Prosody in hate speech perception: A step towards understanding the role of implicit prosody

Oliver Niebuhr
Centre for Industrial Electronics, University of Southern Denmark, Sønderborg, Denmark
olni@sdu.dk

Abstract
When speaking out hate-speech posts aloud, the prosody that speakers use can drastically change how listeners rate these posts in terms of personal (un)acceptability and consequences for the originator. But, does this also apply to the silent (implicit) prosody that we hear in our head while reading a post? Paving the way to answering this question, the present paper investigated if readers’ explicit prosody is connected to their hate-speech ratings. Results provide evidence for this connection and, moreover, show that migration background plays a role for both the prosody and evaluation of hate-speech posts.

Index Terms: hate speech, German, prosody, perception.

1. Introduction
There is increasing evidence that hate speech is a growing problem of our time, not just in terms of its amount, but also with respect to its psychological and mental consequences for the addressees – both short-term and long-term, see [1]. The harmful effects of hate speech range from psychological distress, depression and suicide attempts in individuals to the disintegration and political radicalization of societies [1,2] – with all their own potentially fatal consequences that happened, e.g., to the Rohingya Muslim minority in Myanmar in 2018 or to the Tutsis in Rwanda in 1994. The language that was used by the Third Reich and that eventually paved the way to the Holocaust [3] would today probably also be considered an example of hate speech. For many years we have seen a sharp increase in hate speech, which accelerated again during the COVID-19 crisis [4,5]. Psychologists already speak of the “hate speech epidemic” [6]. According to a recent survey [7], 79 % of German Internet users stated that Internet comments have become more aggressive and that they have already witnessed hate speech on the Internet. Almost 20 % of all respondents even stated that they have been the victim of hate speech themselves. For men and under 30-year-olds, this proportion was still significantly higher, i.e. up to 37% [7].

The present study is part of a line of research that investigates, in a cross-linguistic German-Danish framework, how prosody shapes the production and perception of hate speech, see [8] for an overview. Our previous research showed that, e.g., speaking out hate-speech posts aloud can make them more severe for perceivers [8,9,10]. This applies in particular to those posts that contain verbal hate-speech indicators, such as Holocaust references, imperatives or pejorative, figurative language. By contrast, written posts that lack such clear verbal indicators can even become less severe for perceivers when they are spoken out aloud. This is especially true for those posts in which the communicative function of the core hate-speech element is mainly conveyed by prosody, as is the case, e.g., with rhetorical questions or irony/sarcasm [9]. An ironic prosody even seems capable of turning posts rated as instances of hate speech in written language into non-hate speech in spoken language [11].

Furthermore, results of previous studies suggest that there is no prosodic pattern specific to spoken hate speech [8,12]. Rather, it is the individual semantic-pragmatic content that shapes the prosody of a hate-speech utterance. Yet, there are still significant correlations between prosodic parameters and how listeners rate spoken hate-speech stimuli in terms of personal (un)acceptability and consequences for the originator [9]. For German, e.g., this means that the lower the f0 minima and the slower the speaking rate in hate speech stimuli, the “louder” gets the perceivers’ call for consequences for the originator [9]. Lower values of HNR and Hammarberg Index (cf. “spectral emphasis” [13]) additionally increase the perceivers’ unacceptability ratings of the hate-speech stimuli [9,12].

Current works explain the increase in hate speech mainly with reference to the “online disinhibition effect” [14,8]: The supposed anonymity on the Internet [15], the lack of a specific addressee, and the asynchrony in communication all make it easier for the originators of hate speech posts to internally distance themselves from their own statements and think of their addressees less concretely as humans. Further explanations are an increased social sensitivity to hate speech (that causes a higher reporting rate) and the still vague, changeable definition of hate speech, due to which statements in everyday use are often referred to as hate speech, although they go beyond the original UN definition of hate speech as “any kind of communication in speech, writing or behavior that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality [...]” [16]. We refer to this UN definition also in the present paper.

In conjunction with the current explanations, we wonder whether the essential role of prosody in producing and, especially, perceiving hate speech could be another reason for why hate speech is becoming a growing problem. Specifically, we want to test the assumption that another important growth factor for the occurrence of hate speech could be a discrepancy in prosody between those who produce hate speech and those who perceive hate speech. In the extreme case, this can mean that, e.g., those who write a comment on the Internet have a different, less hateful prosody on their mind than those who read the comment and, thus, perceive the same comment as more hateful. We will refer to this as the “prosodic discrepancy hypothesis”. Supporting evidence for this assumption would be relevant in that it could serve as a lever against hate speech, e.g., in the form of a prosody-based awareness campaign aimed at all commentators/authors on the internet. In order to be able to search for supporting evidence for the “prosodic discrepancy hypothesis”, a number of assumptions must be met.

