‘SPEECH-SMILE’, ‘SPEECH-LAUGH’, ‘LAUGHTER’ AND THEIR SEQUENCING IN DIALOGIC INTERACTION

Klaus J. Kohler

Institute of Phonetics and Digital Speech Processing (IPDS), University of Kiel, Kiel, Germany
kjk AT ipds.uni-kiel.de

ABSTRACT

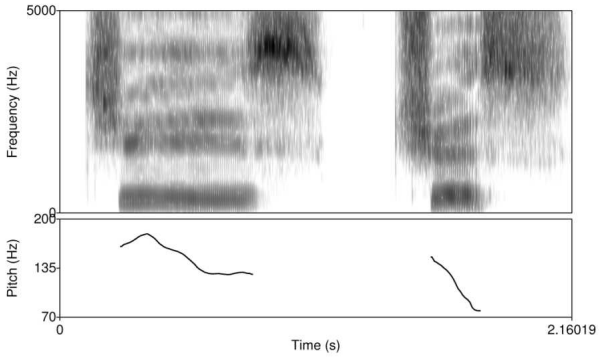
Laughing is examined auditorily and acoustico-graphically, on the basis of exemplary speech data from spontaneous German dialogues, as pulmonic air stream modulation for communicative functions, paying attention to fine phonetic detail.

1. INTRODUCTION

Phonetic analyses of laughing are relatively rare. Existing classification [2,11,12] differentiates *free laughter* and speech-synchronous forms that may be either *speech-laugh* [6], which is characterized by typical sequential laughter attributes coocurring with speech articulation, or *speech-smile*, where f0 is raised and long-term articulatory prosodies of, e.g., lip-spreading and palatalization are superimposed on speech events. *Speech-smile* is likely to be preceded and followed by the facial expression of a smile without sound, speech or paralinguistic vocalization, and is then only visible. This paper focuses on audio aspects of smiling only. For an analysis of smiling in the wider sense it would be necessary to have video recordings, which are not part of the database under discussion.

There has been controversial discussion as to whether speech-laugh and speech-smile form a continuum [12]. Sequences of speech-smile – speech-laugh (– free laughter) and the reverse order are attested in spontaneous speech corpora. But irrespective of this syntagmatic binding, these types of laughing phenomena are different production categories, and therefore they do not vary along one articulatory scale; they are also perceived as belonging to different categories and have different communicative functions. Free laughter and speech-laugh appear to be the expression of *amusement* and *high-key hilarity*, no matter whether real, acted, faked or ironical, although the manifestations may all be different because they can be recognised as such. Speech-smile, on the other hand, is more likely to be a signal of *happiness* and *low-key joy*. It may lead or trail the former, or it may stand on its own, for example as an expression of friendliness. A customer-friendly greeting *“tschüss”* (“bye”) at a shop cash desk may be spoken with a smile as in fig.1 and *audio_file_1.wav*, strengthening high frequencies in spectrum and f0.

Figure 1: *“tschüss”* [tʃyːs] smiling (left), [tʃʏs] neutral (right): spectrogram and f0 trace.



In the literature [2,10,11], the following parameters are listed for the phonetic categorization of different types of laughter:

- voicing, open mouth
- voicing, closed mouth, nasal exit
- voiceless nasal-cavity turbulence
- voiceless laryngeal and oral cavities turbulence
- vowel resonance: close – open, front – back
- pitch of voiced laugh bursts
- number of laugh bursts in a “bout” of laughter
- durations of laugh bursts and burst intervals.
- initial and final inhalations and exhalations are not always included in the analysis of bouts although they are important in the control of breathing for laughing.

Speaking has been described as modified (pulmonic exhalatory) breathing [1], involving complex supraglottal articulation as well as special subglottal pressure settings [5], peripherally supplemented by inhalatory pulmonic and other air stream meachanisms [9]. Pike [9] defined laughter as spasmodic articulation, produced by sudden movements beyond the control of the individual, in the same set as cough, sneeze, hiccough, or belching. This characterization is hardly adequate be-

cause laughter is controlled according to the speech function it is to subserve, even if it may be impressionistically described as spasmodic.

