INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

How long is too long? How pause features after requests affect the perceived willingness of affirmative answers

Lea S. Kohtz[1], *Oliver Niebuhr*[2]

[1] Dept. of General Linguistics, ISFAS, Kiel University, Germany
[2] Mads Clausen Institute, IRCA, University of Southern Denmark, Sønderborg, Denmark
admin@isfas.uni-kiel.de, oniebuhr@mci.sdu.dk

Abstract

A perception experiment involving 28 German listeners is presented. It investigates – for sequences of request, pause, and affirmative answer – the effect of pause duration on the answerer's perceived willingness to comply with the request. Replicating earlier results on American English, perceived willingness was found to decrease with increasing pause duration, particularly above a "tolerance threshold" of 600 ms. Refining and qualifying this replicated result, the perception experiment showed additional effects of speaking-rate context and pause quality (silence vs. breathing vs. café noise) on perceived willingness judgments. The overall results picture is discussed with respect to the origin of the "tolerance threshold", the status of breathing in speech, and the function of pauses in communication.

Index Terms: Pause, silence, duration, breathing, noise, gap.

1. Introduction

It is not least owed to the intensive work of Stephen Levinson and colleagues that "the sound of silence" became a major topic in the speech sciences over the last couple of years [1]. Yet, pauses are still an understudied aspect of speech communication. Established measures of speaking rate simply either include or exclude pauses instead of trying to model their complex role in tempo perception; and those studies that explicitly investigate pauses typically focus on their formal and deliminative aspects. These aspects include, for example, how strongly pause durations vary across native and non-native speakers, speaking styles, and languages [2,3,4,5,6,46], if pause durations can be organized into different classes [4,7], which kinds of phonetic sounds occur in filled pauses [8,9,10, 11,12], and to what extend such pause variables correlate with boundary types of phrases and turns in discourse structure [5,6,9,13,14,15,16,17,18].

Only a few studies, so it seems, go beyond such formal and deliminative analyses and address the contribution of pauses to the pragmatic meanings of utterances. For example, [19] showed for Swedish that pauses are a means to emphasize the following word. The same emphasis strategy is also used in other languages like German, in which the pause duration is so precisely timed that it mirrors the duration of the following emphasized word [20]. Thus, pauses are shorter if the word to be emphasized is "kann" (can) than when it is "konntest" (could). In a different series of experiments, [21] found for English that inserting filled pauses (like "ehm") before decisive pieces of information do not make speakers sound more disfluent in the ears of listeners, but, on the contrary, convey listener-orientation and concern for the addressee. In a similar vein, there is a growing interest in the idea that pauses and their phonetic properties are involved in conveying attitudes and emotions like thoughtfulness, sadness, or anxiety [22].

The present study continues the latter line of research on pauses and attitudinal meanings, extending the recent perception experiment of [23]. Their experiment tested the assumption that the pause duration following a speaker's request subtly indicates the interlocutor's (un)willingness to comply with the request [24,25,26,27,28]. A too long pause conveys unwillingness, even if the interlocutor's response is affirmative. *But, how long is too long?* The experiment of [23] used stepwise manipulations in pause duration to address this question. Their results revealed a "tolerance threshold" of about 600 ms that, when exceeded, abruptly makes the interlocutor sound unwilling to comply with the speaker's request.

It is suggested by [23] that this "tolerance threshold" is caused by top-down inference. It is known from psychological research that the minimal vocal response time to a new, unpredictable stimulus is about 600 ms Thus, if language users are aware of this temporal landmark (e.g, based on implicit learning), then reaction times longer than 600 ms would allow them to conclude that the interlocutor's response to their request was not made spontaneously but resulted from additional cognitive processing due to careful reconsideration; and, no doubt, if an interlocutor has to think twice before saying "yes", then this 'yes' has a taste of "if I have to...".

