The effect of priming on the correlations between prominence ratings and acoustic features

Denis Arnold[1], *Petra Wagner* [2], *Bernd Möbius* [1]

[1]Institute of Communication Sciences, University of Bonn, Germany
[2]Faculty of Linguistics and Literature, University of Bielefeld, Germany
dar@ifk.uni-bonn.de, bmo@ifk.uni-bonn.de, petra.wagner@uni-bielefeld.de

Abstract

In previous research we showed that the priming paradigm can be used to significantly alter the prominence ratings of subjects. In that study we only looked at the changes in the subjects' ratings. In the present study, we analyzed the acoustic parameters of the stimuli used in the priming study and investigated the correlation between prominence ratings and acoustic parameters. The results show that priming has a significant effect on these correlations. The contribution of acoustic features on perceived prominence was found to depend on the prominence pattern. If a dominantly prominent syllable is present in a given utterance, f0 and intensity contribute most to the perceived prominence, while duration contributes most when no syllable is dominantly prominent.

Index Terms: syllable prominence, priming, acoustic correlates

1. Introduction

Syllable prominence has been widely studied. Fry [1] was one of the first to investigate the relation between f0, intensity and duration on the perception of stress. He argued that there must be a complicated interaction of these cues. The studies by Fant and Kruckenberg [2] gave the first evidence that prominence ratings are not only based on acoustic features of the speech signal. They found that subjects' ratings of their own inner voice while reading was close to the ratings of a speech signal comprised of identical text material. Streefkerk [3] argues that the "listener combines bottom-up information from the speech signal with his expectation of prominence on the basis of his knowledge of the language (top-down information)." Studies by Eriksson [4] and Wagner [5] gave further evidence that top-down processes affect the rating of syllable prominence. Eriksson [6] found that the ratings of a linguistic model predicting syllable prominence on the basis of written text came closer to the mean rating than the average subject listening to the actual signal. He argued that one cannot tell how prominence ratings are made by the subjects. In a follow-up study [4] Eriksson used listeners without competence in the presented language to gain insight into the effect of language knowledge on the perception of syllable prominence. The study revealed some differences in the judgment between native and non-speakers. Wagner [5] conducted a set of experiments altering language competence and speech rate. She concluded that listeners "use introspection as a fallback strategy if no reliable acoustic cues to prominence are present, as it is the case in very fast speech. Compared to native speakers, non-natives tend to rely more on acoustic cues."
Both studies [4, 5] varied the usability of bottom-up information for the subjects to gain insights about the contribution of top-down processing on the rating of syllable prominence. This motivated an experiment in which the expectations of the subjects were manipulated directly. Here the psychological priming paradigm was used to manipulate the expectation of syllable prominence. Tillmann [7] reviews studies in which the sound perception of subjects was successfully manipulated, employing the concept of priming. This gave further support to the present approach. Thus, we carried out two experiments, one using a design where within-subject manipulation was applied and one employing a design with between-subject manipulation. In a previous publication we show that priming leads to significant differences in the subjects' syllable prominence ratings. In the present study we had a closer look at the acoustic features and the impact of priming on the correlations between ratings and acoustic features, in particular f0, intensity and syllable duration.

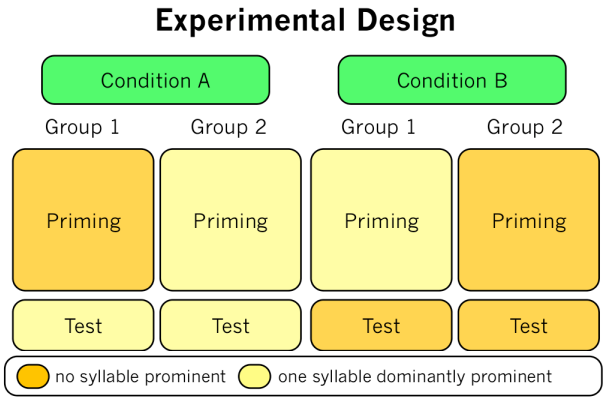

Figure 1: *Design of the experiment.*

2. Hypothesis

We expected to find higher correlations between prominence ratings and acoustic features in those groups of subjects for which the prominence patterns of the priming had the same structure as the prominence pattern of the stimuli. We expected lower correlations in those groups where priming works against the acoustic prominence pattern of the stimuli.