Laughing should be seen as another way of modifying breathing, largely exhalation, but involving inhalation as well, more so than in speaking. Research into the phonetics and communicative functions of laughing needs to analyse breath control (air stream direction, energy time course) as the basic phonetic element, which is modified in fairly simple ways glottally (vibrating/open/closed) and supraglottally (oral/nasal cavity, roughly positioned pharyngeal/oral/labial constrictions).

This is the reversal of the fine-grained phonation and supraglottal articulation superimposed on the relatively simple subglottal setting for speech, which is only modified under special circumstances such as a 'force accent' [3]. In other words, laughing should be analysed as modified breathing in its own right in parallel to speaking, rather than trying to apply speech categories, such as the distinctive vowels and consonants, or CV syllables, of a particular language, while neglecting the fundamental air stream control. Such an independent analysis of laughing will show up the correspondences and divergencies between the two ways of controlling the pulmonic air stream, and will then allow us to make insightful inferences as to how the two are combinable in speech-laughs.

Vocal tract resonances are no doubt important in colouring laughter in various ways for functional purposes, and will therefore need acoustic investigation, but these vocalic qualities are different, phonetically and functionally, from the vowel distinctions in the phonological system of the language. The latter are more numerous and more complex in distinctive feature composition, and serve the differentation of words. The resonances in laughter do not coincide with these qualities in phonological vowel systems and are more elementary, such as 'highish in the front region' vs. 'lowish or rounded in the back region' vs. 'central up-down and front-back' Their function is semantic and pragmatic to distinguish, e.g., 'giggle' and 'chuckle'. This is the same type of difference as vocalic qualities of hesitations, which do not coincide with phonetic ranges in phonological vowel systems either [7]. It is vocal tract resonance as a suprasegmental carrier as against a local segmental differentiator.

Of course, spontaneous laughter cannot be investigated in physiological laboratory experiments, so the direct observation of subglottal activity in laughing is precluded. Thus we have to deduce production, to a certain extent, from the acoustic result, and to this end, need to analyse the acoustic signal, coupled with acute auditory observation, in fine phonetic detail. Up to now, phonetic analysis of laughter has relied on rough acoustic measures, has hardly included the acoustic consequences of air stream direction, dynamics and timing, and has applied descriptive and inferential statistics too quickly to roughly analysed and tagged corpora on the basis of acoustic properties without due consideration of communicative function. The question of synthesizing laughter to make synthetic speech more natural-sounding is totally premature. What we need are studies of fine phonetic detail in dialogic interaction on the basis of exemplary spontaneous speech data. This paper provides a few results of such a (quite limited) investigation. Its aim is programmatic rather than a comprehensive descriptive account of a large database.

2. DATABASE AND METHOD

Two data sources have been used.

- A stereo recording of a dialogue session, consisting of 6 sub-dialogues, between two female speakers (institute secretaries), recorded in the Appointment-making Scenario with overlap [4], labelled but so far not published: **f06**.
- A stereo recording of two male speakers from the Video Task Scenario LINDENSTRASSE [4,8], **l06**, talking about differences in video clips presented to them separately.

In both cases, the speakers knew each other well, and they showed a high degree of spontaneity and naturalness.

In **f061**, speakers jm and mg have to arrange two 2-day business meetings in a 2-month period but cannot find mutually suitable dates because the experimenter inadvertently filled their respective calendars in such at way that there are not enough successive daily slots for both. The only solution mg can suggest is to have the two meetings immediately following each other, turning the two 2-day into one 4-day meeting. jm considers it a possibility but not an appropriate one. She finds this clash between the non-solvable task in the appointment-making game and the hardly conducive adjustment amusing, which she expresses by speech-laugh followed by subdued laughing. It is commented on by mg, with speech-laugh and laughter, as not being important in this kind of appoinment-making. mg's

amusement is in turn picked up by jm, leading to several laugh exchanges between the two speakers.