Two questions are posed here in connection with this tolerance threshold. First, can the findings of [23] be replicated for German? Second, given that most (if not all) interval and timing aspects of the speech signal time are context-sensitively interpreted by listeners, it seems reasonable to ask whether the 600 ms are an absolute value, or whether the "tolerance threshold" is affected by pause-internal and pause-external context factors. Thus, the second question of the present paper is: What happens to the "tolerance threshold" if the conditions within or around the pause change?

Four context factors are tested, two pause-internal and two pause-external factors. The first external factor was the speaking rate of request and response, as local durations are well known to be interpreted in global speaking-rate contexts; pause durations are local durations just as sound segments are [29,30,31]. Moreover, looking at speaking rate was obvious as this is context factor constantly varies in speech. The second external factor was environmental noise, taking into account the fact that everyday conversations typically take place in non-silent lab situations, for example, in a car, in a bar, or on the street. The environmental noise was superimposed on the entire sequence of request, pause, and affirmative answer.

The two pause-internal factors were also selected with respect to ecological validity. Both factors concerned breathing. The study of [6] showed for Czech dialogues that pauses

http://dx.doi.org/10.21437/Interspeech.2017-50

filled by audible inhalation and exhalation are by about 2-3 times more frequent than silent pauses. Own analyses of randomly selected dialogue samples from the Kiel Corpus of Spontaneous Speech [32,33] suggest a similar proportion of pauses with and without audible breathing for German. Thus, the silent pauses in the experiment of [23] may actually not be the best representatives of pauses in speech. For this reason, the present perception experiment included turn-initial audible inhalation (of the addressed interlocutor) and turn-final exhalation (of the requester) as two separate pause-internal factors, in addition to the silent reference condition.

2. Method

2.1. Participants

A total of 28 students at Kiel University, 18 females and 10 males, took part in the experiment. They were between 22 and 33 years old (average age 23.4 years), native speakers of Northern Standard German, and naïve with respect to the actual aim and background of the experiment.

2.2. Stimuli

The experiment was based on 40 different stimuli. Their construction scheme is illustrated in Figure 1. As not all four pause-internal and pause-external factors are fully combined with each other, they were subsumed under two independent variables: Rate (2 levels) and Quality (4 levels). The third and most basic independent variable was Duration (5 levels). The stimulus construction scheme resulted in 20 stimuli per Rate condition or 40 stimuli in total.

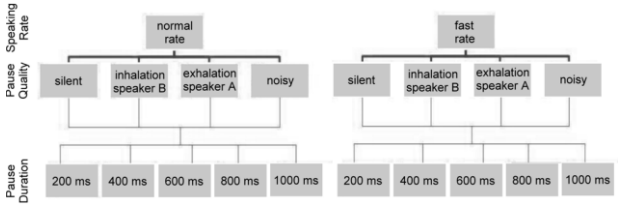

Figure 1: *Construction scheme of the 2x20 stimuli, based on the independent variables Rate (top), Quality (middle), and Duration (bottom).*

The following mini dialogue between two female Standard German speakers provided the basis of the stimuli:

- Requester: "Kannst Du mich nachher zur Uni fahren?" (Can you take me to the university later on?)
- Interlocutor: "Ja, natürlich." (Yes, of course.)

The two speakers were graduate students of phonetics and unknown to all 28 participants of the experiment. The speakers were instructed to produce the dialogue several times in an informal everyday style at two different rates, normal and fast. Moreover, each rate condition was elicited with three different types of pauses: silence, turn-final audible exhalation of the requester, and turn-initial audible inhalation of the interlocutor. All recordings were made digitally at 48 kHz and 24 bit in a sound-treated recording booth at Kiel University.

From the pool of recorded request-pause-answer dialogues, two dialogues were selected. They showed a big, but still natural-sounding difference in speaking rate. Second, the rate difference was similar for requester and interlocutor. Third, the other prosodic characteristics of the requester's and interlocutor' turns were constant (at a phonological level) across the two rate conditions.

