an OL in talk-in-interaction. One plausible challenging explanation is the "invitation-acceptance scheme" proposed by Jefferson (1979):

> Laughter can be managed as a sequence in which speaker of an utterance invites recipient to laugh and recipient accepts that invitation. One technique for inviting laughter is the placement, by speaker, of a laugh just at completion of an utterance, and one technique for accepting that invitation is the placement, by recipient, of a laugh just after the onset of speaker's laughter.

1.3. Research questions

This scheme provides an elegant explanation for how partners in conversation accomplish OL, which might be a common form in social interaction. Nevertheless, there are various questions left unanswered:

- Is there a difference in the phonetic substance between OL and NOL?
- Do interlocutors show proportionally more or fewer OL than NOL?
- Among the OL, how many times does the invitation-acceptance scheme apply?
- How does this scheme work in time? Do laugh invitees wait to be invited?

This paper will take up such questions. Before we present some of our findings we provide some background on the phonetics of laughter and overlapping vocalisations.

2. Phonetic background on laughter and overlapping vocalisations

2.1. Phonetic sciences on laughter

Laughter typically occurs in social interaction, especially in conversations and not so much in read speech or other forms of monologues. Since phonetic research consists to a substantial degree of investigations of controlled speech, conversations are only infrequently the object of this branch of research. In addition, in phonetics typically single words and sentences of read speech are investigated, thus ignoring nonverbal vocalisations at all.

The limited amount of research on laughing performed by phoneticians brought about the phonetics of laughing as a multi-disciplinary research matter. Among them are psychological disciplines with focus on developmental, evolutionary or affective perspectives but also speech technology and other speech-based disciplines such as conversation analysis and interactional linguistics.

2.2. Defining laughter

Although there is no accepted standard definition of laughter, there seems to be an agreement that laughter is a social and/ or affective signal expressed with an acoustic and a visual display in interaction with others. There are various reasons why humans laugh. In the literature, factors such as cooperation, social bonding, affiliation, creating a pleasant atmosphere, feedback or back-channelling, topic termination or humour are mentioned among others (see e.g. Bachorowski et al. 2001, Smoski/ Bachorwoski 2003, Chafe 2007, Holt 2010). This list suggests that laughter is connected to many factors at rather diverse levels.

A question still unsolved is whether laughing and smiling belong to the same category or whether they are two distinct categories. Depending on the definition, smile voice could theoretically also belong to laughter, however, we consider it here as different from laughter. Smiling per se happens in a silent way whereas laughter per se happens in a non-silent way. The acoustic transmission of smiling needs some speech as a "carrier", whereas laughter is usually produced as an autonomous vocalisation. However, laughing can also be produced during articulation of speech (Nwokah et al. 1999, Trouvain 2001). These speech-laughs or laughed speech is again distinct from smiled speech. Often a speech-laugh turns into a laugh vocalisation.

When we look more closely into laugh vocalisations we see rather quickly that there is not only *one* type of laughter (e.g. often transcribed with "hahaha" or "hehehe") but that there is a bundle of various phonetic forms and combinations of these forms (see also Trouvain 2003). The following parameters seem to be relevant:

- voicing,
- duration,
- the number of "syllables" or "calls" (cf. Bachorowski et al. 2001) or "pulses" (Chafe 2007),
- onset noise before and an offset of inhalation and silence after the laughed "syllables" (cf. Chafe 2007),
- intensity,
- fundamental frequency (can be often extremely increased).

It should be noted that there is a huge variability regarding the listed parameters between individuals but also intra-individually (e.g. Vettin/ Todt 2004). This large phonetic variation of laughs also includes a considerable amount of laughs that show a rather mild intensity with a comparably short duration – phonetic conditions that make these mild laughs susceptible to be overheard, particularly when

the listener is focused on speech or when the laugh events occur in overlapping vocalisations.

2.3. Overlapping vocalisations in conversations

Laughter seems to represent an optimal opportunity for joint vocalisation. Various studies show that there is a strong tendency of laughter to overlap with vocalisations (either speech or nonverbal vocalisations) of the interlocutor (Laskowsi/ Burger 2007, Truong/ Trouvain 2012b and also Smoski/ Bachorowski 2003). Joint vocalisations are not uncommon. Think of "collaborative completions" as continuations by the conversational partner with matching prosodic features (e.g. Lerner 1991, Local 2005). Other joint vocalisations that require across-speaker coordination are, for instance, synchronous reading aloud (Cummins 2007), singing in a choir and a common prayer in church.

Fig. 1 illustrates the often observed overlap of dialogue partners when laughing with an example taken from the German Lindenstraße corpus (IPDS 2006) transcribed in (2).

(2) Lindenstraße Corpus (IPDS 2006, Dialogue 4, sec. 164-173)

164 L: is halt son skelett im flur und sie meint dann so
is just such a skeleton in corridor and she says then so

166 mai, was <<interjection> pfn>
well what INTERJ

167 <<laughing>machen die leut [da> (---) °h]
make the people there
There is a skeleton in the corridor and she goes like "well, what are the people doing there?"

