Conference or Workshop Item:
Trouvain, Jürgen, Ogden, Richard and Cantarutti, Marina (2024) On visualising timing aspects of overlapping laughter in conversational speech. In: Laughter and Other Non-Verbal Vocalisations Workshop 2024, 16-17 Jul 2024.

On visualising timing aspects of overlapping laughter in conversational speech

Jürgen Trouvain[1], *Richard Ogden*[2], *Marina Cantarutti*[2]

[1]Language Science and Technology, Saarland University, Saarbrücken (Germany)
[2]Language and Linguistic Science, University of York, United Kingdom
trouvain@lst.uni-saarland.de, Richard.ogden|marina.cantarutti@york.ac.uk

Abstract

New forms of visualisations can help to draw attention to details of analysis. Here, we suggest new combinations of visual information for a better analysis of timing of elements of joint laughter in conversations.

1. Introduction

Visualisation (and notation) play a central role in the development of analysis. New forms of visualisation can help us draw attention to details, and as well as to overcome blind spots when we have become used to one type. Thus, visualisation can be very a powerful way to help us observe, explore and explain details that might otherwise be overlooked. Laughter in conversation is full of such under-researched details.

In conversation, *joint laughter* is especially common. For a better understanding of how laughter in conversations works, details of its position in the talk of the co-participants and of its phonetic composition are needed. The aim of this study is to present a few ways to visualise recordings of the same shared laughter event, and to illustrate what they show.

1.1. Transcripts

A typical way to visualise speech is a written transcript, where laughter as a non-verbal vocalisation can be notated with either an explicit descriptive mark like "(laughs)", or with orthographic syllables such as "hahaha" or "hihi". The latter has the advantage of indicating the number of laughter pulses, but also bears the disadvantage of ascribing to the laugh a specific vowel quality which probably would not stand up to any formant measurement [1].

In conversation analysis (CA), it is usual to organise the transcripts of the different speakers on separate lines, to capture the temporal details of overlapping speech and other vocalisations [5]. In addition, there is the option to enrich the information of phonetic quantity and quality in different ways: speech-laughs, information about lengthening, or specifications of inhalation or exhalation sounds.

Chafe [3] uses special symbols for exhalatory and inhalatory pulses (or calls), e.g. "freaked out /\ /\ /\ \/" for three exhalation followed by one inhalation particle after the speech.

1.2. Speech signal

In phonetics, a common way to visualise speech and other vocalisations is to display the waveform and the spectrogram of an acoustic signal. Sometimes this display comes with annotation and segmentation that can be arranged on different tiers, e.g. phonetic notation of syllables, words in orthographic notation, inter-pause intervals, and many more. A speech-laugh, for example, could be annotated on two tiers, one for speaking and the other for laughing.

Annotation and segmentation of laughs and their sub-units can be rather challenging when going beyond 'simple' laughs or simplistic views on laughs [9]. Challenges include for instance that different annotators disagree on when a given unit, be it an episode, a call or a bout, starts and ends.

An advantage of the acoustic signal over the transcript is that it provides a detailed view of the *exact* timing of certain properties of a laugh and more information on the fundamental frequency, the intensity, the formants and further spectral characteristics. Those acoustic parameters allow interpretations of the speech production regarding voice quality, the involvement of phonation, and, if of interest, changes in the vocal tract. Moreover, information on (perceived) pitch, loudness and rhythmic properties can be used, i.e. characteristics that are very important for describing and analysing laughs [8].

1.3. Recording quality

A prerequisite for a proper and clean visualisation of overlapping conversational data is to have recordings where the recording channels of the different speakers are separate. Channel separation is needed to avoid acoustic masking. Although in real data with separate microphones acoustic masking cannot entirely avoided, it has a tremendous advantage over single-microphone recordings.

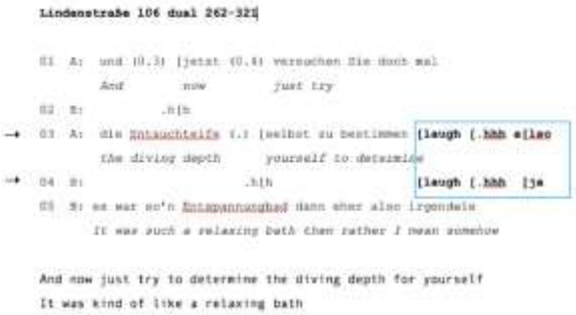
We do not claim that data with masked signals are useless, but they should be interpreted with care. For example, when a speaker joins in with another's laughter, but with a quieter voice than the partner's overlapping laughter, it can be inaudible and invisible in a single-channel speech signal [8].

