A phonetic view on annotating speech pauses and pause-internal phonetic particles

Jürgen Trouvain & Raphael Werner

Abstract: This paper aims to give an overview on the complexities of pauses and their phonetic components. After a brief presentation of the main functions of pauses in speech, the major challenges for the definition, the detection, segmentation and annotation of speech pauses are illustrated with speech signals from conversational corpora. The focus of the phonetic particles in so-called "silent" pauses is on breath noises.

Keywords: pause, prosodic breaks, silence, breath noises, filler particles

1 Introduction

The aim of this paper is to provide an overview on how speech pauses and the phonetic particles within speech pauses can be annotated and described. The most important concepts of pauses are discussed and illustrated by numerous samples taken from different speech corpora. Rather than a research study, this paper is meant to show and discuss some fundamental issues on pauses and pause-internal phonetic particles. It may form the base for a phonetically oriented guideline for the annotation of pauses in data collections.

The structure of this article is as follows: after a brief overview of various functions of pauses (section 2) we distinguish in section 3 between four different domains of pauses: i) articulatory pauses, ii) mute phases as listeners, iii) gaps at turn taking, and iv) typical speech pauses within turns. Section 4 summarises approaches of the transcription of pauses whereas the focus of section 5 is on the annotation and segmentation of pauses in acoustic signals. The subsequent sections treat important aspects for pauses in general: the distinction of silent vs. filled pauses in spontaneous speech (section 6), the acoustic breath information in many pauses (section 7), the consideration of pause-internal phonetic particles beyond silent phases (section 8), and perceptual aspects of pauses that are often exclusively regarded from a speech

production perspective (section 9). In the final section we draw some conclusions and give an outlook on important research questions and matters of annotation not treated here.

2 Functions of pauses

The temporal structure of speech is determined – among other timing parameters – by pauses. They are an essential part of speech production, because it is impossible not to pause after a given time of speaking. Likewise, pauses are an essential part of speech perception, because listening to a longer stretch of speech without any pauses would give the listener a very hard time for processing.

Pauses can reflect many different functions only three of which are listed here. Speech pauses often reflect *syntactic-prosodic boundaries* that delimit prosodic phrases. In the framework of *Prosodic Hierarchy*, prosodic breaks (or boundaries) reflect (morpho-)syntactic boundaries (cf. Shattuck-Hufnagel & Turk 1996). Higher-level syntactic boundaries are linked to stronger prosodic breaks, whereas lower-level syntactic boundaries are linked to weaker prosodic breaks. While the hierarchical structure of syntactic boundaries remains the same in a given text, the hierarchical structure of prosodic breaks may change. This optionality is also true for the phonetic marking of prosodic breaks, for instance by producing pauses, especially breath pauses which motivated the use of the term breath-groups for prosodic phrases (Lieberman 1967) five decades ago. The idea of a hierarchy of prosodic break strength is for instance taken up in annotation approaches like *Tone and Break Indices* (ToBI) where the break indices range from 1 (weakest) to 6 (strongest) (Beckman & Ayers 1994). However, most applications of ToBI seem to be restricted to the break indices 3 and 4 which correspond to minor and major breaks in other approaches. One problem is that the phonetic descriptions of the different strength levels with respect to pauses are rather vague in ToBI and other approaches.

Pauses are one of the main features of *tempo* and *fluency* of speech. Fluently produced read speech is ideal to study the relationship between syntactic and prosodic boundaries and its acoustic markers because the spoken versions of written texts can be investigated for different variables, e. g. comparisons within and across individuals or across languages. Comparisons between different tempo categories (Trouvain & Grice 1999, Werner et al. 2020) show that there are several options for which syntactic boundaries should be reflected by which prosodic breaks (as mentioned above). In addition, there are many individual strategies as to how this prosodic break should be marked with which pause.

Different levels of fluency and pausing can be observed in second language speech and in spontaneous speech (also in the first language). Speakers use more

and longer pauses in their second language compared to their native language (Lennon 1990, Cucchiarini et al. 2002, Trouvain & Braun 2020). In spontaneous speech we encounter disfluencies of various kinds in addition to speech pauses, sometimes in incomplete sentence constructions. In a recent approach, Brugos et al. (2019) suggest a ToBI annotation scheme (for English) to account also for disfluencies.

Pauses can play a key role in *expressivity*, e. g. to signal different emotional states with pauses (Viola & Madureira 2008). A higher degree of arousal is normally reflected by a higher (perceived) speech tempo. For instance, horse race commentaries are usually described as getting faster the closer the horses are coming to the finish. However, acoustic analyses of those commentaries (Trouvain & Barry 2000) reveal that the articulation rate remains the same over the race and that the number of pauses increases towards the end instead of decreasing as expected. The main characteristic of these pauses in the final part is that they are shorter and filled with strong inhalation noise – together with an immense increase of the mean pitch. Pauses can also be used to express emphasis as Strangert (2003) showed with experimental data, but dramatic effects can also be observed in spontaneous conversational data (e. g. Ward 2019).

It must be stressed that pauses in speech can reveal multiple origins. In their historical overview on speech pause research (with comprehensive parts on "filled pauses") O'Connell & Kowal (1983: 222) note that

> [p]auses are determined by breathing, embarrassment, weariness, anxiety, confusion, anger, interruption, pain, syntactic complexity, mendacity, availability of lexical items, emphasis, boredom, and a host of other situational, organismic, intersubjective, linguistic, and conventional factors.

3 Defining a pause in speech

Although speech pauses and what they constitute often seems to be taken for granted, presumed ad-hoc definitions along the lines of "a speech pause is an interruption of speech" often fall short when it comes to the annotation of pauses in spoken data. We distinguish here four types of "interruptions of speech": i) articulatory pauses, ii) pauses as listeners ("mute" phases in talk-in-interaction), iii) gaps at turn changes in conversations, and iv) pauses in connected speech sections when e. g. having the turn in a conversation. Only the last type is considered here as a typical speech pause. However, a clear distinction from the other types is not always possible and should be borne in mind when annotating pauses.