168 S: [hehehehehehehe °h]
169 echt nee hier wars so, dass wir ne kurze szene wo
really no here was it so that we a short scene where

sie in der straße entlanggehen und dann °h
they in the street go along and then
Really? No, here we had a short scene where she walked along the street and then …

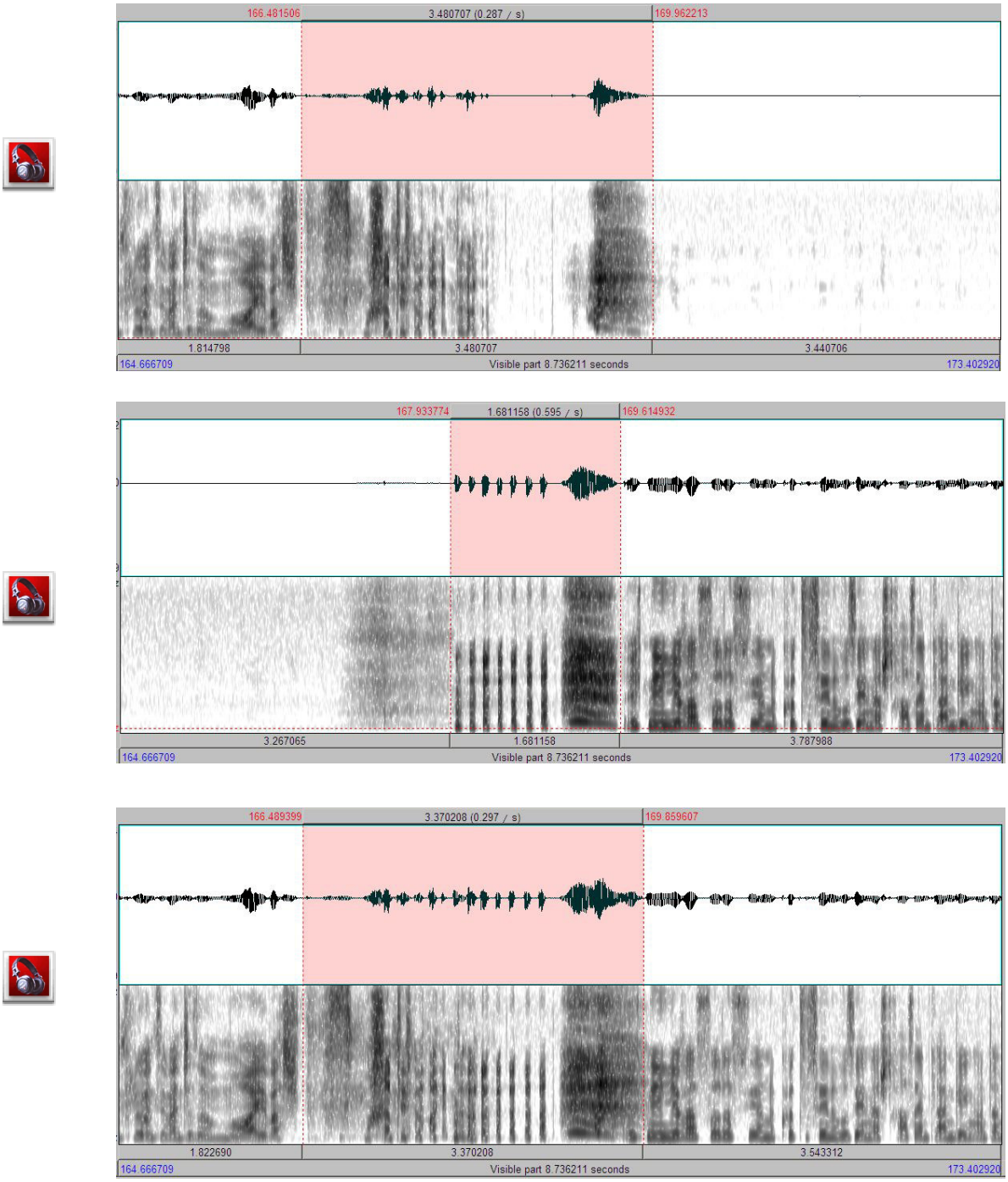


Fig. 1: Example of laughter at turn-taking (from Lindenstraße Corpus (IPDS 2006), Dialogue 4, from 164 to 173 sec) represented as wave form and spectrogram (0-8 kHz). Top: left channel speaker. Mid: right channel speaker. Bottom: mixed speech signal. Marked with colour: the laugh production.

Since each speaker was recorded with a different microphone, their vocalisations can be represented separately. Such a recording with separate channels allows us to determine the *exact timing* of the laughter in the dialogue flow and the *exact phonetic-prosodic realisations* of the laughter by each speaker. In recordings without channel separation (which seems to be the case in most dialogue corpora, e.g. the audio file that corresponds to the example transcript in Selting et al.

2009), these details of prosody are masked. In addition, some less intense vocalisations may not even be noticeable when masked in a recording without channel separation.

3. Method

We report here selected aspects from previous studies on laughing based on already annotated corpora of conversational speech (Truong/ Trouvain 2012a, b and Trouvain/ Truong 2012, 2013). In contrast to purely auditory approaches, which are usual in conversation analytic research, we are interested in the acoustic properties of various prosodic parameters including the proportion of voicing, the fundamental frequency, the intensity and the duration of laugh tokens. These acoustic characteristics have the advantage that they can be compared across corpora and independently of the interpretation of transcribers. A further advantage is that *large* amounts of data can be analysed by automatic procedures.

The size of data we are working with goes beyond the data set usually considered in qualitative studies as used in conversation and interactional analysis. We would like to stress that this difference is not meant to disqualify qualitative studies. Rather, quantitative studies are considered a way to *complement* qualitative studies in order to strengthen the findings and theoretical statements therein.