- (1) When reading aloud a written stimulus, there is a connection between the prosodic characteristics of that stimulus and its hate-speech evaluation by the reader [17].
• (2) Assumption (1) also applies to implicit prosody [18].
• (3) Assumption (2), in turn, requires that the silent (implicit) prosody that emerges in reading text goes beyond the “default” (matter-of-fact) implicit prosody (DIP) [19] that is shaped by and aimed at syntax and semantics alone. The traditional “Implicit Prosody Hypothesis” of Fodor [18] postulates the existence of this DIP only. However, many recent results indicate that implicit prosody is indeed more than DIP and includes prosodic colorings related to, e.g., arousal, expressiveness and direct/indirect speech [19,20].
• (4) Changes in the sense of precondition (3) can be measured validly and reliably.

The starting point of this research is obviously complex. Therefore, the present paper will only take the first step and deal with assumption (1). The experimental method used to this end is described in more detail below. It is based on a combination of read-aloud task and subsequent rating of individually elicited statements in a 2D rating space that was successfully used in many previous studies [8,9,11].

2. Method

2.1. Stimulus Material

Ten excerpts (15-20 s) of speeches given in 2018-2019 in the German Bundestag on the subject of migration served as stimulus material for the experiment. All excerpts were more or less controversially discussed in the (social) media. Of the 10 excerpts, however, there was only one that can be considered a clear hate-speech example in terms of the UN definition. The example was uttered in 2018 by the then AFD parliamentary group leader in the Bundestag Alice Weidel, see Figure 1.

Figure 1: Alice Weidel (AFD parliamentary group leader) during her Bundestag speech in 2018. Source: https://www.youtube.com/watch?v=ZEGj1T0pnR0.

“[…] But I can tell you: Burqas, headscarf girls, alighted knife men and other scapegraces will neither ensure economic growth for our Germany nor will they secure our welfare system. Who pays your considerable pensions […] after all? Your immigrant gold pieces? You don't seriously believe that!” (‘[…] doch ich kann ihnen sagen: Burkas, Kopftuchmädchen, alimintierte Messer-Männer und sonstige Taugenichts werden unserem Deutschland das Wirtschaftswachstum vor allem den Sozialstaat nicht sichern. Wer zahlte denn ihre staatlichen Pensionen […]? Ihre eingewanderten Goldstücke etwa? Das glauben Sie doch nicht im Ernst!’). Supporting the classification of this utterance sequence as hate speech, Alice Weidel received an official reprimand from the then Bundestag President Wolfgang Schäuble for the discriminatory nature of her choice of words.

Weidel’s speech excerpt given above constitutes the target stimulus of the experiment. Firstly, this is because it is an authentic hate-speech example that comes from a German politician who is widely known as a person and for her political views. Secondly, the excerpt is long enough to carve out its prosodic characteristics, also in contrast to those of other (reference) stimuli. Thirdly, the excerpt is an example of those cases in which the author/speaker does not separate the hate-speech section prosodically from the surrounding, less acute sections of the speech; here: the one minute after the excerpt. This is summarized in Table 1 for seven prosodic parameters measured manually with PRAAT; p-values refer to t-tests per parameter based on measurements per prosodic phrase in the hate-speech section (N=13) and the surrounding speech (N=30). Thus, in view of the “prosodic discrepancy hypothesis”, the selected example is an ideal starting point to investigate whether other people, if not Alice Weidel herself, especially people with a migration background, do produce a prosodic difference between the target (hate-speech) stimulus and other reference stimuli – and whether such a difference coincides with a difference in the rating of target stimulus and reference stimuli.

Table 1: Similar prosody (p<0.05) in the hate-speech and reference sections of Alice Weidel’s speech.

