Types of hate speech in German and their prosodic characteristics

Jana Neitsch[1*] and Oliver Niebuhr[1]

[1]*Centre for Industrial Electronics, Mads Clausen Institute, University of Southern Denmark, Sønderborg/DK*
* neitsch@mci.sdu.dk

The present pilot study investigates spoken hate-speech items, produced by a professional German speaker after actual Twitter and Facebook posts and reflecting characteristic morpho-syntactic features of German hate speech. Acoustic-prosodic signal analyses reveal a considerable feature-specific variation in the production of hate speech. Thus, at least at the level of phonetics and spoken hate speech, no evidence was found that hate speech is a separate communicative function conveyed by a set of uniformly realized prosodic characteristics.

INTRODUCTION

Hate speech is becoming more and more of a concern in societies around the world [1]. The term "hate speech", however, remains highly controversial. There is a lack of consensus on its definition and impact, while the motivation and justification for its criminalization and regulation are inextricably linked to allowing freedom of speech on the one hand and protecting the human rights of equality and dignity on the other [2].

Given the pressure that hate speech exerts on the pillars of modern civilization, it is striking how little is known about the linguistic and communicative mechanisms underlying the expression and perception of hate speech. The Velux-funded XPEROHS project is intended to address some of these gaps. The project is based on a large German-Danish corpus of about 3.5 million real instances (posts) of hate speech on Twitter and Facebook, see [3,4] for further details on the compiled corpus and the goals of the cross-linguistic project. The present study is concerned with the German subset of this corpus, which consists of ~1.7 million hate-speech posts.

Although hate speech is primarily associated with written language, at least in the way it is experienced and discussed in the media [5], there is, of course, also hate speech in spoken language, for example in political discourse [6] or in connection with bullying on school yards and football fields [7]. The present study takes the first step into the field of spoken hate speech by asking the question, at an acoustic-phonetic level, whether hate speech has a uniform manifestation in the sense of a context-independent phonetic core and/or how large the phonetic variation of hate speech is with respect to major lexical and morpho-syntactic features that have emerged – for the German subset of the corpus – from a preceding linguistic processing of the Twitter and Facebook data.

For providing an empirical definition and scientific guidance in identifying and handling (e.g., reporting, deleting) hate speech, relatively or absolutely constant characteristics are required. At the level of speech communication, this would mean that phonetic parameters are not entirely shaped by the morpho-syntactic features underlying the functional and semantic variants of hate speech but also by the phenomenon of hate speech itself.

METHOD

The acoustic-phonetic analysis conducted here is based on German hate speech. A set of 12 hate-speech items (posts) were selected from the German part of the Twitter and Facebook corpus. They were all similarly short. That is, they consist of less than 25 words and include between 20 and 30 syllables. Moreover, semantically they are all directed against the minority group of immigrants. Six items use the more general and relatively neutral term "Ausländer" (foreigners) to refer to this group, the other six use the more specific, religious and – in this context – negatively connoted term "Muslime" (Muslims).

These 12 original hate-speech items have been modified or supplemented in terms of major lexical and morpho-syntactic features that crystallized as being characteristic of hate-speech items posted on Twitter and Facebook [3], thus leading to 6 feature conditions in addition to the original baseline condition: (1) rhetorical questions (RQs, e.g., "who wants/needs/would ever..."), (2) irony (IRO, modal particles expressing irony in German), (3) imperatives (IMP, directed either at the writer's peer group "we must..." or at the minority group "go/stop...."), (4) metaphors (MET, e.g., "Muslims are like..."), (5) holocaust reference (HOL), and (6) indirectness (IND e.g., "I am not against foreigners/ Muslims, but..."). In this way, 6 items sets have been created in addition to the original set. Thus, the total number of items was 84 (7x12).

The 84 hate-speech items were evaluated by two independent groups of people, an expert panel of researchers working on hate speech and a panel of ordinary users of Twitter and Facebook. The 84 items passed this pre-test in that they were consistently identified as hate speech. Moreover, both experts and non-experts rated all items similarly as to the degree of expressed hate.

The 84 items were realized, in line with [8], by a male native speaker of German (BP, 47 years old). He is a professional speaker as well as an experienced public-speaking trainer with an academic education and a PhD degree in phonetics and linguistics. BP is able to control the phonetic characteristics of his speech and to deliberately choose and produce phonetic patterns in order to create specific semantic-pragmatic effects.