At the beginning of **l06**, speaker mpi sets the theme of "the utter stupidity" of the German TV soap series LINDENSTRASSE, capping his appraisal with hilarious laughter. The whole dialogue then revolves round exchanges on this theme between speakers mpi and tra about the episodes they have been presented with separately. This leads to several exchanges of laughing.

The recordings were listened to, and the laugh sections excerpted, in *cool edit*. The excerpted sections were then acoustically processed (spectrogram, f0, energy) in *praat* and descriptively analysed by close auditory and visual inspection.

3. RESULTS

3.1. Sequencing of 'speech-smile', 'speech-laugh', and 'laughter'

Any sequencing of the three laughing phenomena is possible. In an elaborate form, a speech-smile can develop into a speech-laugh and in turn into laughter, or contrariwise laughter continues into speech and then trails off as a speech-smile.

Figure 2: f061_jm *"glücklich"* + subdued laughter spectrogram, f0 (plain), energy (dotted).

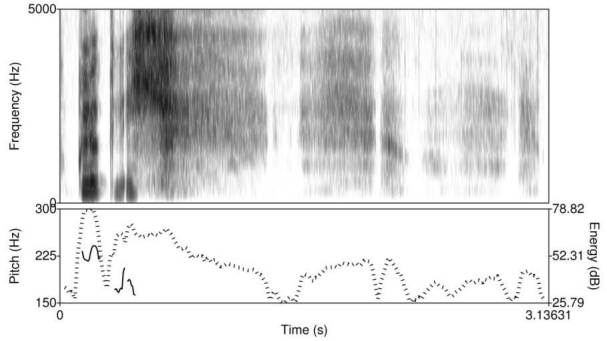


Figure 2 and *audio_file_2.wav* provide an example of the former order by speaker jm. Here *"glücklich"* ("suitable") is preceded by *"nicht so"* ("not so") and a breath intake, and shows a strong energy increase in the accented vowel as well as an f0 rise, indicating low-key joy over the clash between task and executability. This is followed by a renewed strong energy increase in the second syllable *"-lich"*, accompanied by a phonation (and f0) tremor, signalling incoming laughter. There is then long voiceless oral exhalation with high front vowel colouring before two oral cycles of voiceless long breathing in and short breathing out, energy decreasing progressively. This indicates subdued laughter, terminating the laugh section before resuming normal articulation for *"gut. also machen wir das so."* ("good. let's do it that way.").

Figure 3a: f061_mg *"ja"* + laughter + *"ja, gut"* + laughter: spectrogram, f0 (plain), energy (dotted).

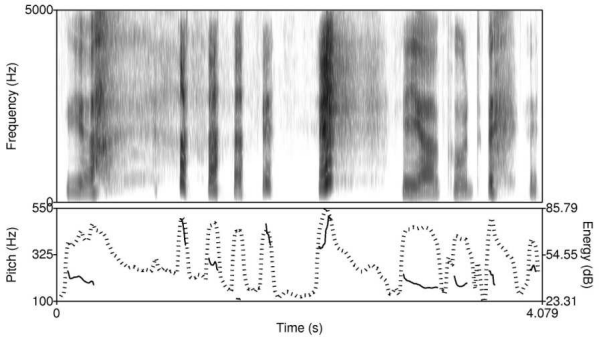


Figure 3b: f061_mg *"ja, gut"* vs. f065_mg*"na gut"* spectrogram and f0 trace.

