

Description for the general public

Hate speech is a form of verbal aggression targeting members of various social groups. It is widespread in everyday life – we hear in interactions with other people, in public spaces, on the TV, radio, and the Internet. Hate speech is a topic of great social significance as it has detrimental consequences not just for the targeted individuals and groups, but for everybody exposed to it. In the project, our objective is to understand the underlying properties of hate speech and its impact on the listeners. We will work with audio recordings of hate speech, and refer to it as *auditory hate speech* to discriminate it from textual representations, which most of the previous research concentrated on. We will create an audio database and develop acoustic, linguistic, and paralinguistic features relevant to fully automatic identification of hate speech. We expect that auditory hate speech will be a more powerful elicitor of emotional responses than its textual counterpart, and people will thus not get used to it as easily as to textual hate speech. We plan on conducting studies in which we will manipulate the content of stimuli, their modality, and participants' response options. We will then measure participants' psychophysiological and behavioral responses (e.g., activity of skin, heart, brain). We will also use signal processing and machine learning methods to identify and differentiate hate speech from other types of speech, on the sides of both speakers and listeners. This will be achieved through the creation of a vast amount of annotated data from hateful utterances, and subsequent development of automatic learning classifiers, which will draw on the acoustic, linguistic, and paralinguistic features of hate speech. Initially, annotations will be performed manually, and rapidly extended by means of automatic transcriptions based on automatic speech recognition. We believe that other researchers in the social sciences will be interested in using the outcomes of the project for their own work, e.g., to study blogs containing auditory hate speech. For the speech community in computer science, the manually annotated database will provide a highly useful ground truth for the study and classification of extreme emotions on the Internet.