<table>
<thead>
<tr>
<th>Prosodic parameter</th>
<th>Target Stim.</th>
<th>Sur. Speech</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean f0 minimum</td>
<td>178.8 Hz</td>
<td>166.5 Hz</td>
<td>0.52</td>
</tr>
<tr>
<td>mean f0 level</td>
<td>256.6 Hz</td>
<td>268.9 Hz</td>
<td>0.40</td>
</tr>
<tr>
<td>mean f0 range</td>
<td>11.3 s</td>
<td>12.5 s</td>
<td>0.65</td>
</tr>
<tr>
<td>mean (RMS) intensity</td>
<td>75.1 dB</td>
<td>76.2 dB</td>
<td>0.69</td>
</tr>
<tr>
<td>mean speaking rate</td>
<td>4.9 syl/s</td>
<td>5.1 syl/s</td>
<td>0.32</td>
</tr>
<tr>
<td>mean HNR</td>
<td>9.4 dB</td>
<td>11.0 dB</td>
<td>0.06</td>
</tr>
<tr>
<td>mean Hbg-index</td>
<td>14.2 dB</td>
<td>15.5 dB</td>
<td>0.33</td>
</tr>
</tbody>
</table>

For the experiment, the target stimulus and the other nine non-hate-speech reference stimuli (see Appendix) were transcribed, i.e. transformed into written stimuli. These written stimuli were then read aloud and then again transformed into spoken stimuli with subjectively matching prosodies by the participants.

2.2. Participants

The experiment was based on 68 participants. Half of these participants, 20 men and 14 women (21-77 years old), were born and raised in Germany and, accordingly, spoke German as their native language. Their parents, however, immigrated to Germany. Thus, 34 of our 68 participants had a 2nd generation migration background. This should make the target stimulus particularly explosive for them. The migrant sample was supplemented by an equally large sample of 34 German native speakers of the same age group (17 men, 17 women, 23-28 years old) but without a migration background.

2.3. Procedure

The experiment was carried out as an alternation between reading and rating task. Each stimulus was dealt with first in the reading task and immediately afterwards in the rating task. In the reading task, participants “performed” the 10 written stimuli, i.e. Alice Weidel’s target stimulus and the 9 reference stimuli in individually randomized orders. The stimuli were read from a LED screen. Participants were asked to give each stimulus “an individually appropriate vocal sound shape”. It was emphasized that there would be no right or wrong solution to this. Rather, the vocal sound shape given to each stimulus would solely depend on the individual reader’s interpretation.

After having read each stimulus, a mouse click brought the participants to the rating task for this stimulus. With a single click of the mouse in a 2D rating space (see Fig. 2), each stimulus was evaluated simultaneously along two gradient scales: “personal (un)acceptability” (x-axis) and “consequences for the originator” (y-axis). After this 2D hate-speech rating had been carried out, a further mouse click sent the participants back to
the reading task where they continued with reading (“performing”) the next written stimulus. The decisive element in this procedural framework was that both groups of participants – those with and those without a migration background – were further divided into equally sized subgroups of 17 people each, keeping the male/female percentage largely constant.

For one subgroup, the interpretation of the target stimulus as hate speech and thus a corresponding prosodic realization was stimulated. This stimulation was absent for the other subgroup. The presence/absence of this stimulation was implemented solely by the temporal order in which the participants saw the written stimulus and the name of the corresponding author (speaker) on the screen. The author-first (AF) subgroup saw first the name of the author on the screen in the reading task and then, 3 seconds later, under this name the author’s written stimulus that they were supposed to perform. By contrast, the author-last (AL) subgroup first saw and performed the written stimulus, and only after that appeared the name of the corresponding author underneath the written stimulus on the screen. Pilot tests suggested that simply through the name of Alice Weidel and her well known political attitudes the AF condition created an elicitation context that was able to stimulate a more hateful prosodic realization of the target stimulus.

All 68 participants took part in the experiment in individual sessions in the CIE Acoustics Lab of the SDU, where the stimulus readings were recorded under studio conditions, i.e. in a sound-attenuated booth. The same headsets were used for all participants at a constant, pre-adjusted loudness level. Using headsets also ensured valid intensity measurements due to a constant mouth-to-microphone distance.

2.4. Prosodic and statistical analyses

The selected stimuli were prosodically analyzed in terms of the same parameters that are also listed in Table 1 and for which significant correlations with the 2D stimulus ratings have been found in previous studies. The prosodic analysis was done automatically through ProsodyPro [21]. Obvious measurement errors or measurement gaps were manually corrected, though. To normalize for sex-specific differences in F0 measurements, all values were converted into semitones (st) relative to 100 Hz or 200 Hz for male or female speakers, respectively.

For the 2D ratings, it was calculated in % how far to the right (x-axis) and how far up (y-axis) in the ratings space the mouse clicks were made per stimulus and participant.