On this basis, BP received the instruction to familiarize himself thoroughly with the set of 84 separate items and to practice their elicitation with different phonetic realizations in order to find one that suits each individual item and makes it sound like authentic and natural spoken hate speech. Note that BP was not informed about the 7 feature conditions included in the 84 items and according to a debriefing interview, he also did not become aware of these conditions during the speech-production task.

The speech-production task itself was conducted in the sound-proof booth of the Kiel Phonetics Lab [9]. Recordings were made with a Microtech Gefell M940 microphone at a 44.1 kHz sampling rate and a 16-bit quantization. BP produced the 84 items as isolated hate-speech utterances.

The acoustic-phonetic analysis of the 84 items included two pitch (f0) parameters, i.e. f0 mean (Hz) and f0 range (semitones, st), as well as two voice-quality parameters, i.e. mean HNR (Harmonics-to-Noise Ratio measured in dB) and the Hammarberg index (dB). Loudness was included in terms of mean RMS intensity (dB). Duration and tempo parameters were excluded due to the items' different morpho-syntactic make-up. Analysis was done automatically by means of PRAAT scripts written by [10,11,12].

RESULTS

Analyzed were 84 recorded stimuli that were produced by the phonetically trained speaker. Results were analyzed for mean f0, HNR, Hammarberg Index, and f0 range. P-values were adjusted using the Benjamini-Hochberg correction [13] to account for the multiple variables that were tested. All statistical models showed significant interactions between *target group* (*Muslims* vs. *foreigners*) and all of the *feature conditions* (see above; all p-values < 0.03). Interactions were split up according to the two targets and compared against the original items. Results for *foreigners* indicate that RQ, IRO, IND, IMP, and HOL were realized with a significantly lower mean f0 than their original base items (all p-values < 0.04). For *Muslims*, IRO and HOL items were realized with a significantly lower mean f0 (all p-values < 0.0003), whereas IMP, IND and RQ were produced with a higher mean f0, see Fig.1. HNR results for *Muslims* show significantly higher values compared to the original base items for all the feature conditions except for HOL (all p-values < 0.03), indicating less breathiness than in the original base items. For *foreigners*, IMP, IND and RQ were also produced with higher HNR values, indicating less breathiness than in the original base items (all p-values < 0.03). The Hammarberg index yielded higher values for RQ and HOL and lower values for MET (all p-values < 0.03) when the minority group concerned were *foreigners*. For *Muslims*, all feature conditions had higher Hammarberg indices than the original base items (all p-values < 0.04; see Fig. 2). Finally, f0-range results indicate lower values for

IMP, MET and RQ compared to original items (all p-values < 0.02) in the *foreigner* sub-set, but higher values in the *Muslim* subset (all p-values < 0.0003).

DISCUSSION

A main result of our data is that items addressing the target minority group of *foreigners* differ prosodically from those addressing target minority group of *Muslims*. In particular, the prosodic differences caused by the six feature conditions compared to the original items were a lot more pronounced when the hate speech targeted Muslims rather than foreigners. That is, the communicative functions that were involved in the tested hate speech items and embodied by the six lexical and morpho-syntactic features were conveyed more clearly and consistently, i.e. with a greater phonetic effort, in Muslim-oriented speech. The greater phonetic effort includes a greater vocal effect, see the higher HNR and Hammarberg-index values for the Muslim items. This makes sense given that Muslims are a smaller, more specific and negatively connoted target group than foreigners in general. A further implication of this finding is that the "intensity" of hate speech can vary at the phonetic level.

However, as was stated above, the nature of this variation points in the direction of a more or less strong signalling of the communicative functions (irony, rhetorical questions, etc.) conveyed by the different feature conditions. For example, rhetorical questions turned out to be the condition that differed most in its prosodic characteristics from the original base items, which is plausible insofar as this communicative function is, unlike indirectness and holocaust reference, primarily conveyed by means of prosody. The prosodic characteristics of rhetorical questions show such a coherent co-variation across speakers and contexts that researchers assume the existence of a 'prosodic construction' for this communicative function in the sense of a bundle of linked parameters and settings [14]. The parameters and settings we find here for the realization of rhetorical questions in hate speech are in accord with those reported in [14].