3.1 Articulatory pauses in stop consonants

Interruptions of the acoustic flow of speech can also be observed in the closure phases of unvoiced stops (Fig. 1). These intermissions that are often clearly visible in the speech signal are part of the articulatory movement and should therefore not be considered a speech pause (Hieke et al. 1983).

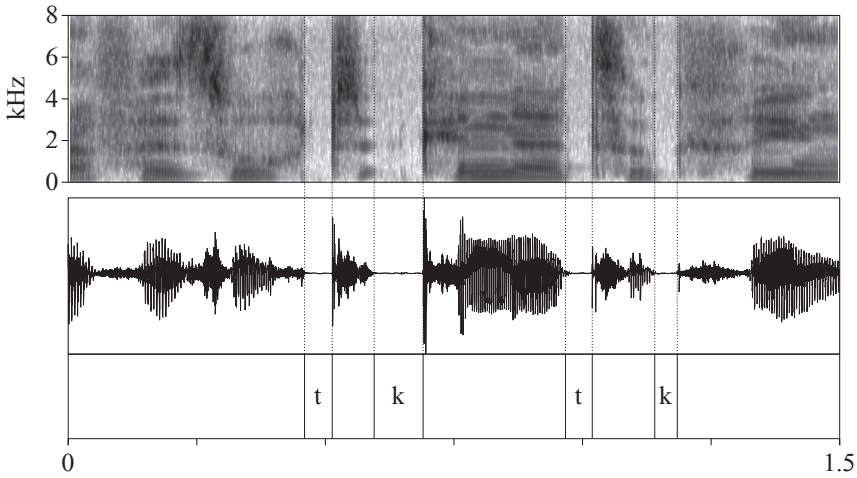


Fig. 1: Audio signal (duration: 1.5 sec) with four "articulatory pauses", i. e. interruptions of the exhalatory flow in the closure phase of voiceless stops in the section "versucht Kinder zu kriegen" (speaker: l03kpa (at 150.9 sec) from the Lindenstrasse corpus (IPDS 2006)). (► eContent_TR_a and ► eContent_TR_b.TextGrid)

eContent_TR_a.wav

eContent_TR_b.TextGrid

3.2 "Mute" phases as listener in conversations

In conversations it is typically the case that, most of the time, only one interlocutor is speaking, while the other/s do/es not speak (Fig. 2), though there are also phases where speakers overlap each other. Should the "mute" phases be considered as speech pauses? It could be argued that the time from one feedback utterance (backchannel) to the next one or the next opportunity to take the floor could be seen as a pause – which seems to be very different to speech pauses when speakers have the floor.

These "listener pauses" can be considered from two different perspectives: one is taking into account the conversational role of the interlocutor, here as the listener who is not claiming or attempting to take the turn. The other perspective is the speech production of the "active listener". Listening often involves articulatory activity, for instance feedback (or backchannel) utterances which can show a great variation: from simple phonatory "grunts" (often transcribed as "hm") to "ja" and reduplications like "jaja" and "jajaja" up to short

phrases (in Fig. 2 the speaker uses "ach so" and "nee nee nee"). These short articulatory sections are part of the listener role but they are not silence.

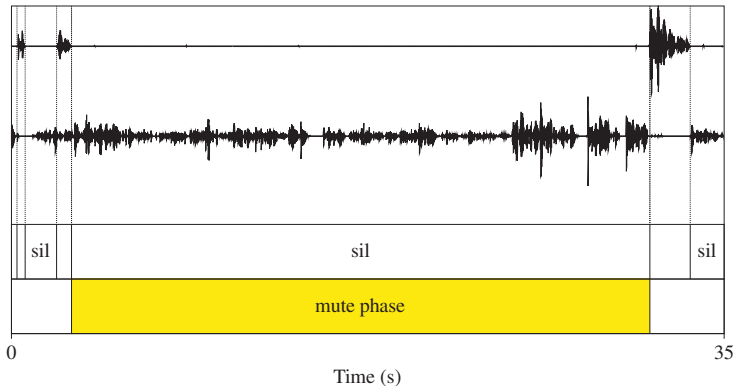


Fig. 2: Audio signal (duration 35 sec) of both speakers from a dialogue. The top speaker showed a longer "mute phase" of nearly 30 sec. She uttered "ach so [silence] nee nee nee [long silence] ja im Schweinestall die Szene hab ich wieder gesehen" (speaker l03ape at 94.0 sec from the Lindenstrasse corpus). (▶ eContent_TR_c)

eContent_TR_c.wav

3.3 Gap between speakers at turn changes in conversation

Another pause-like phenomenon in talk-in-interaction can be observed between speakers when the transfer of the floor is organised. These stretches of silence between interlocutors are usually called gaps (Fig. 3) and there seems to be a universal tendency to apply a "no gap, no overlap" principle (Sacks, Schegloff & Jefferson 1974, Stivers et al. 2009, Heldner 2011).