The selected two dialogues had speaking rates of 4.6 and 7.3 syllables per second and served as bases for the further stimulus construction. Requests were produced in both rate conditions with a final rise based on a L*+H H-% nuclear contour starting on "Uni" [34]. Prenuclear H* accents were realized on "Kannst". The affirmative answers were in both rate conditions produced as a single prosodic phrase with a H* accent on each word, the second one being downstepped.

The pauses between the two turns in each base stimulus were removed and replaced by silence, exhalation, and inhalation. The inserted pauses were selected from the other mini dialogues and cut out such that they were exactly 1000 ms long. Then, the inserted pauses were shortened in four equal-sized steps of 200 ms using the PSOLA algorithm in PRAAT [35]. A stimulus was resynthesized for each pause condition, including the original 1000 ms condition. The minimum pause duration of 200 ms was selected to be above the gap detection threshold in dialogues [36,37].

As a final step, the sequences of request, silent pause, and affirmative answer were overlaid in PRAAT with environmental noise of a street café (www.salamisound.de/1020114-aussengastronomie-viele-cafes). The signal-to-noise ratio was +4 dB. Studies on Lombard speech showed that speech presented at +4 dB SNR is still easily comprehensible [38].

2.3. Experimental setup and procedure

The experiment was conducted in three parts created with PRAAT-MFC scripts. The first part was a short practice session with 10 randomly selected normal and fast stimuli. The second and third parts of the experiment presented the actual stimuli, but separately for the normal and fast speaking rate conditions. The presentation order of the two rate conditions, i.e. normal-fast or fast-normal, was equally distributed across the 28 participants. The 20 stimuli in each rate condition occurred five times in an overall pseudo-randomized order that avoided stimulus doublets. Stimulus randomizations were different for each participant.

Prior to the practice session, each participant was instructed as follows: *The perception experiment you participate in is about judging speaker attitudes. You will hear variants of a short AB dialogue with a constant wording. The dialogue consists of a request (A) and a subsequent affirmative response (B). Your task is simply to "read between the lines" and judge after each dialogue on a scale from 1 (weak) to 5 (strong) how willing the interlocutor actually is to comply with the request, i.e. how sincere "Ja, natürlich" (B) sounds to you.*

Note that this experimental procedure differed from that of [23], who distributed their stimulus conditions (i.e. pause durations) across many AB dialogues and groups of participants. This complex stimulus design was not adopted in the present study for three reasons. First and foremost, the aim of the present study was not to find out, *if* perceived willingness is affected by pause duration. This has already been shown by [23], and so it was not necessary to actively disguise this variable in the present experiment. Second, the present experiment included more factors than just pause duration. Thus, the stimulus differences were less monotonous and harder to figure out than in the experiment of [23]. Third, as a consequence of the multiple factors included here, the number of stimuli would have simply been too high to use the same sophisticated design that had been used by [23].

The experiment was run on individual desktop PCs in a silent, sound-treated lecture room of the General Linguistics Dept. at Kiel University. The participants listened to the stimuli via headphones at a constant pre-adjusted loudness level and made their judgments after each stimulus by ticking numbers (1-5) on a screen in front of them. A whole experimental session (incl. instructions and practice stimuli) took about 30 minutes.

3. Results

Judgments of perceived willingness were pooled across the five repetitions of each stimulus and analyzed in a three-way repeated-measures ANOVA, using Rate, Quality, and Duration as fixed within-subject factors. P-values reported in the following include Greenhouse-Geisser corrections, if the corresponding Mauchly test was significant.