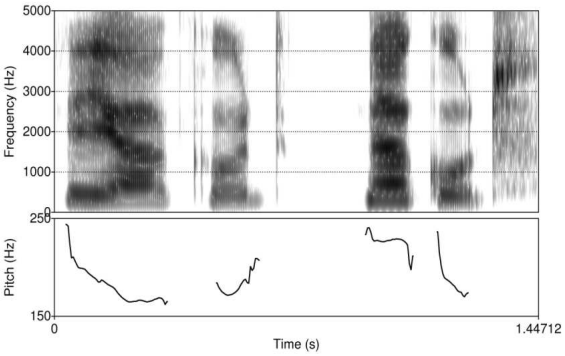


Figure 3a and *audio_file_3a.wav* provide an example of a more complex sequencing by speaker mg. *"ja."* ("yes") shows energy modulation: the energy rises into the vowel, then falls, and rises again, dividing the vowel into two. This is followed by strong voiceless exhalation, with open vowel resonance, setting in at the energy level of the sonorant vowel and trailing off over 400 ms. There are then, embedded in continuing exhalation, 4 voiced energy bursts of approximately the same duration, 70-80 ms, and of the same abrupt rise-fall of energy, evenly spaced, 140-170 ms, creating a very rhythmical pattern of strong laughter. The first 3 bursts have half-open central vowel resonance and descending f0, the 4th has high-front vowel resonance and a high upward f0 jump. The sequence is followed by a 400 ms pause and another, longer voiced energy burst, 120 ms, on an ingressive air stream, with an abrupt rise to a level 14 dB above the previous 4 bursts, accompanied by an abrupt f0 rise from 355 Hz to 512 Hz. The vowel resonance is less high and less front than the preceding burst.

This terminates the laughter and is followed by voiceless exhalation, which turns into speech with

a smile: *"ja, gut."* ("well, all right") has fronting throughout, most obvious in [u:] and [t]. The speech-smile then develops into a concluding short laugh with which the speaker hands over her turn: it consists of a short voice burst, trailing off in strong voiceless exhalation, followed by another much weaker voice burst, all of mid-central vowel resonance. This concluding laugh lacks the rhythmicity of a laughter pattern.

Figure 3b and *audio_file_3b.wav* compare mg's phrases *ja, gut,* of Figure 3a, and *na gut*, with and without a speech-smile, respectively. In the speech-smile, F1 and F2 of [u:] are raised, [t] has a higher locus frequency as well as an energy concentration of the burst at this raised locus, and f0 rises, thus high frequencies are strengthened.

Similar variation in the sequencing of the three types of laugh phenomena is found in the dialogue **l06** of the two male speakers (cf. 3.5 for examples and further discussion).

3.2. Air stream direction

As illustrated in 3.1, voiced and voiceless breathing occurs both egressively and ingressively in laugh turns. This is also found in **l06.** Final exhalation or inhalation should thus be treated as part of such turns and not be ignored in the analysis.

3.3. Oral and nasal air streams

Figure 4 and *audio_file_4.wav* illustrate nasal and oral air streams in laughs of f061_jm. After the utterance *"ja, dann hätten wir's, ne."* ("well, that's it, isn't it"), there is oral + nasal exhalation, which is followed by an oral closure. In turn, a nasal air stream is modulated, first by strengthening – weakening, then by weak glottal vibration, then by a strong voice burst, followed by a weaker one, of [m] colouring, and finally by mouth opening and a voice burst with schwa resonance. This results in a double iambic pattern, each with ascending pitch, a different rhythmicity from the one discussed in 3.1.

Laughing on a nasal air stream conveys somewhat subdued hilarity, a chuckle. In the present case, it occurs in preparation of an unrestrained oral laugh together with the dialogue partner.

Speaker mpi of **l06** also shows this difference between unrestrained and restrained laughter in figures 5 and 6 and in *audio_file_5-6.wav*.

The unrestrained laughter occurs right at the beginning of the dialogue after mpi has emphatically stressed the utter stupidity of the TV series, by saying *"ich hatte schon 'n bisschen vergessen, wie extrem unglaublich schwachsinnig die LINDENSTRASSE ist."* ("It had already somewhat slipped my mind how extremely unbelievably idiotic LINDENSTRASSE is."). He gives *"schwachsinnig"* a force accent [3]. (He is the speaker who has the highest number of force accents for negative emphasis in the whole corpus.) He then highlights his own characterization of the series by unrestrained laughter with a wide-open mouth.

Later-on, he reports on scenes that were presented to him in his video clip and refers to one by the non-word *"didelidu"*, which he again finds hilarious but is less emphatic about, so restrains his laughter to a chuckle by closing his mouth and modulating a nasal air stream.