There are two important differences between qualitative and quantitative approaches of the phonetic/ prosodic analysis of vocalisations. First, for a quantitative analysis the data will be treated automatically and not manually as it is usual for a qualitative analysis. There are some prerequisites for an automatic processing: the electronic availability of annotations, the time stamps of the start and the end of the annotated laugh token, and the corresponding speech signal files for all speakers of the conversation. As stated before, the channels of both speakers must be separated in order to clearly recognise the overlap, which is important to perform an acoustic-phonetic analysis.

Second, employing a quantitative analysis of already existing corpora means that the annotations were made by others than the current researchers. This fact can sometimes lead to some disagreement between the existing annotations of certain laugh tokens and how the researchers would have annotated the same token (for more details on these issues see Truong/ Trouvain 2012a). When working with smaller sample sizes the mentioned disagreements could affect the findings. For our rather big sample size (more than 18,000 laugh tokens, see also Table 1) we consider this risk marginal.

3.1. Corpora used

Four English-speaking corpora of conversational speech were selected for analysis.

1. The AMI Meeting Corpus (Carletta 2007) contains audiovisual recordings (100 hours) of elicited design meetings, in which a team of 4 persons are discussing the design of a new remote control. In these meetings, the par-

ticipants were put together in a room and were assigned certain roles to play (project manager, marketing expert, industrial designer, and user interface designer).

2. The ICSI Meeting Corpus (Janin et al. 2003) contains audio recordings of 75 natural research meetings that were actually held at the International Computer Science Institute (ICSI) over a period of 3 years. The number of participants per meeting varies between 3 and 11, yielding an average of 6 participants per meeting. The topics of the meetings concern, among others, natural language processing, the development of the ICSI Meeting Corpus, and methods to compensate for noise for automatic speech recognition.
3. The Diapix Lucid Corpus (Baker/ Hazan 2011) contains audio recordings of unscripted task-based dyadic interactions. Each participant is given a different version of a cartoon picture and is seated in a separate room. The two participants communicate via headsets to locate the twelve differences between the two pictures.
4. The HCRC Map Task Corpus (Anderson et al. 1991) contains audio recordings of unscripted task-based dyadic interactions (with and without eye contact). Each participant is given a different version of a map and a role, that of "instruction giver" or "instruction follower". The instruction giver's map contains a route that should be reproduced on the instruction follower's map with as few deviations as possible.

As shown in Table 1 the four corpora differ severely with respect to number of recorded speakers, number of recorded conversations and duration of each conversation. Consequently, the number of annotated laughs differs as well. Differences in the annotation practice and on the technical level (more details in Truong/ Trouvain 2012a) led to a reduction of the number of laughs used in the current analysis. The main differences concern:

- speech-laughs which are sometimes ignored and sometimes inconsistently labelled (for this analysis we ignored them altogether).
- the definition of what counts as a 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.
- technical errors: sometimes the annotated laughs show incorrect time stamps for beginning and/ or end.

Despite the listed drawbacks of annotation we consider the existing corpora as usable – conceding, though, that we are not considering completely correct data. A very time-consuming re-annotation would be necessary to obtain more homogeneous laughter annotations across corpora, that in turn will lead to more consistent and reliable research results.

On a qualitative level the corpora are not homogenous either. AMI and ICSI are multi-party conversations whereas Map Task and DiaPix are dialogues. The level of acquaintance is rather different, with the exchanges including friends, colleagues and also strangers. The sensual modalities also differ: two corpora (AMI

and ICSI) used face-to-face conversations, one corpus (DiaPix) used only the acoustic channel and one corpus (Map Task) used both conditions. There are also differences regarding the communicative settings in which the recordings were situated with the ICSI corpus containing authentic conversations and the other three corpora with naturalistic but task-based speech.

	AMI	**ICSI**	**Map Task**	**DiaPix**
no. of annotated laughs	16477	12574	1002	582
no. of used laughs	8587	8268	966	575
no. of speakers	679	494	250	114
no. of conversations	165	74	95	52
no. of speakers per convers.	4	3-11	2	2
mean duration of convers. (in mins, standard deviation in parentheses)	35.1 (13.5)	55.0 (15.9)	6.8 (3.1)	7.7 (2.3)
total durat. of all convers. (in hours)	94.5	67.9	11.4	6.8
visual contact	yes	yes	yes/ no	no
relationship between speakers	mostly strangers	colleagues	friends, strangers	friends

Table 1: Descriptive features of the four inspected corpora.

Apart from these differences, we can assume that *single* tokens of laughter from these (and other) corpora may heavily depend on conditions like visual contact, social distance, and communicative task. There is no doubt that a *qualitative* analysis has a high level of explanatory power – however, it is restricted to an *extremely limited* number of tokens (sometimes n=1). When we are interested in generalisations, for instance to check a certain hypothesis based on a qualitative analysis, then certain qualitative prosodic features of laughs should occur across various corpora and the inherent differences described above. Corpora with annotations with hundreds and thousands of tokens can represent a solid base for such an approach.

3.2. Analysis

We carried out a semi-automatic acoustic analysis of the prosody of corpora of conversational speech which are already annotated. Ideally, the annotated labels "laughter" are aligned with the corresponding section in the speech signal. These sections of speech can be automatically analysed with self-authored Praat scripts for prosodic parameters such as duration, fundamental frequency (F0), intensity and also the amount of voiced frames.

The temporal alignment was important to see whether speakers overlap or not. Please note that two laughs which overlap are always *partially* overlapping. Of course for some laughs one can have the auditory impression that both speakers start laughing at the same time. However, with the temporal granularity of milliseconds which is usual for acoustic-phonetic studies the delay of time is always nicely visible. Schematically such overlaps can be pictured like in Fig. 2, where

the aforementioned "invitation-acceptance-pattern" by Jefferson (1979) is illustrated.