The conclusions we can draw from these findings with respect to our research questions is that there is obviously no uniform, context-independent phonetic core pattern that characterizes hate speech (we have also tested other prosodic parameters with the same result). The opposite is true. The contextual variation of hate speech is substantial. The nature of this variation suggests that it is externally driven, i.e. due to signalling communicative functions that can be involved in hate speech, such as rhetorical questions. Unlike for the latter, we found no evidence for something like a separate 'prosodic construction' of hate speech. Of course, this conclusion is still very preliminary since the analyzed production data only relies on a single speaker (although his skills argue in favor of the generalization of the produced items), and since we still need to compare the different hate-speech conditions to a baseline condition of no hate speech from the same speaker.
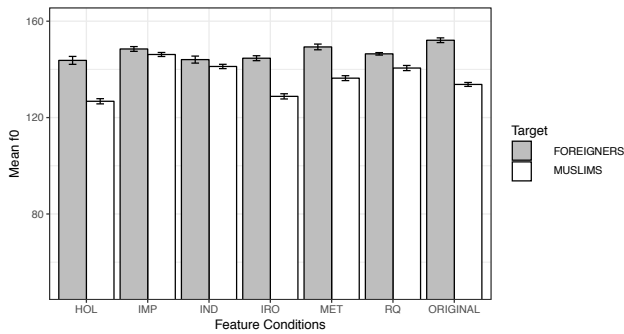
ACKNOWLEDGMENTS

Fig. 1. Mean f0 levels across all feature conditions for the two target groups (foreigners and Muslims).
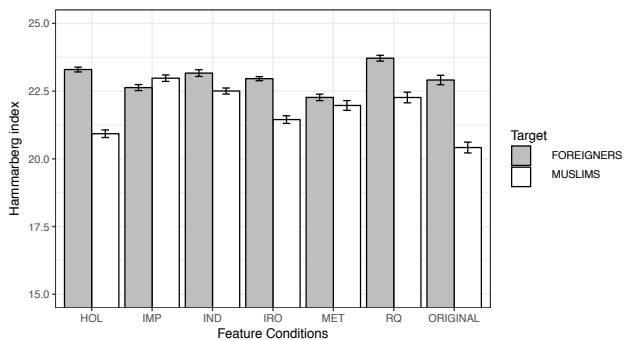

Fig. 2. Hammarberg index across all feature conditions for the two target groups (foreigners and Muslims).

REFERENCES

[1] Guterres, A. (2019). United Nations Strategy and Plan of Action on Hate Speech. Taken from: https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf

[2] Herz, M., & Molnár, P. (Eds.). (2012). *The content and context of hate speech: rethinking regulation and responses*. Cambridge University Press.

[3] Baumgarten, N., Bick, E., Geyer, K., Lund Iversen, D., Kleene, A., Lindø, A. V., Neitsch, J., Niebuhr, O., Nielsen, R. & Petersen, E. N. (Submitted). Towards Balance and Boundaries in Public Discourse: Expressing and Perceiving Online Hate Speech (XPEROHS). *RASK – International journal of language and communication*.

[4] XPEROHS Project. (2019). https://www.sdu.dk/en/om_sdu/institutter_centre/isk/forskning/forskningsprojekter/xperohs

[5] Assimakopoulos, S., Baider, F. H. & Millar, S. (2017). *Online Hate Speech in the European Union: A Discourse-Analytic Perspective*. Cham: Springer.

[6] McClay, R. (2017). Us and them: A descriptive analysis of Donald Trump's campaign speeches. *Unpublished Master Thesis. University of Birmingham. Retrieved by https://www. birmingham. ac. uk/Docum ents/college-artslaw/cels/essays/appliedlinguistics/McClay2017. Trump-Speech-Discourse-Analaysis. pdf*.

[7] Darmstadt, A., Prinz, M., Rocholl, F. & Saal, O. (2018). *Hate Speech und Fake News: Fragen und Antworten.* Amadeu Antonio Stiftung und Berliner Landeszentrale für politische Bildung, Berlin, Germany.

[8] Hrdina, M. (2016). Identity, activism and hatred: Hate speech against migrants on Facebook in the Czech Republic in 2015. Naše společnost, 14(1), 38-47.

[9] Niebuhr, O. & Peterson. J. M. (2011). *KALIPHO: Kieler Arbeiten zur Linguistik und Phonetik*. Kiel, Germany.

[10] Xu, Y. (2013). ProsodyPro – A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France. 7-10.

[11] De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, 41(2), 385-390.

[12] De Looze, C., & Hirst, D. (2008). Detecting changes in key and range for the automatic modelling and coding of intonation. In *Proceedings of International Conference on Speech Prosody*, 4, 6-9.

[13] Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, Series B 57, 289-300.

[14] Neitsch, J., & Niebuhr, O. (2019). Questions as prosodic configurations: How prosody and context shape the multiparametric acoustic nature of rhetorical questions in German. In *Proceedings of the 19th International Congress of Phonetic Sciences*. 2425-2429.