3. Experiment

3.1. Design

There were two conditions with two groups per condition (Fig.1). Each group was primed with a different set of prim-

ing material but exposed to the same test sentences. The ratings of the test sentences of both groups of one condition were compared. In condition A all test sentences contained one syllable that is dominantly prominent, i.e. they contain contrastive stress deviating from citation form. In group 1 all priming sentences have a neutral realization, where no syllable is dominantly prominent. In group 2 the priming sentences have the same prosodic pattern as the test sentences. In condition B all test sentences have a neutral realization. The priming sentences for group 1 contain one syllable which is dominantly prominent. The priming sentences for group 2 have the same neutral realization as the test sentences.

3.2. Speech Material

Each condition had four test sentences with a length of ten syllables each. There was a set of four priming sentences for each test sentence. Each priming sentence had exactly the same syntactic structure and a similar semantic structure as the test sentence. All priming sentences belonging to a given test sentence shared the same prosodic pattern. However, this pattern differed from the pattern of the pertinent test sentence in the accentuation of one particular syllable for one group within a condition. For the other condition the priming sentences had the same prosodic pattern as the test sentence. The same set of test sentences was used for each group of a given condition. Each group had its own set of priming sentences. This resulted in 2*4 = 8 test sentences and 2*4*4 = 32 priming sentences. The following listing gives an example for a test sentence and its priming sentences. Italic typesetting indicates the dominantly prominent syllable in the test sentence:

test sentence: Die *jun*ge Frau geht in das rote Haus.
priming sentence 1: Der alte Mann stieg in den vollen Bus.
priming sentence 2: Das kleine Kind ging in das kleine Haus.
priming sentence 3: Die alte Frau steigt in den leeren Bus.
priming sentence 4: Der junge Mann geht in das gelbe Haus.

The speech material was spoken by a trained speaker and was not manipulated in order to keep the material natural. The speech signals were recorded in a sound-treated studio and stored as 16-bit, 44.1 kHz wave files.

3.3. Perception Experiment

72 subjects were asked to rate the syllable prominence of 20 sentences. Most of the subjects participated in the test at our lab's computer pool. The other subjects participated at their homes. The experiment was conducted by means of a software coded for the study by the authors. All instructions were presented on the computer screen. The stimuli were presented via headphones and subjects were asked to rate the prominence my means of one slider per syllable, presented on the computer screen. An orthographic representation of each syllable was presented above the corresponding slider (Fig.2).
Subjects were allowed to listen to a given stimulus again if necessary. They were instructed to listen again only if the signal was not presented correctly - which could be the case due to network problems - or if they felt the strong need to listen to the stimulus again to give a valid rating. They were instructed to keep the number of repetitions as low as possible. This point is crucial because listening to a stimulus again can interfere with the intended priming.

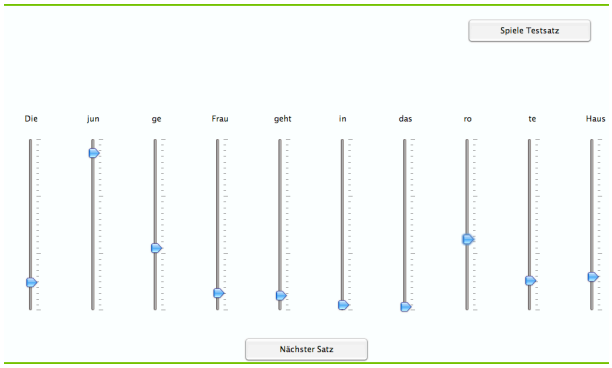

Figure 2: *Rating for the sentence "Die Frau geht in das rote Haus." - The woman enters the red house. - with the graphical user interface. Each syllable is rated with its own slider. There is one button to repeat the signal "Spiele Testsatz" and one to complete the task and move on "Nächster Satz".*