All three within-subject factors yield highly significant main effects. In terms of partial eta-squared, the strongest main effect was that of Quality ($F[3,81]= 99.540$, $p<0.001$, $\eta_p^2=0.787$), followed by Duration ($F[4,108]=16.096$, $p<0.001$, $\eta_p^2= 0.373$), and Rate ($F[1,27]=7.927$, $p=0.009$, $\eta_p^2= 0.227$). There was only one significant interaction between Quality and Rate ($F[3,81]=5.252$, $p=0.008$, $\eta_p^2=0.163$). It resulted from the fact that the effect of Rate was larger in the exhalation condition than in all other conditions of the factor Quality.

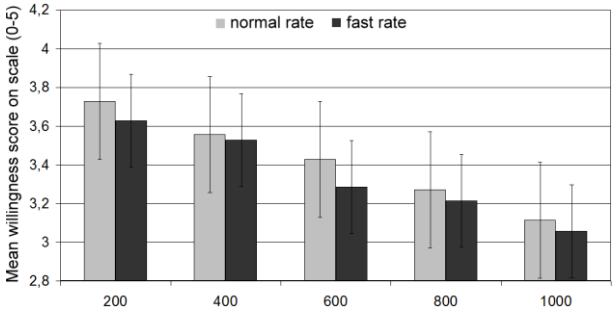

Figure 2: *Means and standard deviations (on the willingness scale from 0-5) for the normal-rate stimuli (light gray bars) and fast-rate stimuli (dark gray bars) with silent pauses from 200 ms to 1000 ms. Each bar represents 116 judgments.*

A descriptive results summary is provided in Figures 3(a)-(b). The combined effects of Rate and silent pauses on perceived willingness are separately displayed in Figure 2. It shows that the perceived willingness is lower across all pause durations when the speaking rate in the utterances surrounding the pause was higher. The decrease is in the range of 0.2 to 0.3 scale points, which is in the same order of magnitude as an increase in pause duration of 200 ms.

A detailed look at the factor Quality in Figures 3(a)-(b) shows that adding background (café) noise to the sequence of request, pause, and affirmative response yields the overall lowest level of perceived willingness. In contrast, perceived willingness is overall highest in those stimuli in which the pauses are filled by the interlocutor's audible inhalation prior to her affirmative response. Pauses filled by the requester's turn-final audible exhalation do not have the same effect; on the contrary, they result in a considerably lower level of perceived willingness. The silent pauses that were used in the original experiment of Roberts & Francis (2013) cause a perceived willingness level that falls in between the two breathing conditions. According to multiple post-hoc tests (with Sidak corrections), all four conditions of the factor Quality differ significantly from each other at $p<0.05$.

As regards Duration, both Figures 2 and 3 show decreasing willingness levels for increasing pause durations. Yet, perceived willingness does not decrease as gradually across the pause-duration continuum as it may seem at first glance. In actual fact, the continuum consists of two parts. Multiple post-hoc tests between all factor levels of Duration (with Sidak correction) reveal across all Quality conditions that *only* the duration step from 600 to 800 ms caused a significant decrease in perceived willingness. There are *no* further significant decreases between adjacent duration steps within the pause ranges of 200-600 ms and 800-1000 ms.

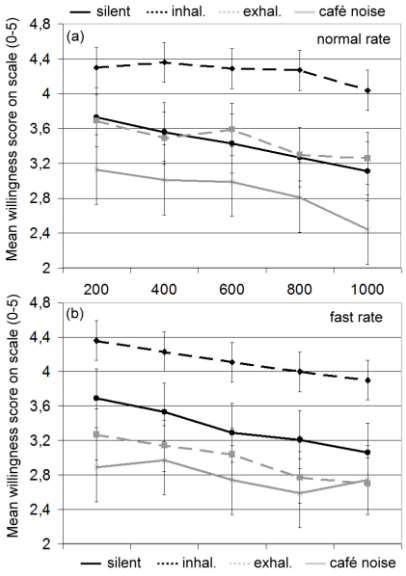

Figure 3: *Means and standard deviations for the (a) normal-rate and (b) fast-rate stimuli of the 5 Duration and 4 Quality conditions. Each point represents 116 judgments.*