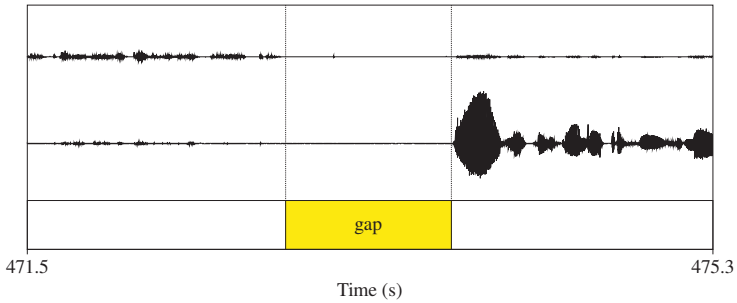


Fig. 3: Pause as a gap (920 ms) between speakers (with separated channels) in a conversation (audio signal (3.8 sec) taken from the IFADV corpus (van Son et al. 2008)). (▶ eContent_TR_d)

eContent_TR_d.wav

3.4 Typical speech pauses

The speech pauses we consider here do neither fall under the articulatory pauses nor the pauses without having the turn nor the between-turn pauses. The speech pauses considered here are always produced by speakers within their turn sections (as illustrated in Fig. 4). These pauses are usually the main markers of syntactic-prosodic breaks between sentences (in read speech).

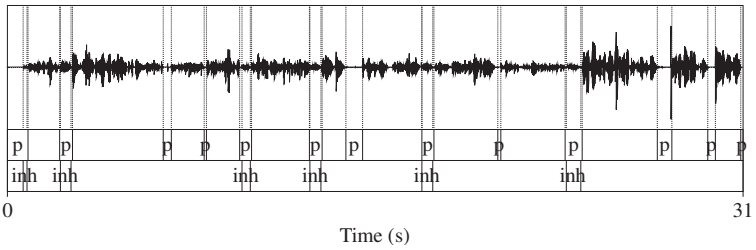


Fig. 4: Typical speech pauses (label "p" in the top tier) at the beginning and in the middle of turns. These pauses contain silence and often inhalation noises, the latter marked on an extra tier (label "inh" in the bottom tier). Please note that the boundaries of an inhalation noise within a pause do often not align with the pause boundaries (see section on breath information). Audio signal (duration: 31 sec) taken from the Lindenstrasse corpus (speaker l03kpa at 195.0 sec). (► eContent_TR_e)

eContent_TR_e.wav

4 Transcription of pauses

In phonetic notation it seems not very popular to mark pauses, presumably because phonetic alphabets like the *International Phonetic Alphabet* (IPA) are mainly used for the transcriptions of segments, very often for single words. In the IPA overview there is a section on "suprasegmentals" and the double bar symbol "‖" represents the boundary of a "major (intonation) group" which comes closest to a pause if IPA is used for stretches of speech that span more than a single prosodic phrase.

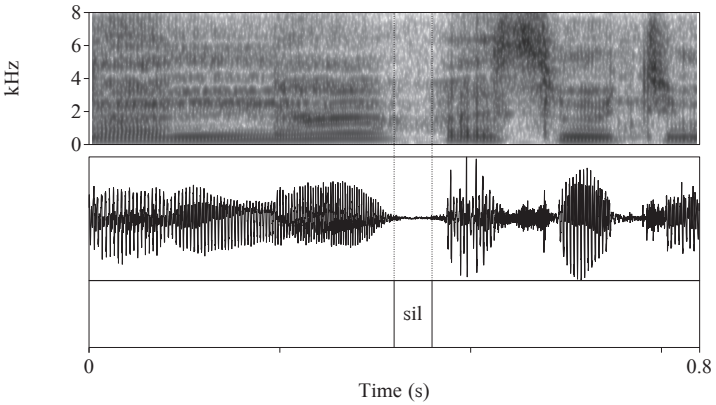
Hualde & Prieto (2016) follow this tradition with the development of an *International Prosodic Alphabet* (IPrA). If pauses are considered at all in IPrA, then as markers of prosodic breaks. This line of transcription is continued in the German prosodic transcription scheme *Deutsche Intonation: Modellierung und Annotation* (DIMA) by Kügler et al. (2019) (see also Kügler et al. in this volume) where the focus is on intonation, less so on prosodic phrasing, and not at all on pausing.

When transcribing conversations often more detailed approaches of marking pauses are applied. For instance, Crystal & Davy (1976: 11) distinguish between different pause length categories: "Four lengths of pauses are marked, the

shortest with a dot (·), the next longest with a dash (–), the next with two dashes (– –), and the longest with three (– – –).” This length categorisation is also reflected in other transcription schemes for conversations such as *Gesprächsanalytisches Transkriptionssystem* (GAT), e. g. in the updated version GAT2 (Selting et al. 2009). Pause duration can either be estimated or measured. If estimated, GAT2 follows a similar typographic notation as Crystal & Davy (1976) for (.) ‘micropauses’ (shorter than 0.2 sec), (-) ‘short pauses’ (between 0.2 and 0.5 sec), (– –) ‘medium pauses’ (between 0.5 and 0.8 sec) and (– – –) ‘longer pauses’ (between 0.8 and 1 sec). If measured, then deciseconds of pauses are given, e. g. ‘(0.4)’ for a measured pause duration of around 0.4 sec.

5 Annotation and segmentation of pauses in the speech signal

There is no generally agreed upon cut-off point for a pause, it varies e. g. from 100 ms (Trouvain 2004) to 200 ms (Lennon 1990, Cucchiarini et al. 2002) and 400 ms (Tavakoli 2011), just to mention three different values. It is proposed here not to have a fixed threshold in the acoustic correlate of a pause but to define a pause as a perceived pause plus a silence (excluding the closure phases of plosives). Such a definition may also include pauses shorter than 100 ms (Fig. 5). This would have the advantage of taking very short pauses into account which would be missed otherwise. Using a threshold can have a substantial influence on the results as a large-scale multilingual study of pause duration by Campione & Véronis (2002) could show.



eContent_TR_f.wav

eContent_TR_g.TextGrid

Fig. 5: Very short pause within an utterance at a minor prosodic phrase boundary with a silence shorter than 100 ms (here 52 ms). Audio signal (duration: 0.8 sec) taken from the Lindenstrasse corpus (speaker l0kpa at 195.0 sec utters “(ausei)nander [sil] weißt du mit dem”). (▶ eContent_TR_f and ▶ eContent_TR_g.TextGrid)

A pertinent problem with the segmentation of pauses in an acoustically based speech signal is that the closure phases of plosives after a pause and at the beginning of an inter-pause stretch can usually not be visually detected. The pragmatic solution we propose is to take a constant value of a plausible duration, e. g. 50 ms, as a differential value to subtract consistently from the entire pause duration.

