

The role of duration as a phonetic correlate of focus

Frank Kügler

Institute of Linguistics
University of Potsdam, Germany

f.kuegler@googlemail.com

Abstract

The aim of the present paper is to collect quantitative data on durational patterns of target words in different information structure conditions. The acoustic data will be the basis for the manipulation of stimuli to conduct perception tests that will deal with the role of duration as a correlate of focus prominence. For that purpose, a production study is presented that investigates the effects of information structure (wide, narrow, and contrastive focus, prefocal and postfocal givenness), sentence length and position in a sentence on duration of a target constituent in German. Duration of target constituents was measured in 400 utterances produced by 10 speakers. The predictions that focus increases target word duration has been confirmed, while the expected decrease in duration due to givenness has only been confirmed for prefocal given constituents. Postfocally, duration is equivalent to wide focus duration. The effects of sentence length and position have only partly been confirmed; a constituent seems to be shorter in longer sentences than in short ones, and target words occurring early in a sentence appear to be longer than late occurring ones.

1. Introduction

Communication is a process in which a speaker transfers information to a hearer. In this transfer, any utterance has a certain information structure. Any utterance contains at least a focus [7]. According to [7], an utterance may further contain a non-focus part that may branch in a topic and/or a background. Focus is here understood as ‘the presence of alternatives that are relevant for the interpretation of linguistic expressions’, and topic as a referent or constituent which the remainder of the sentence is about [16]. Backgrounded information in this experiment refers to *given* information, defined as previously mentioned in the discourse [2, 3].

Based on the assumptions of Focus-Prominence-Theory (e.g. [17, 18]) the expression of abstract focus prominence is language specific. Many languages use F0 as a prominent cue. In these languages duration almost ever accompanies focus prominence. This paper is the first step in an investigation to disentangle the role of F0 and duration as phonetic correlates of focus prominence. The general question behind is whether intonation languages like German employ the correlate of duration as a consequence of using higher or later F0 peaks, or whether duration is a functionally relevant prosodic cue on its own.

As has been reported in the literature, focus causes an increase in duration on focused constituents, see for instance [1] for Swedish, [8, 9, 10] for English, [4, 11] for German, and [15] for Korean. In addition to the fact that duration increases in focus, the duration of a constituent also appears to depend on its position in a sentence [11, 12]. Longer durations are found in

earlier positions of the sentence. Also, if sentences are longer, i.e. contain more words, the duration of constituents appears to be longer in earlier positions [11].

The relation of duration to different focus domains has been investigated by [4] who found a negative correlation with the size of the focus domain, i.e. the smaller the focus domain the longer the duration of a particular constituent. In general, focus seems to be closely connected to longer duration. This can also be seen in situations where no other tonal correlates of focus are present, i.e. in the case of second occurrence focus (SOF) [5, 13, 14]. In English and German constituents that are in the scope of a focus operator such as ‘only’ but do not receive the nuclear stress of the sentence and thus are not associated with pitch accents show longer durations as equivalent given constituents without any focus particle.

This paper reports on durational patterns in German. The aim of this paper is to provide quantifiable data of the effects of ‘position’, ‘length’, and ‘information structure’ (focus and givenness). Based on these data perception experiments are planned that will investigate the role of duration and focus in more detail. For the purpose of the production experiment sentences were created which contain a target word in different positions in the sentence. In addition, sentences varied in total length (counted as number of syllables per sentence). The sentences were embedded in contexts to elicit different information structures (see below).

2. Method

2.1. Speech materials

Two target words were embedded in carrier sentences of different length in different positions. The two target words are nonsense words, contain only sonorant segments in order to ensure easy pitch tracking. One of the words is monosyllabic and one is bisyllabic with word stress on the first syllable. These nonsense words were used as surnames in order to make their occurrence more natural. The target words are *Mohn* [ˈmo:n] and *Liehner* [ˈli:nə]. The carrier sentences are of the following structure.

- (1)
 - a. Frau Mohn will ein Lamm malen.
 - b. Ein Lamm will Frau Mohn malen.
 - c. Frau Mohn will ein Lamm im Berliner Tierpark malen.
 - d. Im Berliner Tierpark will Frau Mohn ein Lamm malen.

Two target words were embedded in short (1-a), (1-b) and long sentences (1-c), (1-d) either early (1-a), (1-c) or late (1-b), (1-d) in the sentence. The set of sentences in (1) was put in five different contexts yielding five different information structures of the target words.