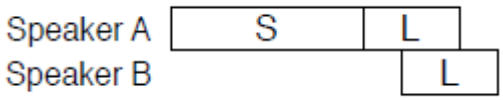


Fig. 2: Schematic illustration of two overlapping laughs (L). The current speaker ends the phase of articulated speech (S) with a laugh. The speaker in the role of the recipient joins in with a laugh after the speaker's laugh.

Transcript 3 from Jefferson (1979: 81) can serve as an example where the speaker laughs at the completion of an utterance and that the recipient accepts this invitation by laughing after the onset of the speaker's laugh.

(3) Excerpt from Jefferson (1979: 81)

```
Ellen:   He s'd well he'd said I am cheap he said, 'hh
         about the big things. he says but not the liddle
         things, hhhHA HA [HA HA HA
Bill:                     [heh heh heh
```

The advantage of taking already annotated data is that the annotation of thousands of laughs is already available (cf. Table 1). The already mentioned disadvantage is that many various labellers performed the annotation in possibly different ways. As a laugh we consider a span of vocalisation of one speaker annotated as "laugh". This means that an overlap of laughing as illustrated in Fig. 2 counts as two laughs (one laugh from speaker A and one laugh from speaker B). It is important to note that speech-laughs were not considered here because in some corpora they were not regarded at all and in others they were not annotated in a consistent way.

The frequencies of occurrences of overlapping (OL) and non-overlapping laughs (NOL) were determined in all four corpora with the help of Praat scripts. OL are laughs like those in Fig. 2. Please note that NOL are laughs that do not overlap with the laugh of the interlocutor but they can overlap with the speech of the other. For each laugh annotated and used, the mean values of the fundamental frequency (F0), the intensity, the duration and the number of voiced frames were automatically extracted, again using Praat scripts. Thereafter the values for F0 and intensity were normalised (using a z-transformation as the standard procedure of normalisation).

A closer analysis of the timing schemes was performed with the data of the Map Task corpus only. For this purpose we first created plots of "speech and laugh activities" like the one in Fig. 3.

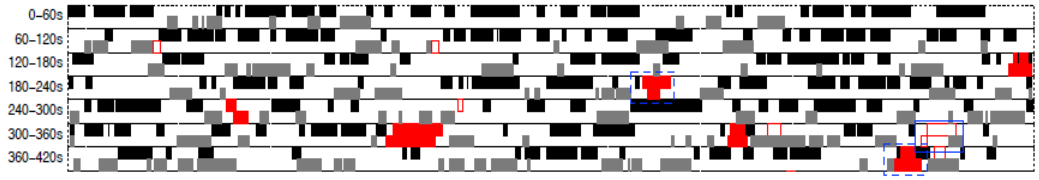
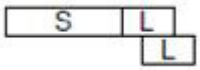
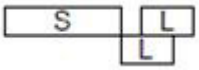


Fig. 3: Speech and laugh activity of the first seven minutes of the dialogue q1nc2 from the Map Task corpus. Each line represents 60 seconds of speech with filled dark bars for the speech activity of the speaker with the instructor role and filled grey bars for the speech activity of the speaker with the receiver role. Empty sections represent silences. Filled red bars stand for overlapping laughs, empty red bars for non-overlapping laughs. Blue rectangle with solid border marks an example of a clear laughter pair in which the laughs do not overlap. Blue rectangles with dashed borders mark examples of Jefferson's invitation-acceptance laughter pairs.

Theoretically, two overlapping combinations are possible: either the current speaker starts laughing (pattern 1, see Fig. 2 and Table 2) or the co-participant (pattern 2). Likewise, two non-overlapping combinations are theoretically possible when either the current speaker shows some "solo laughter" (pattern 3) or his/ her interlocutor (pattern 4).

1 speaker A ends the turn with a laugh, speaker B starts laughing *after* A's onset of the laugh (= the "invitation-acceptance scheme" proposed by Jefferson (1979)); 
`A: und sich mit den NACHbarn angelegt, he[he]`
`B: [phhh hohoho]`

2 speaker A ends the turn and speaker B starts laughing *before* A's onset of the laugh; 
`A: und sich mit den NACHbarn angelegt, [hehe]`
`B: phhh [hohoho]`

3 speaker A ends the turn with a laugh, no laugh of speaker B overlaps or follows; 
`A: und sich mit den NACHbarn angelegt, hehe`
`B:`

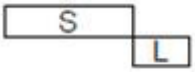
4 speaker A ends the turn without a laugh, but speaker B laughs. 
`S2: und sich mit den NACHbarn angelegt,`
`S1: phhh hohoho`

Table 2: The four theoretically possible patterns are illustrated with the help of the first example in the transcript from the introductory section.

4. Results

4.1. Frequency of occurrence

Fig. 4 shows the absolute number of occurrences of OL and NOL in the four corpora. In one corpus, ICSI, the majority of laughs are OL (65%). For the three other corpora the picture is reversed with OL between 35% (AMI), 38% (Map Task) and 41% (Diapix).