A further problem comes with the detection of pauses, and subsequently with their segmentation. In linguistics there are also recordings analysed with an acoustic quality that is too low to detect all pauses. This is particularly problematic for pauses that are shorter than expected. In addition, rather often conversational corpora contain data where only one channel was recorded for all speakers. Such a signal makes it impossible to determine whether and when exactly which speaker produced a pause. A similar problem comes with data where individual speakers were recorded with separated microphones but where the vocalisations of the other speaker(s) make it difficult, and sometimes impossible, to have an exact segmentation of pauses. Those acoustic masking effects are often ignored or underestimated in the design of conversational corpus recordings.

6 "Silent" vs. "filled" pauses

Filler particles like [əː] and [əːm] are often denoted as "filled pauses". Those filler particles, that can be observed in many languages, are displayed with very different orthographic transcriptions in different languages such as "äh-ähm", "euh-euhm", "uh-u(h)m", "er-erm", and there seems to be no standard orthography within the same language for those particles. It is often overseen that there seems to be a variation of phonetic forms for the sequence of oral vowel plus nasal consonant across languages, e. g. in Chinese there is no "um" but "un" (Tian et al. 2017). In addition, there are more filler particles than just a lengthened central vowel that can be followed by a nasal consonant, for instance glottalisations (Belz & Trouvain 2019).

In contrast to these "filled pauses", the term "silent pauses" (or "empty pauses") is in use. Both terms, "silent pauses" and "filled pauses" are highly problematic when looking at pauses from a phonetic perspective.

The presumed idea of a "filled pause" is that there is a silence that is enriched with a phonetic particle, and that this filler particle plus the surrounding silence represents the "filled pause". However, most of the time the term is used rather loosely in research, and whenever more concrete definitions are given it becomes clear that a "filled pause" is used synonymously with the term filler particle – and not the entire silence. Fig. 6 shows an example of such a filler particle (here "uhm") – often denoted "filled pause" – nested in a pause with breath noise and silence before the speech part.

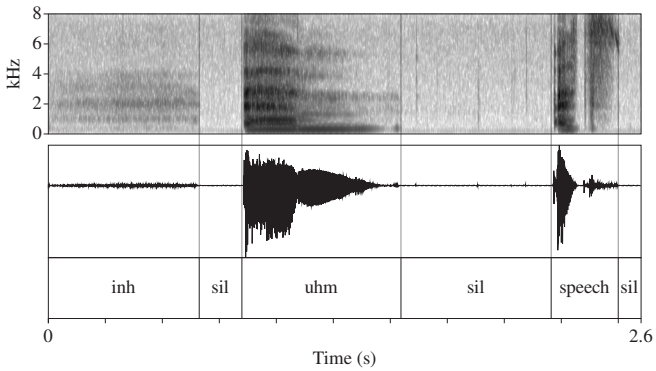


Fig. 6: Pause with a filler particle ("uhm") in a disfluent section containing two silent phases ("sil") and an inhalation noise ("inh") before speech (the word "als") starts. Audio signal (duration: 2.6 sec) taken from the Lindenstrasse corpus (speaker l03kpa at 121.16 sec)). (▶ eContent_TR_h and ▶ eContent_TR_i.TextGrid)

eContent_TR_h.wav

eContent_TR_i.TextGrid

In addition, those filler particles/filled pauses can occur in fluent sections of speech without any silence at all before or after, or even be cliticized onto adjacent words (Clark & Fox Tree 2002). Fig. 7 displays a filler particle with an extremely short silence before the particle and no silence at all after it, so that listeners would not perceive it as disfluent here. Those cases make it clear that the term "pause" for a filler particle is a misnomer. Also, denoting a filler particle in a fluent section of an inter-pause unit as a "disfluency" does not match the core idea of fluency vs. disfluency.

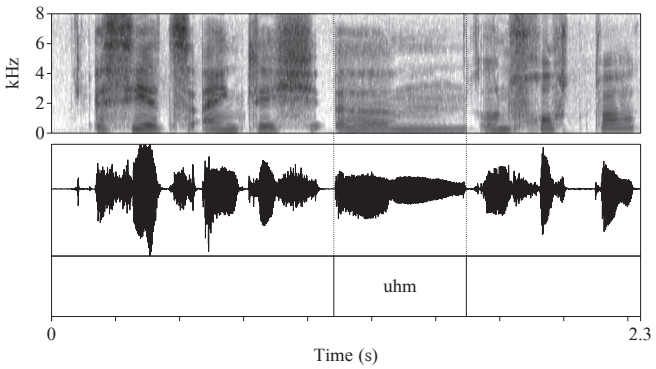


Fig. 7: Filler particle in a fluent section with a very short pause (56 ms "sil") before. Audio signal (duration: 2.3 sec) taken from the Lindenstrasse corpus (speaker l03kpa at 93.2 sec who utters "ist sie denn nun eigentlich [ähm] unfruchtbar"). (▶ eContent_TR_j and ▶ eContent_TR_k.TextGrid)

eContent_TR_j.wav

eContent_TR_k.TextGrid

The label "silent" in "silent" pauses is a misnomer as well. From an acoustic-phonetic point of view "silence" would be the absence of phonetic activity which excludes the acoustic correlates of inhalation and, to a lesser extent, exhalation in speech pauses. In other words, many "silent pauses" are in reality breath pauses with inhalation noises, i. e. not silent (or empty) at all.