3.4. Rating Scale

Streefkerk [3] argued that an optimal rating scale for syllable prominence has not been found yet. Jensen and Tøndering[8] discussed which scale should be used for prominence rating. They compared 2-point, 4-point and 30-point scales for the rating of syllable prominence. They argued that a 30-point scale is hard to handle for non-expert listeners, it is more time consuming and the subject has to listen to signals more often. Eriksson [4, 6] used an open scale with sliders and prominence was rated as the percentage of the complete range of a slider. We decided to mix the approaches and use sliders and transformed the values into a 31-point scale.
The subjects were instructed to put the slider all the way up, if they believed that the syllable was maximally prominent and all the way down if they believed the syllable was minimally prominent. The subjects were encouraged to make use of the full range of the rating scale. No numbers indicating the value of the rating were displayed next to the sliders. Instead, 30 tick marks helped to interpret the position of a slider while rating. The position of a slider was transformed into a numerical value ranging from 0-30 using the standard routines of the J-Slider Class of Java-Swing.

3.5. Results of the perception study

Following the definition of syllable prominence by Wagner [9], we compared the difference D_n between the prominence rating of the manipulated syllable P_n and their immediate neighbors P_{n-1}, P_{n+1}.

$$D_n = \frac{2P_n - P_{n+1} - P_{n-1}}{2} \quad (1)$$

We compared D_n of the manipulated syllable in both groups of each condition. We mostly found significant differences in the ratings in both conditions (Table 1). In group 2 we found that the ratings of the unmanipulated syllables were lower than in the other group for condition A. Fig. 3 and Fig.4 show an example of the mean rating for test sentences and priming sentences for both groups of condition A, and Fig.5 and Fig.6 show an example for condition B.
We found that most of our subjects had no problems using the rating scale. Taking all ratings together the full range was used by our subjects. A high number of ratings are equal to zero.

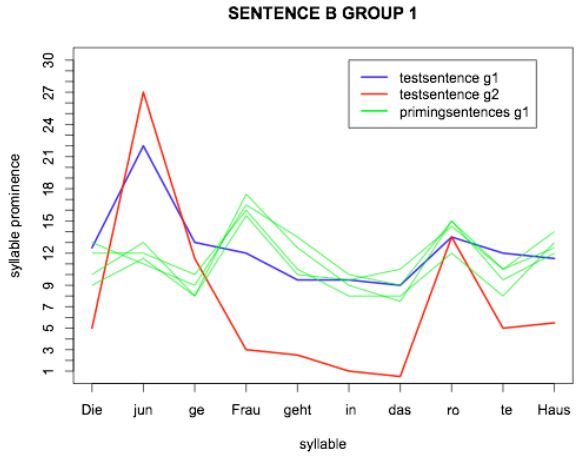

Figure 3: *Syllable, prominence rating, condition A. The red line shows the test sentence for group 2, the blue the result for the same sentence in group 1. The green lines show the ratings of the priming material in group 1. The second syllable "jun" is manipulated. Priming causes a lower rating in the manipulated second syllable in group 1.*

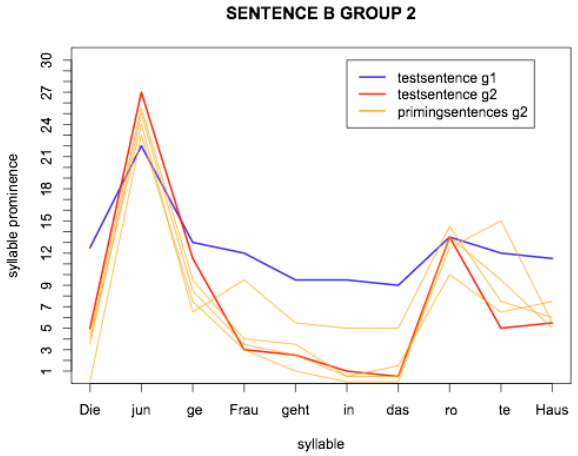

Figure 4: *Syllable, prominence rating, condition A, same sentence as in Fig. 3. The orange lines show the ratings of the priming material in group 2. Lower ratings in syllables 4-7 are in line with the ratings of the priming material.*

We believe that this effect is due to the experimental design, as the sliders were initialized on zero for each trial. Since there was no option in the java class to initialize with the first click on a slider, we decided to initialize on zero rather than in the middle of a slider, which would have created a large number of "15" values, which in turn would have been harder to distinguish from intended ratings.
Subjects listened repeatedly 1.31 times on average per sentence. This number is lower than the number of repetitions for all the rating scales evaluated by Jensen and Tøndering [8].