The statistical analysis comprised three independent variables (each a 2-level fixed factor): Stimulus (target stimulus vs all reference stimuli), Recipient (participants with vs without a migration background) and Elicitation Context (AF vs AL). Note that Stimulus was a within-subject variable, whereas Recipient and Elicitation Context both were between-subject vari-ables. Participants were included as a random factor. Dependent variables were the measurements of the seven prosodic parameters and the % values for two (x/y) axes of the 2D rating space.

3. Results

We run a mixed-model repeated-measures MANOVA to test the effects that the three independent variables (fixed factors) Stimulus, Recipient, and Elicitation Context had on the seven prosodic measurements and the two (x/y) ratings. Results of the MANOVA showed a clear and significant main effect of the within-subjects factors Stimulus (F[9,56]=122.7, p<.001, \(\eta^2=.95\)). Moreover, Stimulus yielded the largest effect size. The two-between-subjects factor Recipient and Elicitation Context also came out clearly significantly, but with smaller effect sizes (Recipient: F[9,56]=25.8, p<.001, \(\eta^2=.81\); Elicitation Context: F[9,56]=21.4, p<.001, \(\eta^2=.78\)). All two-way interactions were significant, too; and the same applied to the three-way interactions. The two strongest interactions in terms of effect size both involved the factor Stimulus, i.e. Stimulus*Recipient (F[9,56]=15.6, p<.001, \(\eta^2=.72\)) and Stimulus*Elicitation Context (F[9,56]=4.2, p<.001, \(\eta^2=.41\)). The weakest interaction concerned the three-between-subjects factors, i.e. Recipient*Elicitation Context (F[9,56]=3.9, p<.001, \(\eta^2=.39\)).

Due to the limited space of a proceedings paper we cannot report the entire results statistics. Instead, we briefly summarize the key patterns and findings in the following that are also illustrated in Figures 3(a)-(d) and Figures 4(a)-(d).

Figures 3(a)-(b) illustrate the two different patterns of how prosodic measurements changed as a function of Stimulus and Recipient. In the first pattern (Fig.3a), participants with and without a migration background produced the reference stimuli with a similar prosody and then deviated from that reference prosody in opposite directions when producing the target stimulus. For example, speakers with a migration background increased their speaking rate, whereas speakers without a migration background decreased it. The same also applied to the results of F0 range.

The second pattern (Fig.3b) is one in which the participants with and without a migration background produced the reference stimuli with a similar prosody and moved away from that prosody in the same direction for the target stimulus, with migration-background speakers showing a stronger change than speakers without this background. This results pattern emerged for all dependent variables except for speaking rate and F0 range, including the x-axis and y-axis ratings. For the prosodic parameters besides HNR (Fig.3b), the change manifested itself as an increase in intensity, F0 minimum, and F0 level from the reference stimuli to the target stimulus. For the Hammarberg index, the change meant a decrease from the reference stimuli to the target stimulus.

Figures 3(c)-(d) display two examples of the interactions between Stimulus and Elicitation Context. We found only one general results pattern for this type of interaction: Besides the fact that the reference stimuli showed on average the same prosody across participants, the author-last (AL) condition yielded a small (Fig.3c) or no significant prosodic change (Fig.3d) from the reference stimuli to the target stimulus. By contrast, in the author-first (AF) condition we only found strong and significant prosodic changes (again more strongly pronounced for the migration-background speakers, which caused the above mentioned three-way interactions).

Figures 4(a)-(d) summarize the results of the rating task. We see that the target stimulus received more severe hate-
speech ratings than the reference stimuli in terms of both personal (un)acceptability (x-axis) and call for consequences for the originator (y-axis). In particular, unlike for all reference stimuli, only the target stimulus consistently obtained ratings in the upper right corner of the rating space (Fig. 2), i.e. values > 50%. Moreover, we see that participants with a migration background rated the target stimulus more extreme on both axes than participants without a migration background (Fig. 4a-b). More extreme target-stimulus ratings also emerged in the AF condition compared to the AL condition (Fig. 4c-d).

Table 2: Pearson correlations between ratings and parameters for participants without (top, blue) and with a migration background (bottom, green).