7 Breath information in pauses

Phonetic studies have shown that duration and intensity of inhalation noises can be indicators of utterance planning in speech production and inform listeners about the length of the upcoming phrase (Fuchs et al. 2013, 2015). A recent study also suggests that in read speech, duration and intensity of inhalation noises are linked to a 'recovery' from the effort of the prior utterance (Kallay et al. 2019). When speakers are under physical stress they show different breathing patterns and forms of breath noises in speech pauses, e. g. with many exhalation noises that are otherwise infrequent in speech (Trouvain & Truong 2015).

A typical non-verbal vocalisation in spontaneous speech is laughter of which various forms can be described with characteristic noises of ex- and inhalation (Bachorowski & Owren 2001, Truong et al. 2019). A strong inhalation noise can mark the offset of a long and complex laugh (Chafe 2007, Truong et al. 2019). Also in (other) affect bursts, breath noises can play a crucial role, e. g. when startling or crying (Trouvain 2011).

On the level of pragmatics, breath noises can be used as discourse markers, signalling an intent to take the turn, and in some cultures respiratory noises are markers of politeness, e. g. in Korean (Winter & Grawunder 2012). Breath noises also have a high potential of signalling individuality, either by idiosyncratic acoustics, e. g. by inhalation noises with [s↓], an ingressive alveolar fricative (Trouvain 2015), or by different patterns of inhalation and exhalation (Lauf 2001, Kienast & Glitza 2003). The incomplete list above shows that breath noises are a rather rich source of information on the linguistic but also on the non-linguistic level. Fig. 8 and 9 show typical examples of speech pauses with nasal and oral inhalation noises which are sandwiched between edges of silence.

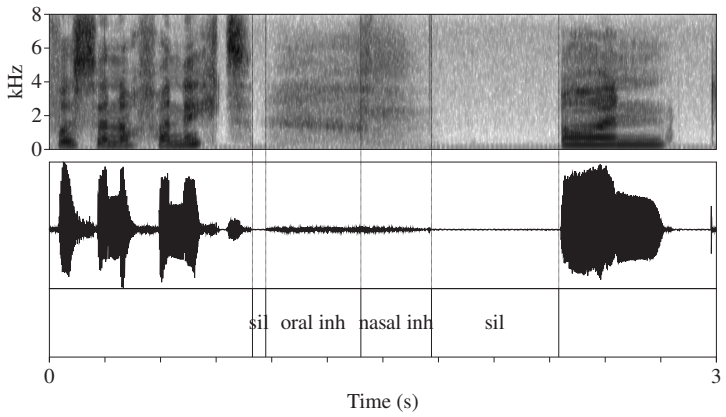


Fig. 8: Speech pause with a breath noise between silences that consists of an oral inhalation noise followed by a nasal inhalation noise. Audio signal (duration: 3.0 sec) taken from the Lindenstrasse corpus (speaker l03ape at 14.14 sec who utters "wusste noch von nichts [pause] und"). (▶ eContent_TR_l and ▶ eContent_TR_m. TextGrid)

eContent_TR_l.wav

eContent_TR_m.TextGrid

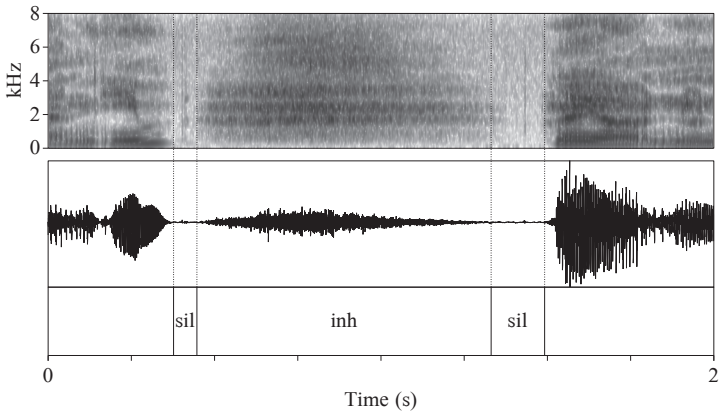


Fig. 9: Speech pause with an oral inhalation breath noise between silences. Audio signal (duration: 2.0 sec) taken from the Lindenstrasse corpus (speaker l03ape at 54.8 sec who utters "an ihm [pause] naja und nun"). (▶ eContent_TR_n and ▶ eContent_-TR_o.TextGrid)

eContent_TR_n.wav

eContent_TR_o.TextGrid

8 Silence and phonetic particles in "silent pauses"

As elaborated above, so-called "silent pauses" do most of the time also contain phonetic particles other than silence. Speech pauses can be enriched with one or more particles of, for instance, the following categories: breath noises, tongue clicks, glottal reflections. So far, there is no consistent use of the annotation of pause-internal phonetic particles (Trouvain & Werner 2020), and thus, there is no established set for those particles that could be recommended to be used for annotation.

Breath noises can be divided into those stemming from inhalation (with ingressive airstream) and those from exhalation (with egressive airstream). Breath noises should be made distinct from articulation with ingressive airstream, e. g. used for feedback utterances (backchannels) in various languages (see e. g. Eklund 2008), and of course to articulation with egressive airstream which is the normal way of speaking. Another distinction in breath noises can be made between the airways, i. e. whether it is only nasal on the one hand or oral (and potentially nasal at the same time) on the other hand. For inhalation, there are sometimes cases where speakers change their airways in the same breath noise, e. g. oral followed by nasal, or vice versa (see e. g. Kienast & Glitza 2003). Although breath noises are very often clearly visible in the spectrogram of speech signals, they might be of such a low intensity that their annotation should be primarily based on the audible inspection.

Tongue clicks can occur rather frequently in languages in which click sounds do not have a phonemic status (cf. Wright 2011, Trouvain 2014). They are not only used as word-like vocal gestures expressing disapproval (and other meanings) but they also occur in an unconscious way for word search and at beginnings of new discourse units in conversations. Often, they co-occur with inhalation noises (Trouvain 2014), see Fig. 10.