4. Discussion and Conclusion

The present findings lend further support to the conclusion of [23]: The duration of the pause following a request changes the pragmatic meaning of the interlocutor's answer. If the pause is too long, then it signals to the requester that his/her interlocutor is less willing to comply with the request, even if the answer s/he gives is affirmative. Regarding the question how long a pause can be before it gets too long, the present results are also in line with those of [23]. That is, there is an abrupt and significant decrease in the answer's perceived willingness for pauses longer than 600-800 ms. This is exactly in the order to magnitude of the "tolerance threshold" determined by [23]. However, our listeners' ratings largely remain in the upper half of the willingness scale (> 2.5, probably due to our simpler stimulus presentation). Thus, the ratings decrease from 'more willing' to 'less willing' rather than from 'willing' to 'unwilling'. Not least for this reason, it is important to understand that the "tolerance threshold" mirrors perceptual sensitivity rather than categorical linguistic structure. Nevertheless, if you want to affirm a request and sound sincere, then, in order to be on the safe side, you should not wait longer than 600-800 ms with starting your answer.

Refining and qualifying the latter rule of thumb, our results clearly show that duration is not the only feature if a pause that can make a following affirmative answer sound

more or less willing. Pause-internal breathing is another important feature. Compared to the silent reference condition, interlocutors can take *more* time to give an affirmative response and still sound equally willing if they bridge (i.e. fill) the pause by audible inhalation. Audible exhalation of the requester cannot be exploited in the same way by interlocutors. The opposite is true, in fact. If the requester fills the pause by audible exhalation, then interlocutors have *less* time for giving an answer, if this answer is supposed to convey the same degree of willingness as in the silent reference condition.

Besides the fact that listeners are obviously able to distinguish between inhalation and exhalation noises (from different speakers), the present findings on breathing suggest that listeners counted the interlocutor's inhalation as part of the following affirmative answer. This shortened the perceived pause duration and, in consequence, caused an increase in perceived willingness. If this explanation holds, it means at the same time that the exhalation was not analogously considered part of the request, as this would have shortened the perceived pause duration and, as a result, increased willingness ratings as well. However, willingness ratings remained unchanged or even slightly decreased in the exhalation condition. That is, exhalation is treated more or less like a silent pause and hence as a part of the speech signal that is neither associated with any utterance nor linked with the interlocutor's reaction.

So, a more general implication of our results on breathing is that inhalation and exhalation pauses do not have the same status in speech communication, but are differently embedded in the surrounding syntagmatic structure. This conclusion would have considerable consequences for many areas of speech research, provided that it is supported by further studies. A promising starting point for these studies is that corpora based on KIM [32,33] and GAT-2 [39] have separate labels for inhalation and exhalation.

The speaking rate turned out to be a relevant pause-external feature. If request and response are produced at a higher rate, then the pause in between the two utterances also needs to be shorter in order to make the interlocutor's answer sound equally willing. This means that the communicative interpretation of pause durations is relative rather than absolute. What probably counts for perceived willingness is not the pause duration itself, but the ratio between this duration and the duration of other local context elements like words or syllables. Such a ratio-based functional interpretation is probably a general characteristic of speech perception that also applies to VOT, vowel quantity [29,30], and final lengthening [31].

The effect of background café noise could be explained with reference to the fact that acoustically variable signals are perceived to be longer than acoustically steady signals, cf. [40,41]. A silent pause is obviously a steady signal that, however, becomes very variable when filled by café noise. We assume that the high variability of café-noise pauses made listeners overestimate the actual physical pause duration; and this overestimation, in turn, triggered overall lower willingness ratings. For example, silent pauses in our study could be as long as 1000 ms and still conveyed about the same degree of perceived willingness as café-noise pauses of only about 200-400 ms. It is consistent with this explanation that a recent Lombard study provides evidence for shorter pause durations when dialogues are produced inside a loud car [42].

