# Optimizing creak detectors for Polish speech

Phoneticians might want to annotate creaky portions of audio recordings either in order to exclude them from analysis, or, conversely, to focus on them specifically, depending on the research goals of a study. Given the time-consuming nature of manually annotating creaky voice, automatic detection tools offer a practical alternative, or at least a helpful addition to a semi-automatic annotation pipeline. Two freely available Python-based creak detection tools are: `creapy` (Paierl et al., 2023) and `DeepFry` (Chernyak et al., 2022). `creapy` is based on a Random Forest classifier while `DeepFry` employs a deep learning model composed of an encoder and a classifier trained together. While their operation is, by definition, automatic, each of them leaves the analyst with some choice to make. For instance, `creapy` has a threshold parameter: the classifier first estimates the probability of a certain portion of signal containing creak, and if that probability exceeds the threshold specified by the analyst, then that portion is annotated as creaky. `DeepFry`, on the other hand, allows the selection of one of two pre-trained models, each trained on a different data set.

To find the optimal tool for detecting creak in Polish speech, that is to first choose between the two tools, and then to find the optimal setting of the winner, we compared the performance of these two tools against the gold standard, that is, human annotations. We used a corpus of recordings of twenty Polish women reading 119 Polish sentences each, in which creak was annotated by an expert phonetician for a separate study. We then used these annotations as gold standard against which the classifier performance was assessed. We annotated creak with `DeepFry`, with both its available models, as well as `creapy`, with the threshold parameter spanning the full range from 0 to 1, at 0.05 increments. Then, we verified whether the automatic annotation agreed with the gold standard annotation by comparing the value of the annotation (creak / no creak) in 10ms increments.

Using the resulting data, we then calculated the F1 metric and the Matthews Correlation Coefficient (MCC), both commonly used to assess classifier performance. The results are presented in Figure 1. Both metrics point to `creapy`, specifically with its threshold setting of 0.75, as the optimal tool (F1 = 0.685; MCC = 0.668) for detecting creaky voice in Polish female speakers.

# References


Chernyak, B. R., Simon, T. B., Segal, Y., Steffman, J., Chodroff, E., Cole, J. S., and Keshet, J. (2022). Deepfry: Identifying vocal fry using deep neural networks.

Paierl, M., Röck, T., Wepner, S., Kelterer, A., and Schuppler, B. (2023). Creapy: A Python-based tool for the detection of creak in conversational speech. In Skarnitzl, R. and Volín, J., editors, *20th International Congress on Phonetic Sciences : ICPhS 2023*, pages 1716–1720, Prague. GUARANT International. ISBN 978-80-908 114-2-3.

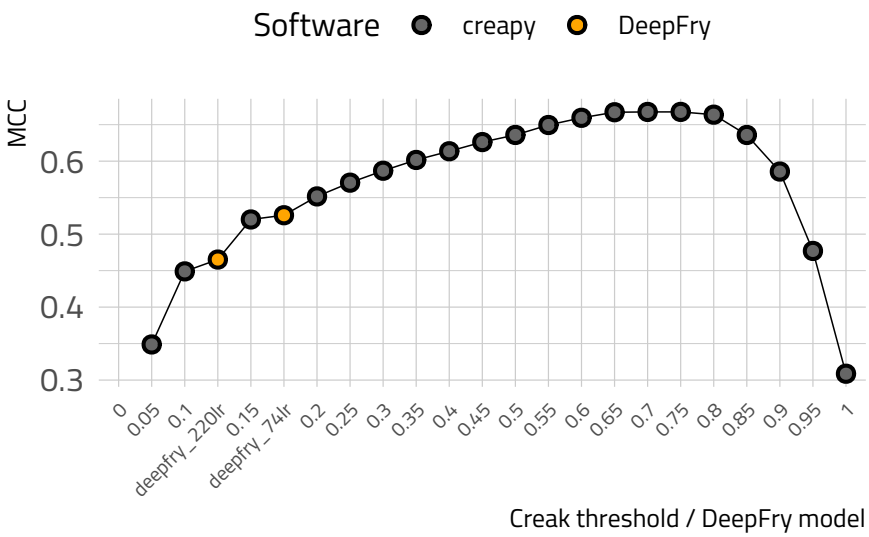


Figure 1: Performance of two creak detection tools: `creapy` and `DeepFry` on Polish as measured by the MCC metric.