The information structural baseline is considered to be an all-new sentence, with no particular part of the sentence in fo-

cus, i.e. the answer of a question like (2-a). Narrow focus was elicited by (2-b), and contrastive focus by (2-c), where *Drahner* is contrasted with the target word *Liehner*. The variable of givenness was tested in prefocal and postfocal position. The corresponding questions were either asking for a constituent after the target word, prefocal givenness (2-d), or before the target word, postfocal givenness (2-e).

- (2) a. What happen?
 b. Who does want to paint a lamb?
 c. Does Mrs Drahner want to paint a lamb?
 d. Does Mrs Mohn want to paint a hoarse?
 e. Does Mr Mohn want to paint a lamb?

A total of 8 unique question-answer pairs (2 items \times 2 positions \times 2 sentence lengths) were constructed and each pair was realized in the 5 information structure conditions, resulting in $8 \times 5 = 40$ sentences per speaker. All the 40 sentence-pairs were presented to each speaker in a pseudo-randomized manner; items from four other unrelated experiments were interspersed as fillers. Four pseudo-randomized lists were prepared to minimize order effects.

2.2. Recording procedures

The experiment was carried out using presentation software. Participants were seated in a sound proof booth in the recording studio at the University of Potsdam in front of a condenser microphone. Participants were familiarized with the task through written and verbal instructions. Each trial consisted of a presentation of the question and its answer on the computer screen. As soon as the sentences were presented, participants heard the pre-recorded question, spoken by a male voice. Participants were instructed to speak out the answer displayed on the screen as a response to the question. If the question was answered without any hesitations or false starts, the next trial was presented. If there were hesitations, participants were asked to repeat the answer. Presentation flow was controlled by the experimenter, and participants were allowed to take a break whenever they wanted. The sentences produced by participants were recorded digitally on a computer.

2.3. Participants

18 native speakers of Standard German spoken in the Berlin region participated in the experiment. All were female students at the University of Potsdam. Each speaker was paid or given course credits for participation and took approximately 35 minutes to complete the experiment.

2.4. Data pre-processing and statistical analysis

For this study the first 10 speakers have been selected for annotation and analysis. This resulted in a total of 400 utterances (10 speakers \times 8 items \times 5 conditions). The recordings were digitized at a sampling frequency of 44.1 kHz and 16 bit resolution. Data were labeled in Praat [6] by hand at the level of the syllable. Standard segmentation procedures were applied.¹

For each target word the duration was detected using a Praat script. A repeated measures ANOVA was carried out with duration as the dependent variable. The repeated measures ANOVA disregards the mean of each individual participant (the grand mean of each participant equals to zero), so that there is no additional need of normalizing for the factor *speaker*.

¹A neglectable number of the long sentences (0.4 %) were uttered as two intonation phrases.

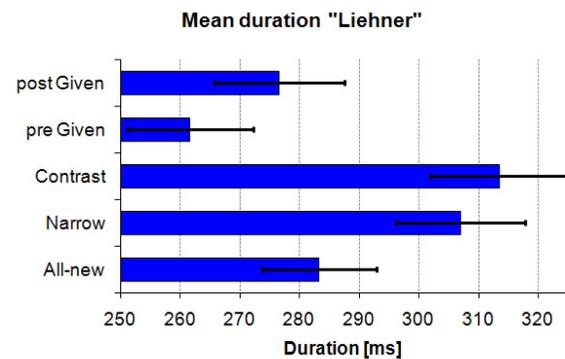


Figure 1: Mean duration pooled across 10 speakers in ms with 95% confidence intervals for the target word *Liehner* given for wide (All-new), narrow, contrastive focus and prefocal (pre-Given) and postfocal (postGiven) givenness.

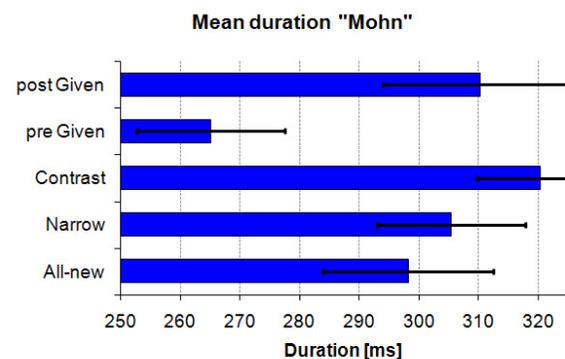


Figure 2: Mean duration pooled across 10 speakers in ms with 95% confidence intervals for the target word *Mohn* given for wide (All-new), narrow, contrastive focus and prefocal (pre-Given) and postfocal (postGiven) givenness.