In summary, even when taking into account a certain drop in perceived willingness at the tolerance threshold, our conclusions and explanations suggest the following: If most pauses in everyday speech are either filled by inhalation, and if read speech is actually faster than spontaneous speech, then interlocutors in everyday conversation situations could have a lot *more* time than 600-800 ms to answer a request before their sincerity is critically questioned by the requester. On the other hand, interlocutors speaking in a noisy environment could have a lot *less* time, also depending on how much a slower speaking rate and audible inhalation can compensate for the presence of background noise. Quantifying the relative power of context factors is a worthwhile aim of follow-up studies.

A further line of follow-up studies should identify other relevant factors that affect the perceived willingness of pauses. One promising candidate would be intonation, or, more specifically, the boundary tone at the end of the request. In German, question-final rises and falls are not an indicator of sentence mode per se, but rather convey the hierarchical relationship of speaker and listener [43]. A final fall means 'please give me the answer that I would like to hear from you', whereas a final rise signals 'you are free to decide yourself'. This meaning difference implies that, compared to rising requests (as they were used in the present study), falling requests should be more restrictive in terms of tolerable pause durations, because the requester superordinates him/herself to the interlocutor and predetermines the answer. Initial supporting evidence for this assumption comes from a randomly chosen sample of 50 question-answer pairs from the German KIESEL corpus [44,45]. As is illustrated in Figure 4, the naturally produced pause durations between yes/no questions and subsequent answers are shorter for falling than for rising questions. Based on this pattern, pilot perception experiments analogous to that of the present paper are currently being created.

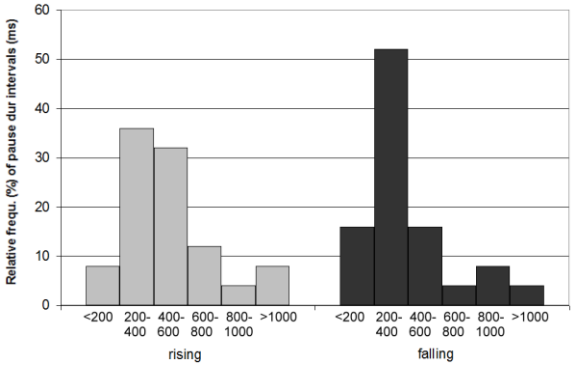

Figure 4: *Histogram of pause durations in between answers and y/n questions with final rising intonation (light gray, left) and with final falling intonation (dark gray, right). Each distribution represents 25 items from the KIESEL corpus [45].*

Future research should also go beyond requests and yes/no questions in general. Wh-questions seem to be an equally interesting subject for analyzing the meaningful role of pause duration and its interplay with other factors. Results of a pilot study show that the pause duration in between a wh-question and a following answer also contributes to the answer's pragmatic meaning in terms of its reliability and credibility. However, a "good" pause in the wh-question context seems to be exactly the opposite of a "good" pause in a request or yes/no question context. Perceived reliability and credibility of an answer to a wh-question *in*crease rather than *de*crease with increasing pause duration. So, longer pauses are better pauses in the wh-context. If and to which duration threshold this is true is currently tested by us in a perception experiment.

5. References

[1] J. Holler, K.H. Kendrick, M. Casillas, S.C. Levinson (Eds), *Turn-Taking in Human Communicative Interaction*. Lausanne: Frontiers Media, 2016.

[2] F. Goldman-Eisler, "The distribution of pause durations in speech," *Language and Speech*, 4, 232–237, 1961.

[3] D. Duez, "Salient pauses and non salient pauses in three speech style," *Language and Speech*, 25, 11-28, 1981.

[4] I. Campione and J. Véronis, "A large-scale multilingual study of silent pause duration," *Proc. Eurospeech, Denver, USA, 2002,* 199–202.