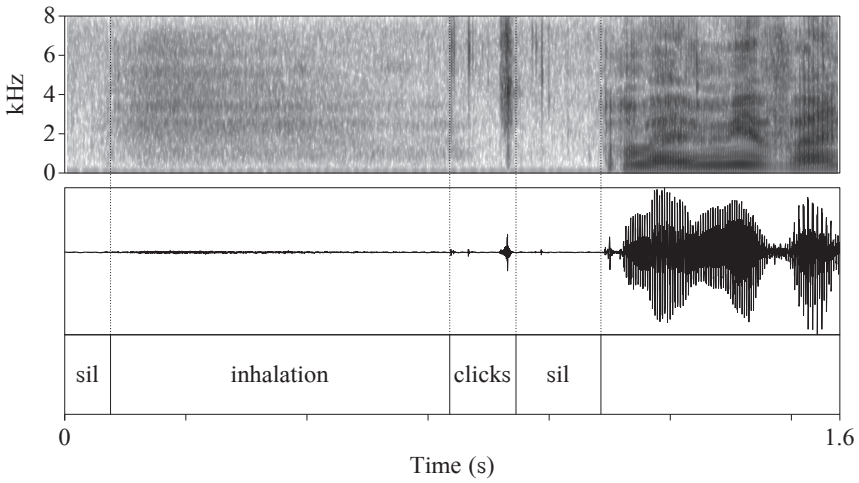


Fig. 10: Speech pause with an inhalation noise followed by clicks (highlighted) and a short silence before the speech. Audio signal (duration: 1.6 sec) taken from the Lindenstrasse corpus (speaker l03ape at 481.9 sec who utters "[pause] bei mir waren"). (▶ eContent_TR_p and ▶ eContent_TR_q.TextGrid)

eContent_TR_p.wav

eContent_TR_q.TextGrid

Glottal reflections can be considered as under-researched phenomena of phonetic particles in speech pauses (Belz & Trouvain 2019). They can appear in different phonetic shapes and can be regarded as interrupted intent of articulation. The example in Fig. 11 shows a hesitation consisting of a silence with additional creaky voice at the end of the first word and at the beginning of the second word including some isolated glottalisations. Acoustically, these glottal articulations are often hard to describe and physiological measurements such as electroglottography would be needed to get a clearer idea of their production process.

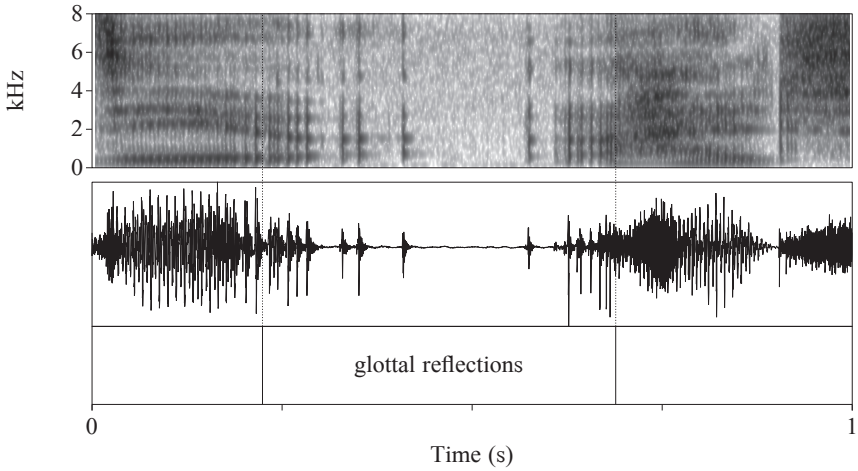


Fig. 11: Speech pause with a mixture of silence and glottalisations. Audio signal (duration: 1.0 sec) taken from the Lindenstrasse corpus (speaker l03ape at 89.3 sec who utters "sie [glott] halt"). (▶ eContent_TR_r and ▶ eContent_TR_s. TextGrid)

eContent_TR_r.wav

eContent_TR_s.TextGrid

One frequent non-verbal phenomenon in spontaneous speech that often happens in pauses is laughter (Trouvain 2014). Very often, laughter itself consists of elements that are treated here as pause particles such as inhalation noises and silence (see e. g. Truong et al. 2019). Although many instances of laughter can show clear links to speech pauses, speech planning and turn-taking, it seems reasonable to consider laughing as a complex phenomenon of its own that can be analysed independently from speech pauses.

9 Perceived pauses

The perception of pauses in speech depends on various factors for different levels of linguistic and phonetic processing (cf. Duez 1993, Strangert 1993, Swerts 1998, Carlson et al. 2005). Apart from silence and the afore-mentioned pause-internal particles such as inbreath noises and filler particles, there are further cues for pause perception in speech: phrase-final (or pre-pausal) lengthening, intonational boundary tones, voice quality (e. g. creaky voice), drops in intensity, and of course syntactic information.

Trouvain & Werner (2021) could show that listeners are able to detect pauses in (spontaneous) speech without any presence of silence or pause-internal phonetic particles at all, though this was valid not for all participants in their

experiment. It is important to say that there was no subject in this study who was able to detect all pauses in the stimuli. It is argued that a pause in speech perception should not be confused with a pause in speech acoustics. Automatic procedures to detect pauses, e. g. de Jong & Wempe (2009), can reliably help when segmenting inter-pause units in the speech signal. In human speech perception, the detection of pauses seems to be strongly linked to the listeners' interpretation of silences and other expected cues in the linguistic message. Arguably, pauses with breath noises (or breath pauses in short) are strong markers of prosodic boundaries, besides the other cues mentioned above. In general, breath pauses seem to occur in longer pauses whereas shorter pauses do usually not contain breath noises (Fuchs et al. 2013, Trouvain et al. 2020). There seems to be also a correlation with pause duration and prosodic boundary strength with longer pauses (and mostly breath pauses) reflecting a higher-level prosodic break (Trouvain et al. 2020).