2.5. Predictions

Based on the above discussed studies, for the factor information structure an increase in duration for focus, and a decrease in duration for givenness is expected. For the factor position, longer durations for earlier target words are expected. And for the factor sentence length shorter duration in longer sentences is expected.

3. Results

3.1. Information structure

Figure 1 and 2 present the results for the analysis of duration in different information structure conditions. Mean duration pooled across speakers in ms with 95% confidence intervals is displayed on the x-axis, and the five different conditions are given on the y-axis. The wide focus condition ('All-new') is considered to be the baseline condition with a mean duration of 283 ms for *Liehner* (Fig. 1) and 298 ms for *Mohn* (Fig. 2).

Comparing the narrow focus condition ('Narrow') with the baseline duration increases by 14 ms for *Liehner* (Fig. 1) and 7 ms for *Mohn* (Fig. 2). For the contrastive focus condition ('Contrast') duration increases by 30 ms for *Liehner* (Fig. 1) and 22 ms for *Mohn* (Fig. 2) compared to the baseline.

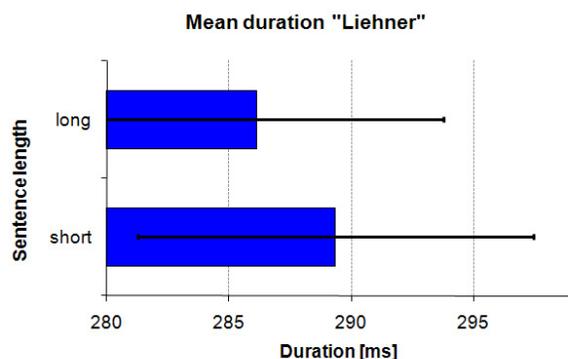


Figure 3: Mean duration pooled across 10 speakers in ms with 95% confidence intervals for the target word *Lieber* given for short and long sentences.

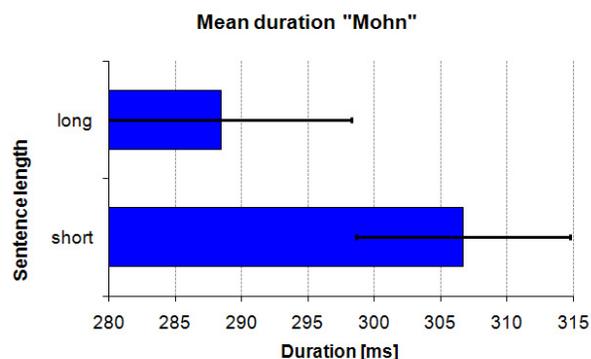


Figure 4: Mean duration pooled across 10 speakers in ms with 95% confidence intervals for the target word *Mohn* given for short and long sentences.

Comparing givenness conditions with the baseline the two givenness conditions differ in their durational patterns. A pre-focally given constituent is on average 22 ms shorter for *Lieber* (Fig. 1) and 33 ms, for *Mohn* (Fig. 2). In contrast, a postfocal constituent is on average 7 ms shorter for *Lieber* (Fig. 1), but 12 ms longer for *Mohn* (Fig. 2). In general, a postfocal given constituent is longer than a prefocal one, the postfocal given constituent being roughly equivalent in duration to the baseline.

The results which are presented for the disyllabic target word are equivalent to measurements taken only on the first (stressed) syllable of that target word.

3.2. Sentence length

Figure 3 and 4 present the results for the analysis of duration for two different sentence length conditions. Mean duration pooled across speakers in ms with 95% confidence intervals is displayed on the x-axis, and the two different conditions of sentence length are given on the y-axis.

Both figures show that the target constituent is longer in shorter sentences. However, this result is statistically only borne out for the monosyllabic item *Mohn* (Fig. 4). The disyllabic target word does not show a significant difference in duration between short and long sentences. Yet, the data show the same trend as for the monosyllabic target word.

The monosyllabic target word is on average 18 ms shorter when uttered in a longer sentence (Fig. 4), while the disyllabic one only about 3 ms (Fig. 3).