[5] K. Weilhammer and S. Rabold, "Durational aspects in turn taking," *Proc. 15th International Conference of Phonetic Sciences, Barcelona, Spain, 2003,*2145-2148.

[6] L. Weingartová, E. Churaňová, and P. Šturm, "Transitions, pauses and overlaps: Temporal characteristics of turn-taking in Czech," *Proc. 7th International Conference of Speech Prosody, Dublin, Ireland, 2014,* 502-506.

[7] J. Sabol and J. Zimmermann, "Komunikačná hodnota pauzy," *Úloha reči a hudby v životním prostředí Xxiii, Akustická conference, České Budějovice, Czech Republic, 1984,* 225–229.

[8] E. Shriberg and R. Lickley, "Intonation of clause-internal filled pauses," *Phonetica*, 50, 172-179, 1993.

[9] A. Batliner, A. Kießling, S. Burger, and E. Nöth, "Filled pauses in spontaneous speech," *Proc. 13th International Congress of Phonetic Sciences, Stockholm, Sweden, 1995,* 472-475.

[10] V. Horváth, "Filled pauses in Hungarian: Their phonetic form and function," *Acta Linguistica Hungarica*, 57, 288-306, 2010.

[11] D. Duez, "Acoustico-phonetic characteristics of filled pauses in spontaneous French speech: preliminary results," *Proc. Workshop on Disfluencies in Spontaneous Speech, Edinburgh, UK, 2001,* 41-44.

[12] M. Karpiński, "Acoustic features of filled pauses in Polish task-oriented dialogues," *Archives of Acoustics*, 38, 63-73, 2013.

[13] K. Ruder, "Duration of silent intervals as a perceptual cue to speech pauses," *Perceptual and Motor Skills*, 36, 47-57, 1973.

[14] A. Butcher, "Aspects of the speech pause: Phonetic correlates and communicative functions," *Arbeitsberichte Institut für Phonetik (AIPUK)*, 15, 3-233, 1981.

[15] D. Duez, "Perception of silent pauses in continuous speech," *Language and Speech*, 28, 377-389, 1985.

[16] B. Zellner, "Pauses and the temporal structure of speech,", in E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 41-62). Chichester: Wiley, 1994.

[17] M.E. van Donzel and F.J. Koopmans-van Beinum, "Pausing strategies in discourse in Dutch," *Proc. International Conference of Spoken Language Processing, Philadelphia, USA, 1996,* 1029–1032.

[18] M. Heldner and J. Edlund, "Pauses, gaps and overlaps in conversation," *Journal of Phonetics*, 38, 555–568, 2010.

[19] E. Strangert, "Emphasis by pausing," *Proc. 15th International Congress of Phonetic Sciences, Barcelona, Spain, 2003,* 2477–2480.

[20] O. Niebuhr, "Formen und Funktionen der Emphase–Grundsteine einer Systematisierung", *Presentation given at Dept. of General Linguistics, Kiel University, Germany, 2014.*

[21] K. Fischer, O. Niebuhr, E. Novák-Tót, and L.C. Jensen, "Hesitation Markers Contribute to Addressee Orientation," *Proc. 43rd Annual Conference of the German Acoustical Society, Kiel, Germany,* 1-4.

[22] I.C. Viola and S. Madureira, "The roles of pause in speech expression," *Proc. 4th International Conference of Speech Prosody, Campinas, Brazil, 2008*, 721-724.

[23] F. Roberts and A.L. Francis, "Identifying a temporal threshold of tolerance for silent gaps after requests," *Journal of the Acoustical Society of America*, 133, EL471-477, 2013.

[24] J. Davidson, "Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection," in J. M. Atkinson and J. Heritage (Eds), *Structures of Social Action: Studies in Conversation Analysis* (pp. 102-128). Cambridge: CUP, 1984.

[25] A. Pomerantz, "Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes," in J.M. Atkinson and J. Heritage (Eds), *Structures of Social Action: Studies in Conversation Analysis* (pp. 57-101). Cambridge: CUP, 1984.