10 Conclusions and outlook

Being inherent in speech, pauses should not be ignored in phonetic annotation of speech material that goes beyond single utterances (even if that means marking them as absence of speech). Their variable character poses challenges for definitions in terms of function, duration, and phonetic components.

Our illustrations and descriptions made it clear that speech pauses can be regarded from different angles. Depending on the perspective, pauses can be defined and categorised in different ways, with consequences for the annotation and segmentation in corpora. We hope that we raised awareness that a "pause" for one line of research is not necessarily identical with a "pause" in another line of research. Although in this article we looked at pauses from a phonetic point of view, it cannot be taken for granted that phonetic studies in general have such a detailed view. Belz & Trouvain (2019) suggest six levels for the annotation of pauses and pause-internal particles, in contrast to most other annotation procedures that consider pauses on just a single level.

In this article, we described pauses with respect to both the acoustic and the perceptual domain but there is also the articulation side to them in natural speech. While their acoustic manifestations may be similar, there are articulatory differences concerning rest postures, speech-ready, and inter-speech pause postures (Ramanarayanan et al. 2013). It should thus be emphasised that there are mismatches and a pause in one domain may not always entail pauses in the other two domains., e. g. closure phases of plosives vs. perceived pauses without stretches of silence.

This article on pauses in speech had the focus on the acoustic signal and ignored visual signals, e. g. from video recordings or motion capture films. It remains an open question how pauses – and pause-internal particles – are manifested visually either in read aloud speech (e. g. newscasters) or in spontaneous conversations. Further rather unexplored research areas concern the pausing behaviour in using a sign language (for an exception see Grosjean 1979) which can likewise be investigated as scripted language (e. g. with an interpreter of a news broadcasting) and of course spontaneous signed conversations.

Acknowledgements

The authors thank Bernd Möbius and Beeke Muhlack but also both editors for their helpful comments on an earlier draft of this article. This research was funded in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID TR 468/3-1.

References

Bachorowski, Joanne A. & Michael J. Owren. 2001. Not all laughs are alike: voiced but not unvoiced laughter readily elicits positive affect. *Psychological Science* 12. 252–257.

Beckman, Mary E. & Gayle A. Ayers. 1994. *Guidelines for ToBI Labelling*. http://www.speech.cs.cmu.edu/tobi/ (last accessed 20 July 2021)

Belz, Malte & Jürgen Trouvain. 2019. Are 'silent' pauses always silent? Proc. 19th *International Congress of Phonetic Sciences* (ICPhS), Melbourne. 2744–2748.

Brugos, Alejna, Langston, Alison, Shattuck-Hufnagel, Stefanie & Nanette Veilleux. 2019. A cue-based approach to prosodic disfluency annotation. Proc. 19th *International Congress of Phonetic Sciences* (ICPhS), Melbourne. 3414–3417.

Campione, Estelle & Jean Véronis. 2002. A large-scale multilingual study of silent pause duration. Proc. *Speech Prosody*, Aix-en-Provence. 199–202.

Carlson, Rolf, Julia Hirschberg & Swerts, Marc. 2005. Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication* 46(3–4). 326–333.

Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84(1). 73–111.

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

Crystal, David. & Derek Davy. 1976. *Advanced Conversational English.* London: Longman.

Cucchiarini, Catia, Helmer Strik, & Lou Boves. 2002. Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America* 111(6). 2862–2873.

Duez, Danielle. 1993. Acoustic correlates of subjective pauses. *Journal of Psycholinguistic Research* 22(1). 21–39.

Eklund, Robert. 2008. Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech. *Journal of the International Phonetic Association* 38(3). 235–325.

Fuchs, Susanne, Caterina Petrone, Jelena Krivokapić & Phil Hoole. 2013. Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41. 9–47.

Fuchs, Susanne, Caterina Petrone, Amélie Rochet-Capellan, Uwe Reichel & Laura Koenig. 2015. Assessing respiratory contributions to f0 declination in German across varying speech tasks and respiratory demands. *Journal of Phonetics* 52. 35–45.

Grosjean, François. 1979. A study of timing in a manual and a spoken language: American sign language and English. *Journal of Psycholinguistic Research* 8. 379–405.

Heldner, Mattias. 2011. Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *Journal of the Acoustical Society of America* 130(1). 508–513.

Hieke, Adolf E., Sabine Kowal & Daniel C. O'Connell. 1983. The trouble with "articulatory" pauses. *Language and Speech* 26. 203–214.

Hualde, José I. & Pilar Prieto. 2016. Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 5. DOI: http://doi.org/10.5334/labphon.11

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

Jong, Nivja H. de & Ton Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41(2). 385–390.

Kallay, Jeffrey E., Ulrich Mayr & Melissa A.Redford. 2019. Characterizing the coordination of speech production and breathing. Proc. 19th *International Congress of Phonetic Sciences* (ICPhS), Melbourne, 1412–1416.

Kienast, Miriam & Florian Glitza. 2003. Respiratory sounds as an idiosyncratic feature in speaker recognition. Proc. 15th *International Congress of Phonetic Sciences* (ICPhS), Barcelona. 1607–1610.

Kügler, Frank, Stefan Baumann, Bistra Andreeva, Bettina Braun, Martine Grice, Jana Neitsch, Oliver Niebuhr, Jörg Peters, Christine T. Röhr, Antje Schweitzer & Petra Wagner. 2019. Annotation of German Intonation: DIMA compared with other systems. Proc. 19th *International Congress of Phonetic Sciences* (ICPhS), Melbourne. 1297–1301.