3.6. Analysis of Acoustic Cues

All test sentences were manually labeled on the syllable and phone level. The duration of the whole syllable, as well as the duration of the nucleus was extracted from the label files and the intensity was computed. We used the ESPS tool get_f0 [10] to estimate f0. We calculated mean f0 for each syllable ($f0mean$) as well as the difference between the maximum f0 of the syllable and the mean f0 for the whole utterance ($f0diff$).

Table 1: *Results of the perception experiment. The t-test has been computed for D_n of the manipulated syllable of each sentence.*

Condition A	Condition B
Sentence 1 t(33.65) = -3.5608, p $<$ .01	Sentence 1 t(33.529) = 2.0652, p $<$ .05
Sentence 2 t(27.353) = -2.1909, p $<$.05	Sentence 2 t(31.096) = -0.0365, p = .5144
Sentence 3 t(28.297) = -1.6834, p = .05165	Sentence 3 t(31,737) = 2.156, p $<$ .05
Sentence 4 t(24.103) = -1.8616, p $<$ .05	Sentence 4 t(32.835) = 0,7846, p = .2192

4. Results

All statistical computations were carried out by means of R [11]. We computed correlations (Spearmans ρ) for the subjects' ratings and f0, syllable duration, and intensity. All correlations were computed on the basis of all syllables from all four test sentences of each group.
The correlations differ strongly between the two conditions (cf. Table 2 and Table 3) . While f0 is the best predictor for prominence in condition A, there are no significant correlations for rated prominence and f0 in condition B at all. The role of syllable and nucleus durations also differs for both conditions. Only intensity receives significant correlations in both conditions.

5. Discussion and Conclusion

Our results show that the contribution of f0 and duration to the perception of syllable prominence differs depending on the prominence pattern of the stimuli. While in condition A, where each presented test sentence contained a dominantly prominent syllable, f0 and prominence ratings show correlations in line

Table 2: *Correlations between x and y Condition A. Spearmans ρ and the corresponding p-values*

Acoustic Feature	Ratings Group 1	Ratings Group 2
f0mean	0.291 p = 0.06	0.561 p $<$ 0.01
f0diff	0.382 p $<$ 0.05	0.651 p $<$ 0.01
duration syllable	-0.012 p = 0.93	0.143 p = 0.37
duration nucleus	0.215 p = 0.18	0.265 p = 0.09
intensity	0.569 p $<$ 0.01	0.695 p $<$ 0.01

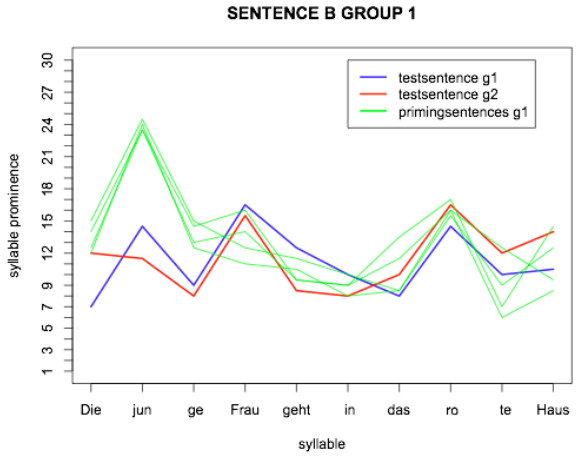

Figure 5: *Syllable, prominence rating, condition B. The red line shows the test sentence for group 2, the blue the result for the same sentence in group 1. The green lines show the ratings of the priming material in group 1. The second syllable "jun" is manipulated*