3.3. Position

Figure 5 and 6 present the results for the analysis of duration for two different position conditions. Mean duration pooled across speakers in ms with 95% confidence intervals is displayed on the x-axis, and the two different conditions of sentence position are given on the y-axis.

With respect to the factor position, the two target words behave different. While for the disyllabic target word no significant difference in duration could be established, the monosyllabic target word differs significantly between these two conditions. The duration of *Mohn* is on average 28 ms shorter when uttered in longer sentences (Fig. 6).

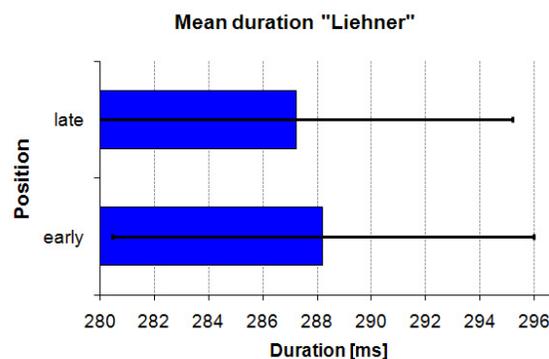


Figure 5: Mean duration pooled across 10 speakers in ms with 95% confidence intervals for the target word *Lieber* given for an early and late occurrence of the target.

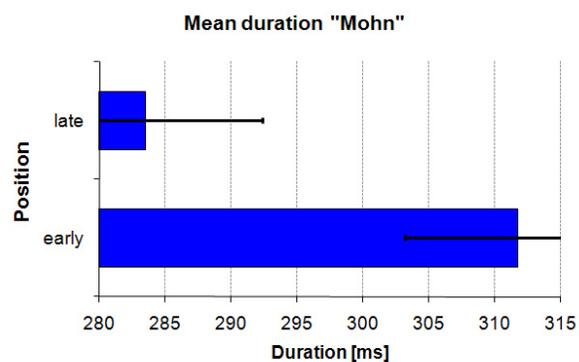


Figure 6: Mean duration pooled across 10 speakers in ms with 95% confidence intervals for the target word *Mohn* given for an early and late occurrence of the target.

4. Discussion and conclusions

The predictions of the present production study are by and large borne out. Based on previous work on duration and focus prominence [1, 4, 8, 9, 10, 11, 15] focused constituents were expected to be longer than compared to the baseline of an all-

new sentence. The data presented here are significantly longer for narrow and for contrastive focus ($p < .05$ for both target words). This confirms the prediction of an increase of duration in case of focus prominence.

The raw data suggest that contrastive focused constituents are longer than narrowly focused ones. However, no significant difference between a narrow and a contrastive focus was found. Based on this finding we can conclude that focus as an information structural category has an influence on the durational patterns, yet semantically different focus types [16] do not show categorical differences in the phonetic implementation. From a phonological point of view, thus, we assume a distinction between focus (no matter of narrow or contrastive focus) and no particular focus (all-new sentences).

As for givenness, the data do not show a coherent pattern for prefocal and postfocal givenness. Based on the literature a decrease in duration was expected for given constituents [11]. However, this is only true for prefocal given constituents. Postfocal ones are about the same in length as in the baseline condition. This distinction has not been reported yet, and the reason for relatively long postfocal constituents is by no means clear.

Apart from the effect of information structure two further effects on the duration of constituents have been investigated, the effect of the length of the sentence and, of position in the sentence on the target word. The former one has to our knowledge not been investigated so far, though it is expected that constituents in a longer sentence overall decrease in duration. The present data proves this effect, yet only for the monosyllabic target word. For the disyllabic target word, only a trend in the same direction can be observed. From the present data set it is not clear to what extent the length of a sentence influences the durational patterns.

As for the effect of position of the target word in the sentence it has been claimed that the duration of constituents is longer if they occur early in the sentence as compared to late [11, 12]. The data of the present study has confirmed this claim only partly. The monosyllabic target word was affected by position while the disyllabic target was not. It is unclear why the effect of position only affects the monosyllabic target word in this study. The only obvious difference between the carrier sentences of the two target words is their number of syllables. The target word *Liehner* was embedded in a sentence containing 20 syllables, while *Mohn* was embedded in a sentence containing 13 syllables. It may be the case that the higher number of syllables, i.e. longer overall sentence duration, obscured the durational decrease of the target word. Another explanation would be that phrasing longer sentences into two intonation phrases might cause this effect; however, this cannot be confirmed by our data since only seven out of 200 long sentences have been uttered in two IPs.