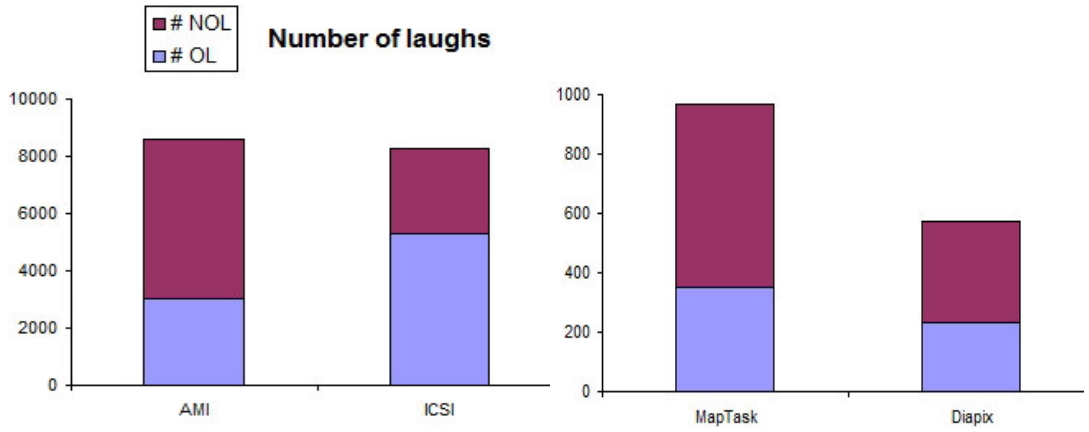


Fig. 4: Absolute number of laughs of the four corpora divided into overlapping (OL) and non-overlapping laughs (NOL). The larger corpora (with more than 8000 tokens) are on the left, the smaller corpora (with more than 500 tokens) on the right.

A look at the relative amount of laughter (related to the total speaking time) reveals differences between the four corpora (Fig. 5). The multi-party conversations show fewer laughs per minute per speaker than the dialogues.

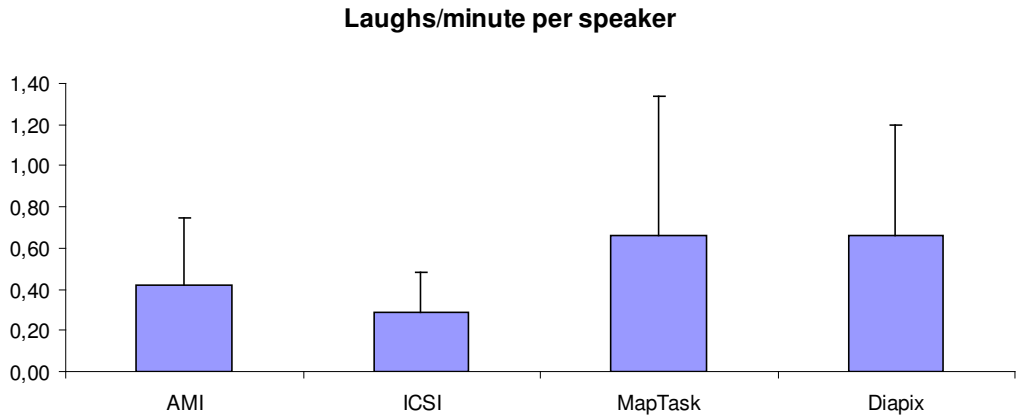


Fig. 5: Laughing rate as the relative number of laughs of the four corpora expressed as an average number of laughs per minute per speaker. Illustrated are the means and standard deviation.

4.2. Prosodic characteristics of overlapping and non-overlapping laughter

The results for the mean duration of OL and NOL are illustrated in Fig. 6. OL are longer in all corpora. This difference in duration is statistically significant based on t-tests for each corpus (for the statistical details the reader is referred to Truong/ Trouvain 2012b). There are also statistically significant differences between the different corpora with ICSI showing the longest durations followed by AMI and both dialogue corpora with the shortest durations.

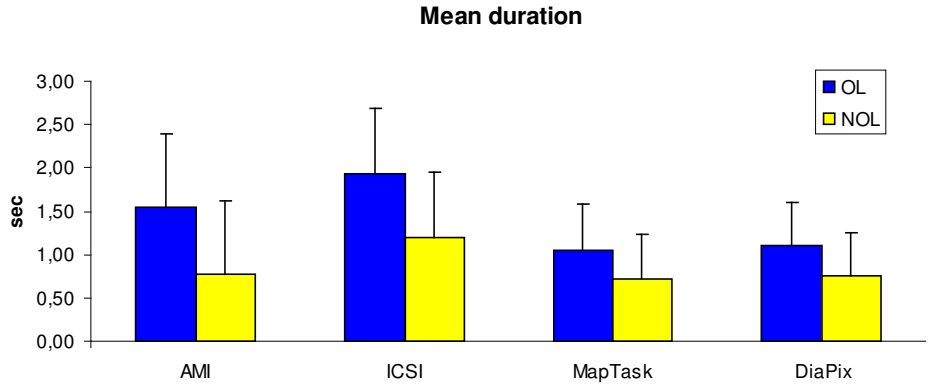


Fig. 6: Values for mean and standard deviation of the duration in the four corpora for overlapping (OL) and non-overlapping laughs (NOL).

The analysis of the fundamental frequency reveals that OL is higher than NOL. The same relationship can be stated for the mean and the maximal intensity as well as for the amount of voiced portions. All differences proved to be statistically significant for all corpora.

When we consider completely unvoiced tokens the number is higher for NOL (see Fig. 7). This finding holds again for all corpora with statistical significance.

