Prosodic variability in suicidal and intoxicated speech: A case study

Lauri Tavi & Stefan Werner

During the last few decades in the field of speech science, numerous studies have aimed to identify acoustic-phonetic indicators of different emotions. Recognizing emotional states using speech analysis would improve, for instance, speaker profiling and speech recognition systems, or benefit different kinds of psychological evaluations, such as depression and suicide risk assessment [1, 2, 3]. Even though previous studies have reported significant associations between various acoustic features and emotions [4], due to high intra-speaker variation in speech prosody, these measures lack clear consistency. Nonetheless, already existing emotion classification models do not usually focus on intra-speaker prosodic variation or an influence of (alcohol) intoxication, which also affects speech prosody in addition to target emotion.

As in many other countries, completing a suicide is a major cause of death in Finland; for instance, in 2017 the total number of suicides was 824 (findikaattori.fi/fi/10). In this talk, we describe the prosodic variability in suicidal and alcohol intoxicated speech using two similar authentic emergency call recordings. The talk is based on our ongoing case study; we investigate prosodic variability (i.e. f0, energy values, rhythm metrics, and voice quality) and possibility of phonetic convergence between suicidal caller and emergency centre operator. The aim is to provide a better insight to prosodic features, which are applied for (semi-)machine detection of depression or suicide risk, in an emergency communication context.

[1] Cummings, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication,* 71, 10–49.

[2] Mundt, J. C., Snyder, P. J., Cannizzaro, M. S., Chappie, K., & Geralts, D. S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *Journal of neurolinguistics*, *20*(1), 50-64.

[3] Williamson, J. R., Young, D., Nierenberg, A. A., Niemi, J., Helfer, B. S., & Quatieri, T. F. (2019). Tracking depression severity from audio and video based on speech articulatory coordination. *Computer Speech & Language*, *55*, 40-56.

[4] Ververidis, D. & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech communication, 48*(9), 1162–1181.