Figure 4: f061_jm *"ne"* + nasal + oral laugh, voiceless, voiced: spectrogram, f0 (plain), energy (dotted)

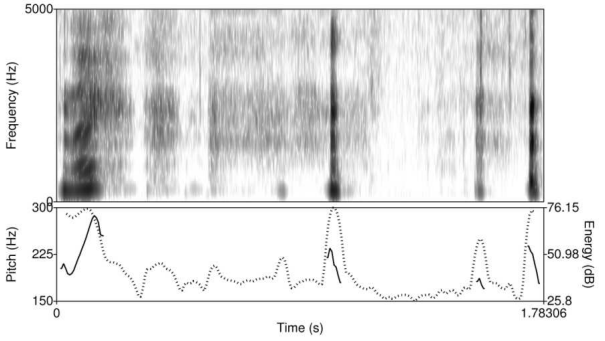


Figure 5: l06_mpi oral laugh, voiceless, voiced: spectrogram, f0 (plain), energy (dotted)

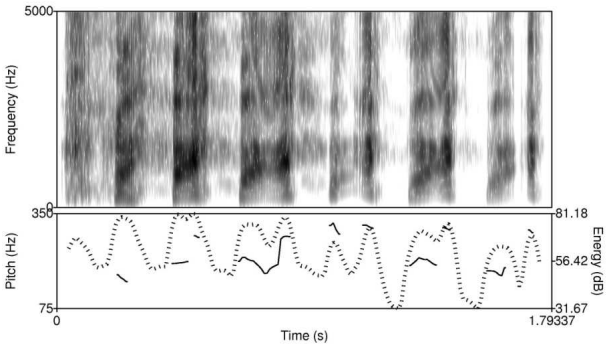
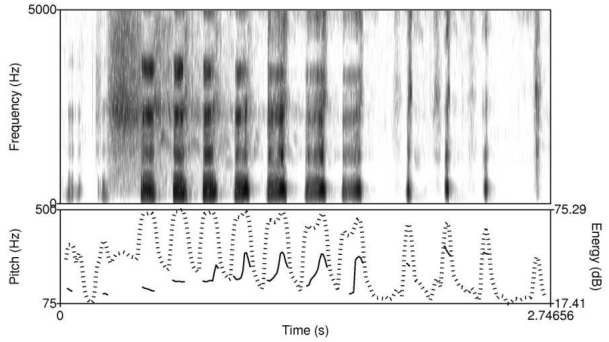


Figure 6: l06_mpi nasal laugh, voiceless, voiced: spectrogram, f0 (plain), energy (dotted)



3.4. Rhythmicity

The nasal or oral air stream is modulated by alternating phonation from glottal opening to various types of vibration and vice versa, and by imposing a dynamic and a duration structure especially on the voiced sections in a sequence. These modulations create rhythmic patterns. In fig. 3, we have seen a sequence of equidistant and equiprominent voice bursts, in fig. 4 an iambic pattern. In fig. 5, the rhythm is even more finely structured: an upbeat of 2 short voice bursts is followed by 4 longer double-peaked ones, grouped in twos of strong – weak energy, i.e. a trochaic pattern, which is clearly perceived in the signal.

In fig. 6, the first 4 voice bursts are evenly spaced and of equal energy on an ascending pitch scale. The next 3 form another block of still evenly spaced but longer bursts, of which the first 2 have well developed f0 upglides (perceivable as such), whereas in the third f0 jumps up abruptly from a low creaky section muffling the rising pitch movement. This together with the decreasing energy in this block creates a dactylic pattern. Then follows a third block of quite short and weaker voice bursts on a high rising pitch level, still evenly spaced.