<table>
<thead>
<tr>
<th></th>
<th>fb min</th>
<th>fb range</th>
<th>fb level</th>
<th>int level</th>
<th>spk. rate</th>
<th>HNR</th>
<th>Hbg ind.</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-un/with</td>
<td>0.23</td>
<td>0.25</td>
<td>-0.11</td>
<td>0.30*</td>
<td>0.33*</td>
<td>-0.26</td>
<td>-0.57*</td>
</tr>
<tr>
<td>y-un/with</td>
<td>0.34*</td>
<td>0.39*</td>
<td>0.08</td>
<td>0.22</td>
<td>0.26</td>
<td>0.34*</td>
<td>0.58*</td>
</tr>
<tr>
<td>x-un/without</td>
<td>0.41*</td>
<td>0.67*</td>
<td>0.36*</td>
<td>0.38*</td>
<td>0.18</td>
<td>0.58*</td>
<td>0.23</td>
</tr>
<tr>
<td>y-un/without</td>
<td>0.55*</td>
<td>0.66*</td>
<td>0.32</td>
<td>0.36*</td>
<td>0.34*</td>
<td>0.49*</td>
<td>0.36*</td>
</tr>
</tbody>
</table>

4. Discussion

First of all, note that the aim of this paper was not to test the “prosodic discrepancy hypothesis”. Only assumption (1) on the way to testing this hypothesis was to be examined here: When reading aloud a written stimulus, there is a connection between the prosodic characteristics of that stimulus and its hate-speech evaluation by the (same) reader.

The three independent variables Stimulus, Recipient and Elicitation context were primarily means of generating sufficient prosodic variation in the production of the written stimuli for connections of this variation with the 2D hate-speech ratings to become visible. Thus, with a view to Alice Weidel and Table 1, it was not to be examined, whether or not readers with and without a migration background would prosodically differentiate the target-stimulus realizations from those of the reference stimuli. Nonetheless, the independent variables provided many new insights beyond searched connections themselves.

The latter clearly support assumption (1). There were significant differences in the prosodic realization of the target and reference stimuli, and these differences had systematic effects on how the stimuli were subsequently rated by the readers. In addition, there were correlations between prosodic parameters and hate-speech ratings both on the x- and on the y-axis of the 2D rating space. These correlations largely replicated those revealed by [9], but are mostly considerably stronger, perhaps because speaker and rater were the same person. Future studies can test that in detail. So, the decisive new finding of this study is that prosody not only affects how we rate hate-speech stimuli that we hear produced by others [8]. Also the prosody that we produce ourselves is linked to how we rate the corresponding utterance in terms of hate speech. Note that whether this is because our prosody heard through the auditory feedback loop [17] shapes these ratings or because the ratings are made first and then shape our prosody is irrelevant for our research aim, but represents a worthwhile question to address by others.

Over and above this main finding, the present study was – to the best of our knowledge – the first showing that people with a migration background, even an indirect (2nd generation) one, react more sensitively to xenophobic hate speech than people without a migration background, particularly in terms of the stimuli’s personal (un)acceptability. The qualitative differences in the prosodic implementation and rating of the target stimulus by participants with and without a migration background could moreover reflect a difference between hot anger and cold anger [9,22,23], respectively, or, more generally, a higher level of arousal or stress felt by the migration-background people.

5. Acknowledgements

The XPEROHS project (project number 95-16416) was funded by the Velux Foundations.
6. References


7. Appendix: Reference Stimuli (German)

“Die Flüchtlingskrise ist die größte Herausforderung, vor der Deutschland je stand. Größer als die Wiedervereinigung. Damals trafen Menschen aufeinander, die alle Deutsch sprachen und einen ähnlichen kulturellen Hintergrund hatten. Das ist heute anders.” Volker Butffer, CDU.

“Es wird an einer Brenz solcher Zuwanderung und damit an der Obergrenze für die Zuwanderung kein Weg vorbei führen.” Horst Seehofer, CSU.


“Es geht um Völkerwanderung, machen wir uns nichts vor. Wenn wir jedenfalls nicht bald reagieren, wird es uns am Ende allen auf die Füße fallen, egal, welches Parteibuch wir haben.” Bows Ramelow, Die Linke.


“Wir stehen vor einem fundamentalen Wandel. Unsere Gesellschaft wird weiter vielfältiger werden, das wird auch anstehen, mitunter schmerzhaft sein. Unser Zusammenleben muss täglich neu ausgehandelt werden.” Aydan Özoguz, SPD.


“Je mehr Bleibeperspektiven wir in den Krisenländern schaffen, umso weniger Flüchtlinge kommen nach Europa. Mit 50 Cent am Tag finanzieren wir heute Essen und Überlebensversorgung eines Flüchtlings etwa im Nordirak oder in Afrika. Bei uns fallen pro Flüchtling Kosten zwischen 50 und 100 Euro am Tag an.” Gerd Müller, CDU.