2. Example visualisations

In contrast to laugh-and-speech activity plots [6], we illustrate here our visualisations with an overlapping 'song-like' laughter by two speakers that also exhibits simultaneous completion. The extract is taken from the Lindenstrasse corpus [4] containing task-based conversations in German. In the selected example two male friends are talking about a certain scene in a soap opera.

2.1. CA transcript

The transcript in Fig. 1 shows the selected extract with some basic detail. Lines 3 and 4 show how precisely coordinated in time A's and B's laughter is: B produces very quiet laughter particles in overlap with A's talk. A then produces post-completion laughter which B joins in with. A and B both finish laughing at the same time. They take an inhalation simultaneously. They then both start talking more or less simultaneously. Even this type of visualisation makes the temporal coordination between speakers visible, while also suggesting details to focus on in what is innocently described as ‘laughter’. How do both speakers coordinate their shift out of laughter so precisely?


Lindenstraße 106 dual 262-321

01 A: und (0.3) [jetzt (0.4) versuchen Sie doch mal
And now just try
02 B: .h[h
→ 03 A: die Eintauchtiefe (.) [selbst zu bestimmen [laugh [.hhh a[lso
the diving depth yourself to determine
→ 04 B: .h[h [laugh [.hhh [ja
05 B: es war so'n Entspannungsbad dann eher also irgendwie
It was such a relaxing bath then rather I mean somehow

And now just try to determine the diving depth for yourself
It was kind of like a relaxing bath

Figure 1: Transcript of the selected extract.

2.2. Speech signal

Phonetic details of the individual laughs of both speakers can be seen in Fig. 2 and Fig. 3, respectively. Both figures are enriched with boxes containing phonetic transcription.

Fig. 2 shows that the exhalation phase of A’s laughter consists of two *phrases*. In each phrase, there are three particles. In the first phrase, A has three pulses, with the rhythm short-short-long. A’s laughter pulses are 180, 210 then 450 ms long, then 220, 230 and 420 ms, i.e. roughly in the ratio [1:1:2] [1:1:2]. This pattern is reminiscent of final lengthening in speech.

B’s laughter particles are also grouped into two phrases of five, with a dip in amplitude in the middle. B’s pulses are all more or less equal in rhythmical weight. It can be seen in Fig. 3 that B’s laughter is not as melodic as A’s.

B’s laughter starts in overlap with A’s talk (cp. Fig. 4). Note that it is not possible to hear this really on listening to the dual track recording; the quiet voiceless ‘snort-like’ laughter is only audible after channel separation.

2.3. Speech signal combined with CA transcript options

In Fig. 4 the laugh events (including some of the before and after context) of both speakers are overlaid together as they occurred in time (taken from [2]. The displayed extract corresponds to the second half of lines 3 and 4 of the transcript (Fig. 1). Fig. 4 shows the waveforms, spectrograms and the F0 contour of both speakers, enriched with the following features: (i) orthographic (for speech) and phonetic notation (for laughing), (ii) display of both waveforms for a better comparison, (iii) coloured highlighting of the laughter bouts, and (iv) durations of each laughter particle.

These visualisations help to make it clear that each speaker’s second phrase of laughter can be interpreted as projecting the end of this shared laughter and help to coordinate the simultaneous inhalations of both speakers, which mark the end point of this laughter bout. What follows is the nearly simultaneous onset of talk.

2.4. Schematic time boxes

Fig. 5 summarises the most important features of the selected example of shared laughter across time in a more 'gestalt' style. The three main stages of the laughter bout are shaded differently. The rhythmical pattern of the exhalation phase in the middle contributes to the projection of the ending. In the second phrase, both participants repeat the rhythm of their respective first phrase. The repetition of the rhythmical phrase provides for and projects the ending of the second phrase.