Lauf, R. 2001. Aspekte der Sprechatmung: Zur Verteilung, Dauer und Struktur von Atemgeräuschen in abgelesenen Texten. In Angelika Braun (ed.), *Beiträge zu Linguistik und Phonetik*, 406–420. Stuttgart: Franz Steiner Verlag.

Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40. 387–417.

Lieberman, Philipp. 1967. *Intonation, Perception and Language*. Cambridge, Mass.: MIT Press.

O'Connell, Daniel C. & Sabine Kowal. 1983. Pausology. In Walter A. Sedelow & Sally Yeates Sedelow (eds*): Computers in Language Research 2*, 221–299. Berlin: de Gruyter.

Ramanarayanan, Vikram, Louis Goldstein, Dani Byrd, & Shrikanth Narayanan. S. 2013. An investigation of articulatory setting using real-time magnetic resonance imaging. *The Journal of the Acoustical Society of America* 134(1), 510–519. https://doi.org/10.1121/1.4807639

Sacks, Harvey, Emanuel A. Schegloff & Gail Jefferson. 1974. A simplest systematics for the organization of turn taking in conversation. *Language* 50. 696–735.

Shattuck-Hufnagel, Stefanie & Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.

Selting, Margaret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elisabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock & Susanne Uhmann. 2009. Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10. 353–402.

Stivers, Tanya, Nick J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan de Ruiter, Kyung-Eun Yoon & Stephen C. Levinson. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) 106(26). 10587–10592.

Strangert, Eva. 1993. Speaking style and pausing. *Phonum* 2. 121–137.

Strangert, Eva. 2003. Emphasis by pausing. Proc. 15th *International Congress of Phonetic Sciences* (ICPhS), Barcelona. 2477–2480.

Swerts, Marc. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30(4). 485–496.

Tavakoli, Parvahneh. 2011. Pausing patterns: differences between L2 learners and native speakers. *ELT Journal* 65(1). 71–79.

Tian, Ye, Jonathan Ginzburg & Takehiko Maruyama. 2017. Self-addressed questions and filled pauses: A cross-linguistic investigation. *Journal of Psycholinguistic Research* 46 (1), 905–922.

Trouvain, Jürgen. 2004. *Tempo Variation in Speech Production. Implications for Speech Synthesis*. PhD thesis, Saarland University. Phonus 8, Phonetics, Saarbrücken.

Trouvain, Jürgen. 2010. Affektäußerungen in Sprachkorpora. Proc. 21st *Konferenz Elektronische Sprachsignalverarbeitung* (ESSV), Berlin. 64–70.

Trouvain, Jürgen. 2011. Zur Wahrnehmung von manipuliertem Weinen als Lachen. Proc. *22nd Konferenz Elektronische Sprachsignalverarbeitung* (ESSV), Aachen. 253–260.

Trouvain, Jürgen. 2014. Laughing, breathing, clicking — The prosody of nonverbal vocalisations. Proc. *Speech Prosody*, Dublin. 598–602.

Trouvain, Jürgen & William J. Barry. 2000. The prosody of excitement in horse race commentaries. Proc. *ISCA-Workshop on Speech and Emotion*, Newcastle (Northern Ireland). 86–91.

Trouvain, Jürgen & Bettina Braun. 2020. Sentence prosody in a second language. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford Handbook of Language Prosody*, 605–618. Oxford: Oxford University Press.

Trouvain, Jürgen & Martine Grice. 1999. The effect of tempo on prosodic structure. Proc. 14th *International Congress of Phonetic Sciences* (ICPhS), San Francisco. 1067–1070.

Trouvain, Jürgen & Khiet P. Truong. 2015. Prosodic characteristics of read speech before and after treadmill running. Proc. *Interspeech*, Dresden. 3700–3704.

Trouvain, Jürgen & Raphael Werner. 2020. Comparing annotations of non-verbal vocalisations in speech corpora. Proc. 6th Workshop on *Laughter and Other Non-verbal Vocalisations*, Bielefeld. 69–72.

Trouvain, Jürgen, Raphael Werner & Bernd Möbius. 2020. An acoustic analysis of inbreath noises in read and spontaneous speech. Proc. *Speech Prosody*, Tokyo. 789–793.

Trouvain, Jürgen & Raphael Werner. 2021. Human pause detection in spontaneous speech in an online experiment. Proc. 32nd *Konferenz Elektronische Sprachsignalverarbeitung* (ESSV), Berlin. 59–66.

Truong, Khiet P., Jürgen Trouvain & Michel-Pierre Jansen. 2019. Towards an annotation scheme for complex laughter in speech corpora. Proc. *Interspeech*, Graz. 529–533.

Son, Rob van, Wieneke Wesseling, Eric Sanders, & Henk van den Heuvel. 2008. The IFADV corpus: A free dialog video corpus. Proc. 6th *International Conference on Language Resources and Evaluation* (LREC 2008). 501–508.

Viola, Izabel Cristina & Sandra Madureira. 2008. The roles of pause in speech expression. Proc. *Speech Prosody*, Campinas (Brazil). 721–724.

Ward, Nigel. 2019. *Prosodic Patterns in English Conversation*. Cambridge: Cambridge University Press.

Werner, Raphael, Jürgen Trouvain & Möbius, Bernd. 2020. Ein sprachübergreifender Vergleich des Pausenverhaltens natürlicher Sprecher in verschiedenen Sprechtempi mit TTS-Systemen. Proc. 31st *Konferenz Elektronische Sprachsignalverarbeitung* (ESSV), Magdeburg. 101–108.

Winter, Bodo & Sven Grawunder. 2012. The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics* 40. 808–815.

Wright, Melissa. 2011. On clicks in talk-in-interaction. *Journal of the International Phonetic Association* 41(2). 207–229.