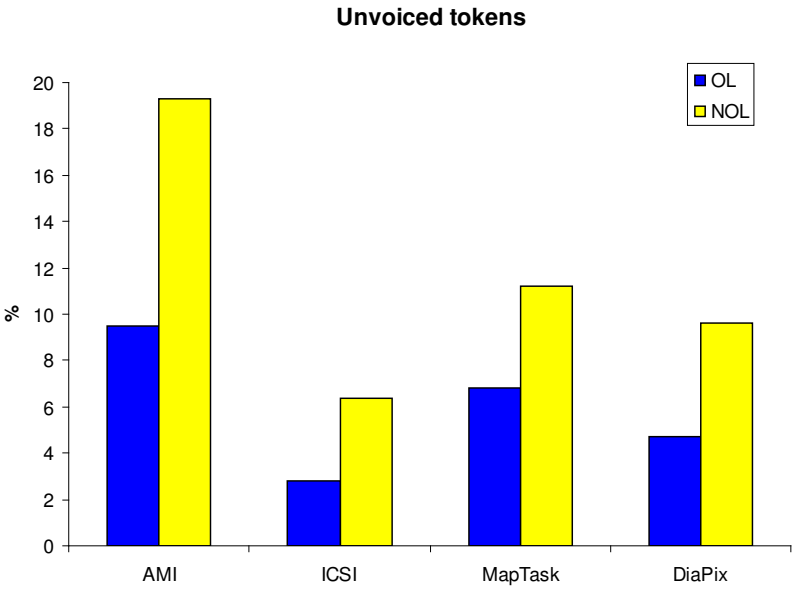


Fig. 7: Percentage of completely unvoiced tokens in the four corpora for overlapping (OL) and non-overlapping laughs (NOL).

To summarise the results, the production of an OL in comparison to a NOL is characterised by a longer duration, a higher fundamental frequency, a higher intensity, a higher amount of voicing and a reduced tendency for complete devoicing. Interestingly, there is statistical evidence for all four acoustic parameters investigated for all four corpora investigated. For this reason we can consider the results as very stable across the different corpora.

4.3. Overlapping laughter as "invitation followed by acceptance"

Table 3 summarises the results regarding the analysis of the timing of laughs with respect to speech in the Map Task corpus. As already seen before, the laughs of the type OL represent the minority of all laughs (38.6%). The two combinations assumed for the analysis of OL (see the first two left-hand rows in Table 3) account for 24.8% of all laughs. Type 1 (speaker invites recipient to laugh) occurs twice as often as type 2 (recipient does not wait with laughing).

Both NOL combinations amount to half of the data (50.8%). Type 3 (speaker laughs without response) and type 4 (only recipient laughs) are balanced.

It must be noted that about one quarter of all laughs could not be assigned unambiguously to one of the four assumed categories. Either the annotated laughs were part of one larger and complex laugh (e.g. with inhalation phases in between which were not annotated as part of the laugh), or there was no speech or other vocalisation immediately preceding, which could be seen as a further category (see Table 3). In eight percent of the cases we were simply unsure as to how to analyse the laughs regarding their timing to speech because we did not see a clear fit in the proposed patterns. All these laughs require a qualitative analysis for further classification.

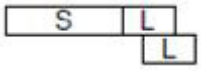
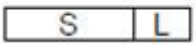
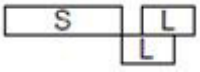
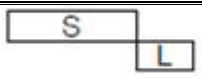
Type	OL	Type	NOL
	16.1%		24.2%
	8.7%		26.6%
'unsure'	3.9%	'unsure'	3.3%
'no speech before'	2.4%	'no speech before'	5.8%
'complex'	7.5%	'complex'	1.5%
total	38.6%	total	61.4%

Table 3: Frequency of speech-laugh patterns observed in the HCRC Map Task corpus (shown in percentages of the total number of laughs). OL= overlapping laugh, NOL= non-overlapping laugh, S=speech, L=laugh (data taken from Trouvain/ Truong 2013).

5. Discussion and conclusions

5.1. Laugh-in-interaction: laughing alone vs. laughing together

The analysis of the annotated laugh tokens in four different corpora shows that a substantial amount of laughter occurs at a time when other interlocutors are laughing as well. The amount of overlapping laughs ranges from one third of all laughs to two thirds. This finding supports the idea that laughter is a joint vocalisation par excellence. The principle "one speaker at a time" which is often assumed to be valid in conversation does obviously not hold for laughing in conversations. Although this is not a new finding, we now have a notion that this is not an infrequent event and can therefore not be considered marginal. Moreover, the results show that types of conversation can heavily vary in their number of OL. This fact has to be taken into account for investigations of laughter in general.

The ICSI corpus, with up to 11 participants in a conversation, shows a remarkably high number of overlapping laughs compared to the other corpora. In this corpus the mean duration of a laugh is substantially longer than in the other corpora. However, each speaker does not show as many laughs per minute as speakers in the other corpora. These observations may reflect the fact that people in a larger communicative community behave differently in situations in which only two persons are engaged. Experiments with the same speakers in dyadic vs. multiparty conversations could shed more light on aspects of how individuals change their laughing behaviour in terms of frequency of occurrence and prosodic parameters.

Overall, the majority of laughs were produced as NOL. One possible explanation for a NOL is that it is a rejected invitation. This would be covered by type 3 of our analysis of the Map Task corpus, which amounts to one third of the entire data. However, a qualitative analysis is needed to decide whether each laugh not responded to by a recipient was really a rejection. The strengths of a quantitative analysis are i) to see how frequent or marginal certain types of laughs occur, ii) what their general acoustic patterns are and iii) to show with the visualisation techniques presented where in the conversation relevant laughs were produced. Thus it can help to find interesting events faster and to gain an overall picture of conversations, in our case regarding laughter.