[26] F. Roberts, P. Margutti, and S. Takano, "Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese," *Discourse Processes*, 48, 331–354, 2011.

[27] K.H. Kendrick and F. Torreira, "The timing and construction of preference: A quantitative study," *Talk given at the International Discourse, Communication, and Conversation Conference, Loughborough, UK, 2012*.

[28] S. Bögels, K.H. Kendrick, and S.C. Levinson, "Never say no… How the brain interprets the pregnant pause in conversation," *PLoS One*, 10, e0145474, 2015.

[29] Q. Summerfield, "Articulatory rate and perceptual constancy in phonetic perception," *Journal of Experimental Psychology: Human Perception and Performance,* 7, 1074–1095, 1987.

[30] R. Smiljanic and A.R. Bradlow, "Stability of Temporal Contrasts across Speaking Styles in English and Croatian," *Journal of Phonetics*, 36, 91-113, 2008.

[31] M. Plüschke and J. Harrington, "The domain of phrase-final lengthening in Estonian," in E.-L. Asu and P. Lippus (Eds), *Nordic Prosody XI* (pp. 293-302). Frankfurt: Peter Lang, 2012.

[32] K.J. Kohler, "Labelled data bank of spoken Standard German: The Kiel Corpus of Read/Spontaneous Speech," *Proc. International Conference of Spoken Language Processing, Philadelphia, USA, 1996,* 73-77.

[33] B. Peters, "The Database – The Kiel Corpus of Spontaneous Speech," *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK),* 35a, 1-6, 2005.

[34] F. Kügler, B. Smolibocki, D. Arnold, et al., "DIMA - Annotation Guidelines for German Intonation," *Proc. 18th International Congress of Phonetic Sciences*, *Glasgow, Scotland, 2015*, 1-5.

[35] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, 5, 341-345, 2001.

[36] M.B. Walker and C. Trimboli, "Smooth transitions in conversational interactions," *The Journal of Social Psychology,* 117, 305–306, 1982.

[37] M. Heldner, "Detection thresholds for gaps, overlaps and no-gap-no-overlaps," *Journal of the Acoustical Society of America*, 130, 508–513, 2011.

[38] P.F. Assmann and Q. Summerfield, "The perception of speech under adverse conditions," *Springer Handbook of Auditory Research*, 18, 231-308, 2004.

[39] M. Selting, P. Auer, D. Barth-Weingarten, et al., "A system for transcribing talk-in-interaction: GAT 2," *Gesprächsforschung: Online-Zeitung zur verbalen Interaktion*, 12, 1-51, 2001.

[40] I. Lehiste, "Influence of fundamental frequency pattern on the perception of duration," *Journal of Phonetics*, 4, 113-117, 1976.

[41] W.A. van Dommelen, "Interactions of fundamental frequency contour and perceived duration in Norwegian," *Phonetica*, 52, 180-187, 1995.

[42] R. Landgraf, "Simulating complex speech production environments," in O. Niebuhr and R. Skarnitzl (Eds), *Tackling the complexity of speech* (pp. 97-112). Prague: Charles University Press, 2015.

[43] C. Petrone and O. Niebuhr, "On the intonation in German intonation questions: The role of the prenuclear region," *Language and Speech*, 57, 108-146, 2014.

[44] Niebuhr, O. (2010). "On the phonetics of intensifying emphasis in German," *Phonetica*, 67, 170-198.

[45] O. Niebuhr, B. Peters, R. Landgraf, and G. Schmidt, "The Kiel Corpora of "Speech & Emotion" - A Summary," *Proc. 41st Conference of the German Acoustical Society, Nuremberg, Germany, 2015,* 1-4.

[46] P. Tavakoli, "Pausing patterns: differences between L2 learners and native speakers," *ELT Journal*, 65, 71-79, 2011.