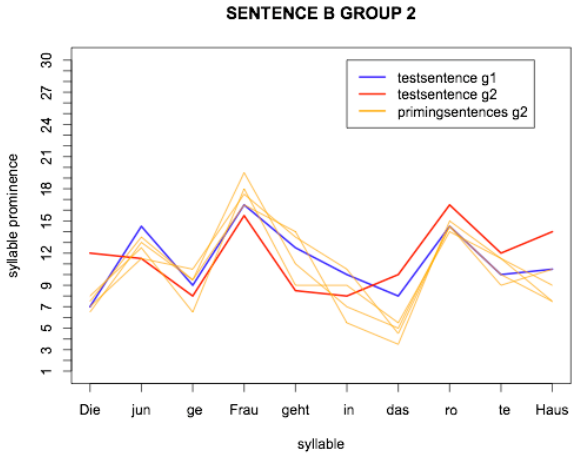

Figure 6: *Syllable, prominence rating, condition B. The same sentence as in Fig. 5. The orange lines show the ratings of the priming material in group 2*

with findings from other studies, there are no significant correlations between f0 and prominence rating in condition B, where no syllable was dominantly prominent in the presented test sentences. Duration shows the opposite behavior. There are significant correlations in condition B, but not in condition A. There are significant correlations between intensity and prominence rating in both conditions. The correlation between intensity and perceived prominence is much higher in condition A. The prediction that the correlations will be lower for those groups of subjects where the priming effect works against the prominence pattern of the stimuli is supported by our data.
The parameters contributing to the perception of syllable prominence differ very much depending on the prosodic structure of the stimulus. If a sentence contains a dominantly prominent syllable, f0 and intensity have a great influence on the perceived prominence, whereas duration does not contribute to it. If the sentence has no dominantly prominent syllable, the duration of the syllable nucleus and intensity contribute to the perceived prominence, whereas f0 has no effect. If our findings are corroborated by further studies, they may contribute to better prominence prediction algorithms. We plan to do a follow-up study where we systematically alter the position of the dominantly prominent syllable to investigate wether some syllable positions are more sensitive to priming effects than other. We plan to refine the routine for prominence rating by using sliders without initializing a value. All our studies only cover the German language. It would be interesting to replicate the study for other languages.

Table 3: *Correlations between x and y Condition B. Spearmans ρ and the corresponding p-values*

Acoustic Feature	Ratings Group 1	Ratings Group 2
f0mean	0.011 p = 0.94	0.098 p = 0.54
f0diff	0.046 p= 0.77	0.096 p = 0.55
duration syllable	0.223 p = 0.16	0.385 $p < 0.05$
duration nucleus	0.416 $p < 0.01$	0.501 $p < 0.01$
intensity	0.218 p = 0.17	0.333 $p < 0.05$

6. References

[1] Fry, D. B. (1958), Experiments in the Perception of Stress. Language and Speech, 1, 126-152.

[2] Fant, G und Kruckenberg, A. (1989), Preliminaries tothe study of Swedish prose reading and reading style. STR-QPSR, 2/1998 KTH, Stockholm, 1-83.

[3] Streefkerk, B. (2002), Prominence - Acoustic and lexical/syntactic correlates, Utrecht: LOT.

[4] Eriksson, A., Grabe E. und Traunmüller, H. (2002), Perception of syllable prominence by listeners with and without competence in the tested language. Proceedings Speech Prosody 2002, Aix-en-Provence, 275-278.

[5] Wagner, P. (2005), Great Expactations - Introspective vs. Perceptual Prominence Ratings and their Acoustic Correlates, Proceedings of Interspeech 2005, Lisbon,2381-2384.

[6] Eriksson, A., Thunberg, G. und Traunmüller, H. (2001), Syllable prominence: A matter of vocal effort, phonetic distinctness and top-down processing. Proceedings of Eurospeech 2001, 399-402.

[7] Tillmann, B. and Bigand, E. (2002), A comparative review of priming effects in language and music, P. McKevitt, S. ONuallain and C. Mulvihill (Eds.) Language, Vision and Music, John Benjamins : Amsterdam, Philadelphia, 231-240.

[8] Jensen, C. and Tøndering, J. (2005), Choosing a Scale for Measuring Perceived Prominence, Proceedings of Interspeech 2005, Lisbon, 2385-2388.

[9] Wagner, P. (2002), Wahrnehmung und Vorhersage deutscher Betonungsmuster, Universität Bonn. PhD-Thesis.

[10] Talkin, D. (1995). A Robust Algorithm for Pitch Tracking (RAPT). In Kleijn, W. B. and Paliwal, K. K. (Eds.), Speech Coding and Synthesis. New York: Elsevier. 495-518.

[11] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org