5.2. Phonetic-prosodic distinction of overlapping and non-overlapping laughter

OL substantially differ from NOL in terms of their prosodic make-up. The distinction between OL and NOL regarding their productions is clear-cut and based on a solid fundament of data. This finding does not exclude that NOL and OL are two separate categories. Moreover, there can be a lot of variation between individuals, text types (or communicative settings) and cultures. However, at least for the English data the *general* pattern of the phonetic-prosodic substance of OL and NOL seems to be clear: The prosodic signaling system of OL compared to NOL uses longer duration, a higher fundamental frequency, a higher intensity and more

voicing. All these acoustic features probably contribute to attract more attention from the listener. In contrast, NOL don't seem to be primarily produced to get attention from the listeners. It thus might be that some laughs occur below our perceptual threshold, especially when overlapped by *speech* of the other – a circumstance which could also affect transcribers of conversational data.

One explanation for this listener-attraction of OL is that laughter is used as a positive social signal to display affiliation and social bonding. OL as a mainly voiced vocalisation with longer duration, higher intensity and higher pitch requires more physiological effort than the virtual default case. In line with this, there are findings that dialogues with voiceless laughs were associated with feelings of social exclusion. Laughing for the purpose of social integration ("laughing *with* somebody") is accomplished best with voiced and prosodically prominent forms of laughter typical of OL. Laughing for the purpose of social segregation ("laughing *at* somebody") seems to work better with unvoiced laughs with a lower physiological effort often found in NOL (Cirillo/ Todt 2005).

5.3. "Laugh to be invited"

The pattern of the "invited laugh" as described by Jefferson (1979) seems to occur substantially more often than the case of the "anticipated laugh", where the recipient laughs before the onset of the "inviting" speaker. Obviously there is a majority of occasions where people show a tendency to "wait" to be invited to a shared laugh rather than anticipating an overlapping laugh. This result could be seen as a type of convention – similar to the widespread convention that somebody entering a room greets first, answered by the person/s already in the room.

An open question remains how laughter is invited. Does in those "invited" cases the laugh of the "inviter" start with the prosodic characteristics of a typical OL, which then signals the recipient to join in or not?

As mentioned above the majority of laughs are non-overlapping. Future studies must reveal the possible communicative functions of these laughs, including rejections of invitations, self-comments (cf. Trouvain 2001) or comments as part of a feedback expression (cf. Ward/ Tsukahara 2000).

5.4. Prosody of laughing beyond transcriptions

We have seen at the beginning of this paper what kind of information is usually recorded in transcripts of conversational analysis. Often these transcripts are based on recordings of only one channel (which can be mono or stereo). Working with signals without a separation of the different speakers carries the risk that acoustic signals in overlapping phases are masked, making the job of human perception of laughs of different persons very hard and sometimes impossible.

A further advantage of separate channels is the possibility to determine the *exact* timing of overlapping vocalisations. This granularity is presumably not relevant for most applications of these transcripts. However, it can play a role for instance in learning more about vocal alignment patterns like the laugh-invitation and who was really first.

The prosodic characteristics of fundamental frequency and intensity have not been studied in conversation analysis so far. Duration is either indicated by the number of laugh "syllables" like [hehehe] or a time span like "1.2 sec". It remains unclear which benefit the indication of the vowel quality bears, laughing with a high front vowel might be different for the interpretation of a communicative situation than laughing vowels that resemble low back vowels. However, acoustic analyses of (many) laugh events show a general tendency that laughter has no clear vocal tract configurations (see e.g. Bachorowski et al. 2001). Instead it could be more helpful for more detailed acoustic analysis to give some indications of the perceived pitch, loudness, duration and also whether the laugh was voiced or voiceless.

5.5. Laughter and interactional prosody

Conversation analysis and interactional linguistics are important research approaches to study the prosodic nature of talk-in-interaction, often ignored by phonetic sciences and speech processing. The analysis of fine phonetic details, be it on an auditive or on an acoustic level, unravel phenomena that are of great interest for phonetic and prosodic research. These phenomena definitively include laughter and other nonverbal vocalisations (see Ogden 2013 for a recent example). Analysis of single cases can help to uncover conversational mechanisms how laughing works in conversations; see, for instance, studies showing that shared laughter is often associated with topic termination (Holt 2010, Shaw et al. 2013). It is extremely important to describe the link between certain pragmatic functions, e.g. topic termination, and prosodic characteristics.

Going beyond singular examples, on the other hand, can help to understand how often and how stable certain features and patterns occur. The latter is what we did with our corpus studies. Another example is the work by Bonin et al. (in press) to investigate in large corpora the aforementioned link between topic changes and laughter. In our view the findings and their interpretations can be useful for interactional phonetics in general and in particular for the prosody of laughing as one of the most frequent nonverbal vocalisations.

6. Acknowledgements

The authors would like to thank the editors of this volume for valuable feedback on an earlier draft of this paper. This work was partly funded by the EU's 7th Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet).

7. References

Anderson, Anne H. et al. (1991): The HCRC Map Task Corpus. In: Language and Speech 34, 351-366.

Bachorowski, Jo-Anne/ Smoski, Moria J./ Owren, Michael (2001): The acoustic features of human laughter. In: Journal of the Acoustical Society of America 110, 1581-1597.

Baker, Rachel/ Hazan, Valerie (2011): DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. In: Behavior Research Methods 43(3), 761-770.

Bonin, Francesca/ Campbell, Nick/ Vogel, Carl (in press). Time for laughter. In: Knowledge-Based Systems. DOI: 10.1016/j.knosys.2014.04.031.