3.5. Laughing interaction in dialogue

Laughing phenomena are not only sequenced and timed carefully within one speaker according to the communicative functions they are to fulfil but also as part of the interaction with another speaker in dialogue. In **f061**, mg makes an isolated laugh burst just before jm's utterance of fig. 2, seeing the funny side of the clash between task and execution, which jm is leading to. During jm's subdued laughter section in fig. 2, mg produces laughter, followed by speech-smile and then by speech-laugh on the utterance *"ich glaub, darum geht es hier nicht."* ("I don't think that's an issue here."), finally turning into laughter. Then jm agrees *"dann machen wir daraus ein viertägiges."* ("In that case we turn it into a 4-day meeting."), ending in a smile, followed by laughter, during which mg joins in with the speech-laugh and laughter of fig. 3. Towards the end of the latter, jm says *"gut. also machen wir das so."*, on which mg comments with a smiling *"ja, gut."* ("All right."). Then both speakers join in laughter finishing off the dialogue. The two speakers' laughing is coordinated action as part of their joint task-solving in the Appointment-making Scenario. This is illustrated by the complete *audio_files_7-jm* and *7-mg*, which are the two recorded channels of the dialogic interaction. They may be listened to, separately or conjointly, by opening them, for instance, in *cool edit.*

In the Video Task Scenario of **l06**, the situation is quite different. The speakers are not engaged in a joint task-solving goal, but simply talk about the differences they have observed in their respective video clips of the TV series. Speaker mpi sets the theme of emphatic evaluation which he embellishes with amusing wordings accompanied or followed by hilarious laughter. In this, he dominates the dialogue and stimulates speaker tra into laughing, which is, however, never so uproarious as his own.

The *audio_files_8-tra* and *8-mpi* provide examples of the two speakers tra and mpi mutually triggering laughter by facetious descriptions of the soap opera excerpts they have seen. tra gives an account of the scene where a gay chap makes advances to the Turkish doctor, and calls it *ein bisschen anbaggern* "a little digging", which sends mpi off into uproarious laughter, partially overlapping tra's continuation of his story, and his summarising comment: *du bist doch auch 'n kleines bisschen schwul, und so* "you are a little bit gay yourself, aren't you, and that jazz". *und so* is said tongue in cheek with a speech-smile running right through it. This gets mpi into a hilarious mood again because he did not have this episode in his video clip, and refers to the person who spliced the videos as withholding the juicy scene from him. He produces the long utterance *w<Z>as? solche Szenen hat <P> Benno mir gar nicht gezeigt* with a speech-smile throughout. It sends tra into hilarious laughter partially overlapping mpi's turn. So, here we have instances of a speech-smile developing into laughter, and vice versa, within one speaker, and laughing phenomena controlling the interaction between speakers.

In the following orthographic transcripts of these interchanges between jm and mg in **f061** and between tra and mpi in **l06**, sequential turns, with the speakers' IDs, are numbered from 01. Partial overlap is symbolized by giving the two speakers' turns the same sequential number. Overlays of speech-smile and speech-laugh are annotated by enclosing the stretch of overlaid speech in <: :>; <P>, <A> = pause, breathing, <Z> = hesitation.

f061 *audio_files_7-jm* and *7-mg*

jm01 und im Dezem=/- <A>
mg01 ja, wenn wir den zweiten und dritten Dezember <P>
mg02 genommen hatten,
jm02 ja, das würde gehen. <P>
mg03 und dann müssen wir eben dann<Z> daraus 'nen viertägigen<Z> Verabredung machen. <P>
jm04 das wären ja zweimal zwei Tage hintereinander. <A>
mg05 ja, wenn wir keinen anderen Termin finden. <P>
jm06 ja, das würde natürlich gehen. aber das ist bestimmt nicht so<Z>
mg07 <laughter>
jm07 <:<speech-laugh>glücklich:> <laughter>. <A>
mg08 <:<speech-smile> ich glaub', darum geht es hier:> <:<speech-laugh> nicht:>
mg09 <laughter>
jm09 gut, also <P> machen wir das so. <A>
mg10 ja. <A>
jm11 denn machen wir daraus ein <:<speech-smile>Viertägiges:>.
jm12 <laughter>
mg12 <:<speech-laugh> ja:> <laughter>.
jm13 <A> ja, dann hätten wir 's, nicht?
mg14 <:<speech-smile> ja, gut:> <laughter>.
jm15 <laughter>
mg15 <laughter>