2.5. F0 traces

A's laughter pulses are voiced, and one of the affordances of voicing is pitch. A narrow band spectrogram and f0 trace are shown in Fig. 6. The three pulses of the first phrase rise in approximately 2.5 semitone steps; on the third, longer, pulse, there is a short high-pitched segment at 312 Hz, 8.8 semitones higher from the preceding low. The second phrase starts in a higher key, about 4 semitones higher than the starting point of the first phrase. The middle pulse ends about 8 semitones higher than the first, but then the final pulse is remarkably like the final pulse of the first phrase. In short, there are rich tonal relations between the pulses in this ‘song-like’ laughter. Speakers modifying the pitch of the second phrase allows for this second ending to be recognised *as* an ending.

3. Discussion

Working with transcripts *and* speech signal analysis allows a fine-grained analysis of details of laughter. By recycling some aspects of the laughter bout, like rhythm, but modifying others, such as pitch, participants can project an ‘ending’ in song-like laughter in much the way that happens in actual songs.

In the example we have shown, the simultaneous ending of laughter and onset of talk need explanation. Schegloff [6] proposes that turns at talk have structurally distinct and recognisable (to participants) ‘beginnings’, ‘middles’ and ‘endings’. While laughter lacks linguistic structure, it does have a structure: an initiating pulse, an exhalation phase, and, commonly, an ending with inhalation [3, 9] as visible in Fig. 4. The ‘middle’ of a bout is often the pulsed spasmodic expulsion of air [3]. The pulses of laughter allow the middle to be recognised as such, i.e. as something which structurally comes before an ‘ending’. One of the affordances of this internal structure is that participants can recognise or ‘parse’ where they are in the laughter bout, and – as here – project its possible ending.

4. Conclusions

We have shown that different ways of visualising audio-recorded laughter in conversation provide different opportunities for thinking about its placement, organisation and structure. Whereas signal-based visualisations provide a lot of detail on a small-scale window of inspection, a conversation analytic-style transcription provides a much wider window, and situates laughter in its conversational context. Schematic representations allow a visualisation of the components of a laughter bout and their structuring, including how they relate to each other and to those produced by other speakers in overlap

Laughter has many phonetic affordances, including pulsing, phrasing, voicing, loudness and pitch. This phonetic richness of laughter is an understudied resource in interaction. The internal structure of laughter bouts provides participants with a structure that can be recognised and used to coordinate their laughter in time, as well as their transition out of laughter into the next activity, such as speaking. By visualising this structure, we may start to understand both its general shapes and particular instantiations on occasions of use, and so account for some of the phonetic richness of laughter.

Acknowledgements: Thanks to Pavel Šturm for his creativity on some earlier visualisations.

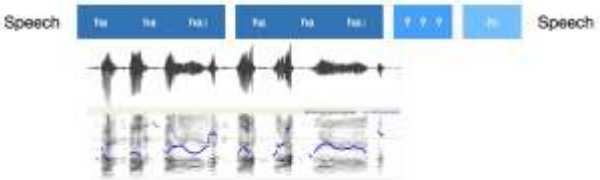


Figure 2: Transcript and speech signal of A's laughter.

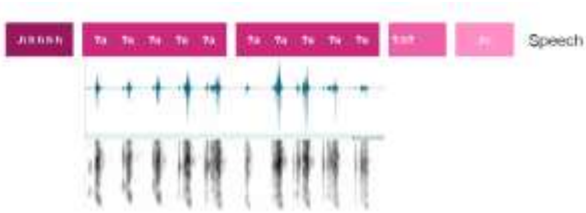


Figure 3: Transcript and speech signal of B's laughter.

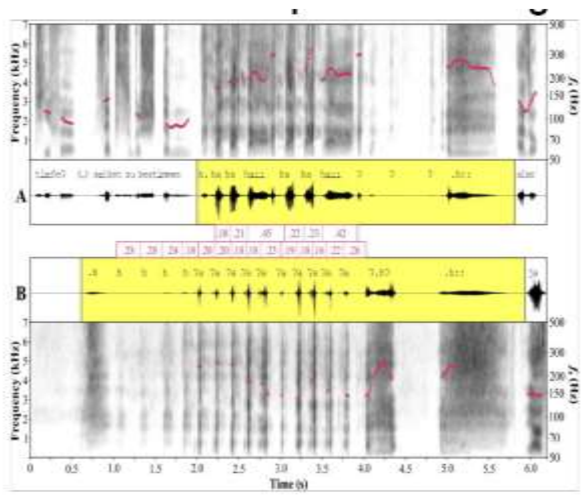


Figure 4: Speech signals of both speakers combined with transcript (taken from [2]).