Overall, this study has confirmed the effects of information structure on duration while the effects of sentence length and position of a target word in the sentence need further investigation. However, this study has provided quantitative data that can be used as a starting point for manipulating speech data in order to perceptually test the role of duration for the recognition of focus prominence.

Acknowledgments

The research presented here was funded by the Deutsche Forschungsgemeinschaft (DFG), through the project D5 of the research centre SFB 632 on Information Structure at the University of Potsdam. Thanks to Anja Gollrad, Sophia Jähnigen, Gerrit Kentner and Bernadett Smoliboeki for assistance.

5. References

- [1] Allwood, J., 1974. Intensity, Pitch, Duration and Focus. *Logical Grammar Reports 11, Department of Linguistics, University of Gothenburg*, 1-17.
- [2] Allerton, D. J., 1978. The notion of "givenness" and its relations to presupposition and to theme. *Lingua*, 44, 2-3, 133-168.
- [3] Baumann, S., 2006. *The Intonation of Givenness. Evidence from German*. Tübingen: Niemeyer.
- [4] Baumann, S.; Grice, M.; Steindamm, S., 2006. Prosodic Marking of Focus Domains - Categorical or Gradient?. In Hoffmann, R.; Mixdorff, H. (eds) *Proceedings of Speech Prosody 2006, Dresden, Germany*, Dresden: TUDpress, 301-304.
- [5] Beaver, D.I.; Clark, B.Z.; Flemming, E.S.; Jäger, T.F.; Wolters, M.K., 2004. When semantics meets phonetics: Acoustical studies of second occurrence focus. Ms., Stanford University.
- [6] Boersma, P.; Weenink, D., 2007. Praat: doing phonetics by computer (Version 4.5) [Computer program], Retrieved from <http://www.praat.org/>.
- [7] Büring, D., 1996. *The 59th Bridge Accent - On the meaning of Topic and Focus*. SFS-Report-05-96, Seminar für Sprachwissenschaft, Tübingen University.
- [8] Cooper, W.E.; Eady, S.J.; Mueller, P.R., 1985. Acoustical Aspects of Contrastive Stress in Question-answer Contexts. *Journal of the Acoustical Society of America*, 77, 2142-2156.
- [9] Eady, S.J.; Cooper, W.E., 1986. Speech Intonation and Focus Location in Matched Statements and Questions. *Journal of the Acoustical Society of America*, 80, 402-415.
- [10] Eady, S.J.; Cooper, W.E.; Kloouda, G.V.; Mueller, P.R.; Lotts, D.W., 1986. Acoustical Characteristics of Sentential Focus: Narrow vs. Broad and Single vs. Dual Focus Environments. *Language and Speech*, 29, 233-251.
- [11] Féry, C.; Kügler, F., under revision. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*.
- [12] Heldner, M.; Strangert, E., 2001. Temporal effects of focus in Swedish. *Journal of Phonetics*, 29, 329-361.
- [13] Ishihara, S.; Féry, C., 2006. Phonetic Correlates of Second Occurrence Focus. *Proceedings of NELS 36. GLSA, University of Massachusetts, Amherst*.
- [14] Jäger, T.F., 2004. Only always associates audibly. Even if only is repeated. The prosodic properties of second occurrence focus in English. Ms., Stanford University.
- [15] Jun, S.-A.; Kim, H.-S., 2007. VP focus and narrow focus in Korean In Trouvain, J.; Barry, W.J. (eds) *Proceedings of the 16th ICPHS*, 1277-1280.
- [16] Krifka, M., 2007. Basic Notions of Information Structure. In Féry et al. (eds) *The notions of information structure*. ISIS 6, 13-55. Potsdam University.
- [17] Selkirk, E., 2004. Bengali Intonation revisited: An optimality theoretic analysis in which FOCUS stress prominence drives FOCUS phrasing. In Lee, Gordon & Bring (eds) *Topic and Focus: A cross-linguistic perspective*. Kluwer, 217-246.
- [18] Truckenbrodt, H., 1995. *Phonological phrases: their relation to syntax, focus, and prominence*. PhD thesis, MIT.