Campbell, Nick (2007a): Whom we laugh with affects how we laugh. In: Proceedings of the Interdisciplinary Workshop on the Phonetics of Laughter Saarbrücken, 61-65.

Campbell, Nick (2007b): Approaches to conversational speech rhythm: speech activity in two-person telephone dialogues. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 343-348.

Carletta, Jean (2007): Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. In: Language Resources and Evaluation 41, 181-190.

Chafe, Wallace (2007): The Importance of Not Being Earnest. Amsterdam: Benjamins.

Cirillo, Jasmin/ Todt, Dietmar (2005): Perception and judgement of whispered vocalisations. In: Behaviour 142, 98-125.

Cummins, Fred (2007): Speech synchronization: Investigating the links between perception and action in speech production. In: Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, 529-532.

Heldner, Matthias/ Edlund, Jens (2010): Pauses, gaps and overlaps in conversations. In: Journal of Phonetics 38, 555-568.

Holt, Elizabeth (2010): The last laugh: shared laughter and topic termination. In: Journal of Pragmatics 42 (6), 1513-1525.

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

Janin, Adam et al. (2003): The ICSI meeting corpus. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 364-367.

Jefferson, Gail (1979): A technique for inviting laughter and its subsequent acceptance/ declination. In: Psathas, George (ed.): Everyday Language: Studies in Ethnomethodology. New York: Irvington, 79-96.

Laskowski, Kornel/ Burger, Susanne (2007): Analysis of the occurrence of laughter in meetings. In: Proceedings of the Interspeech Conference, Antwerp, 1258-1261.

Lerner, Gene H. (1991): On the syntax of sentences in progress. In: Language in Society 20, 441-458.

Liddicoat, Antony J. (2007): An Introduction to Conversation Analysis. New York: Continuum.

Local, John (2005): On the interactional and phonetic design of collaborative completions. In: Hardcastle, William/ Beck, Janet MacKenzie (eds.): A Figure of Speech: a Festschrift for John Laver. New Jersey: Lawrence Erlbaum, 263-282.

Nwokah, Eva E. et al. (1999): The integration of laughter and speech in vocal communication: a dynamic systems perspective. In: Journal of Speech, Language & Hearing Research 42, 880-894.

Ogden, Richard (2013): Forms and functions of clicks in English conversation. In: Journal of the International Phonetic Association 43(3), 299-320.

Sacks, Harvey/ Schegloff, Emanuel A./ Jefferson, Gail (1974): A simplest systematics for the organisation of turn-taking for conversation. In: Language 50, 696-735.

Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 10, 353-402.

Shaw, Chloe/ Hepburn, Alexa/ Potter, Jonathan (2013): Having the last laugh: On post-completion laughter particles. In: Glenn, Phil/ Holt, Elizabeth (eds.): Studies of Laughter in Interaction. London: Bloomsbury, 91-106.

Smoski, Moria J./ Bachorowski, Jo-Anne (2003): Antiphonal laughter in developing friendships. In: Annuals of the New York Academy of Sciences 1000, 300-303.

Stivers, Tanja et al. (2009): Universals and cultural variation in turn-taking in conversation. In: Proceedings of the National Academy of Sciences of the United States of America (PNAS) 106(26), 10587-10592.

Trouvain, Jürgen (2001): Phonetic aspects of "speech-laughs". In: Proceedings of the Conference on Orality & Gestuality (Orage), Aix-en-Provence, 634-639.

Trouvain, Jürgen (2003). Segmenting phonetic units in laughter. In: Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS), Barcelona, 2793-2796.

Trouvain, Jürgen/ Campbell, Nick (eds.) (2007): Proceedings of the Workshop on the Phonetics of Laughter, Saarbrücken (http://www.coli.uni-saarland.de/conf/laughter-07).

Trouvain, Jürgen/ Truong, Khiet P. (2012): Comparing nonverbal vocalisations in conversational speech corpora. In: Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals, Istanbul, 36-39.

Trouvain, Jürgen/ Truong, Khiet P. (2013): Exploring sequences of speech and laughter activity using visualisations of conversations. In: Proceedings of the Workshop on Affective Social Speech Signals (WASSS 2013), Grenoble, 5 pages.

Truong, Khiet P./ Trouvain, Jürgen (2012a): Laughter annotations in conversational speech corpora possibilities and limitations for phonetic analysis. In: Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals, Istanbul, 20-24.

Truong, Khiet P./ Trouvain, Jürgen (2012b): On the acoustics of overlapping laughter in conversational speech. In: Proceedings of the Interspeech Conference, Portland, 4 pages.

Vettin, Julia/ Todt, Dietmar (2004): Laughter in conversation: features of occurrence and acoustic structure. In: Journal of Nonverbal Behaviour 28(2), 93-115.

Wagner, Johannes/ Vöge, Monika (eds.) (2010): Laughter in Interaction. Special Issue of the Journal of Pragmatics 42 (6), 1469-1764.

Ward, Nigel/ Tsukahara, Wataru (2000): Prosodic features which cue back-channel responses in English and Japanese. In: Journal of Pragmatics 32, 1177-1207.

Jürgen Trouvain
Saarland University
Department of Computational Linguistics and Phonetics
Campus C7.2
66123 Saarbrücken
Germany
Trouvain [at] coli.uni-saarland.de

Khiet P. Truong
University of Twente
Human Media Interaction
P.O. Box 217
7500 AE Enschede
The Netherlands
k.p.truong [at] utwente.nl