l06 *audio_files_8-tra* and *8-mpi*

tra01 das ist +/der<Z>/+ der Türke, der auch +/in der/+ in der<Z>/+ in dem Haus wohnt, <A> dieser türkische Doktor.
mpi02 <äh>
tra03 <A> dass er den irgendwie anbaggert?
mpi04 <äh> null Komma null gesehen.
tra05 <A> mhm, das war nämlich irgendwie die zweite Partyszene, die bei mir irgendwann auftauchte so, wo +/v=/+ <äh> <P> der ihn so 'n bisschen anbaggert
mpi06 <laughter> <:<speech-laugh> ah, ej:>
tra06 und meint +/du bist/+ <A> du bist doch auch ‘n kleines bisschen schwul <:<speech-smile> und so:> +/und/+ und <häs>
mpi07 <:<speech-smile> w<Z>as? solche Szenen hat <P> Benno mir gar nicht gezeigt:>, Alter, hat er mir vorenthalten.
tra07 <laughter>
tra08 ja, woraufhin der Türke natürlich die Party irgendwie beleidigt verlässt.

4. OUTLOOK

Starting from sampled instances of three types of laughing phenomena – laughter, speech-laugh, and speech-smile – this paper has looked at their phonetic patterning and communicative function in interactive speech processes, considering laughter pulmonic air stream modulation in its own right, in alternation with, or superimposed on, the air stream modulation in the speech channel. Even such a small database suggests that fine phonetic detail of laughing is highly structured in its link with dialogic interaction. Its fine-grained analysis in instances of communicative settings can provide insights into the acoustic make-up, including rhythmic factors, as well as into the pragmatics of laughing in interaction. The auditory and acoustico-graphic approach advocated here needs to be extended to a much broader database of spontaneous speech from various scenarios. The investigation will also have to consider to what extent the phenomena are determined by the language and by the individual speaker.

5. REFERENCES

[1] Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

[2] Bachorowski, J.-A., Smoski, M. J., Owren, M. J. 2001. The acoustic feature of human laughter. *J. Acoust. Soc. Am.* 110, 1581-1591.

[3] Kohler, K. J. 2005. Form and function of non-pitch accents. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* 35a, 97-123.

[4] Kohler, K. J., Peters, B., Scheffers, M. 2006. The Kiel Corpus of Spontaneous Speech. Vol. IV. German: VIDEO TASK SCENARIO (Kiel-DVD #1). pdf file in *Sound Patterns of German Spontaneous Speech*: www.ipds.uni-kiel.de/kjk/forschung/lautmuster.en.html

[5] Ladefoged, P. 1967. *Three Areas of Experimental Phonetics*. London: Oxford University Press.

[6] Nwokah, E.E., Hsu, H.-C., Davies, P. & Fogel, A. 1999. The integration of laughter and speech in vocal communication: a dynamic systems perspective. *J of Speech, Lang & Hearing Res*, 42, 880-894.

[7] Pätzold, M. and Simpson, A. P. 1995. An acoustic analysis of hesitation particles in German. In *Proc. XIIIth ICPhS*, Stockholm, 512-515.

[8] Peters, B. 2001. 'Video Task' oder 'Daily Soap Szenario' - Ein neues Verfahren zur kontrollierten Elizitation von Spontansprache. *Manuscript.* http://www.ipds.uni-kiel.de/pub_exx/bp2001_1/Linda21.html

[9] Pike, K. L. 1943. *Phonetics.* Ann Arbor: The University of Michigan Press.

[10] Sundaram, S., Narayanan, S. 2007. Automatic acoustic synthesis of human-like laughter. *J. Acoust. Soc. Am.* 121, 527-535.

[11] Trouvain, J. 2003. Segmenting phonetic units in laughter. *Proc. XVth ICPhS* Barcelona, 2793-2796.

[12] Trouvain, J. 2001. Phonetic aspects of “speech-laughs”. *Proc. 2nd Conference on Orality & Gestuality (ORAGE)* Aix-en-Provence, 634-639.