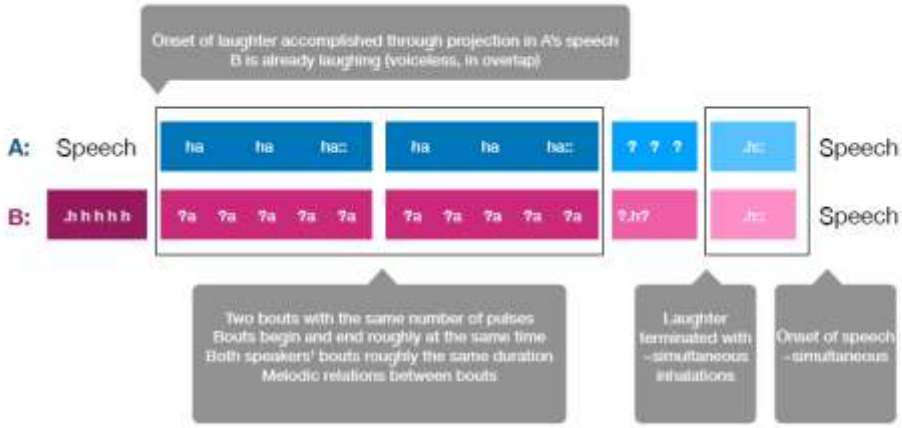


Figure 5: Schematic time box.

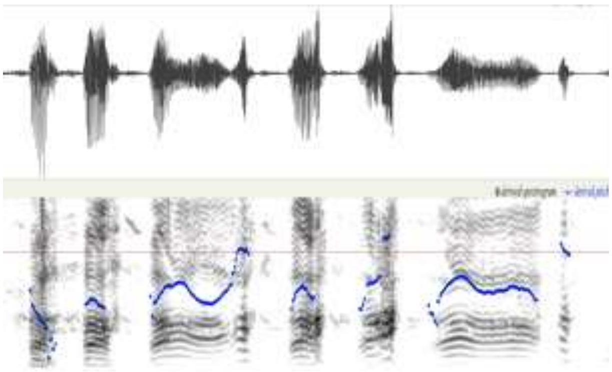


Figure 6: Narrow band spectrogram (0.02 s window); pitch range 100-500 Hz, plotted at semitones re:100 Hz; raw cross-correlation.

5. References

[1] Bachorowski, J.-A., Smoski, M. J., & Owren, M. J. (2001). The acoustic features of human laughter. The Journal of the Acoustical Society of America, 110(3), 1581–1597.

[2] Cantarutti, M., Ogden, R., Šturm, P. & Trouvain, J. 2023. Laughing in interaction: How phonetic details can coordinate action sequences. Proc. Disfluency in Spontaneous Speech (DiSS) Workshop, Bielefeld, pp. 38-42.

[3] Chafe, W. 2007. The Importance of Not Being Earnest. The Feeling behind Laughter and Humor. Benjamins.

[4] IPDS 2006. Video Task Scenario: Lindenstraße – The Kiel Corpus of Spontaneous Speech,”. University of Kiel.

[5] Jefferson, G. 2004. Glossary of transcript symbols with an introduction. In: Lerner, G. (ed.), Conversation Analysis: Studies from the first generation. Amsterdam, pp. 13–31.

[6] Schegloff, E. 1996. Turn organisation: one inteesection of action and grammar. In: Ochs, E, Schegloff, E.A. * Thompson, S.A. Interaction and Grammar. Cambridge, pp. 52-133.

[7] Trouvain, J. & Truong, K.P. 2013. Exploring sequences of speech and laughter activity using visualisations of conversations. Proc. Workshop on Affective Social Speech Signals, Grenoble.

[8] Trouvain, J. & Truong, K. 2014. Towards unravelling prosodic characteristics of speaker-overlapping laughing in conversational speech corpora. In: Barth-Weingarten, D. & Szczepek Reed, B. (eds). Prosodie und Phonetik in der Interaktion – Prosody and phonetics in interaction. Mannheim, pp. 115-135.

[9] Truong, K.P., Trouvain, J. & Jansen, M.-P. 2019. Towards an annotation scheme for complex laughter in speech corpora. Proc. Interspeech, Graz, pp. 529-533.