

The data consists of dyadic face-to-face interactions (role-plays) between professional aggression regulation training actors and non-actors. These role-plays were designed by psychologists and psychiatrists and are similar to the ones practiced in therapy. To ensure realism, the only instruction the participants received were short role descriptions, including the context (involving a degree of urgency), their role (e.g., bus driver), and their goal (e.g., do not let the passenger travel without paying). Since no scripts were used, the interactions emerged as the participants reacted to one another, and could, therefore best be described as improvisations. Besides the degree of urgency in the specified scenarios, the actors played an important role in manipulating the emotional content of the database. Being proficient in giving aggression management training in forensic clinics and in showing and eliciting emotions, they were instructed to vary the degree of aggression for different scenarios.

Speaking setup: (file=SpeakSetup.csv): In total, 16 subjects participated in the recordings: 4 actors (3 male, 1 female) and 12 non-actors (3 male, 9 female). Three different scenarios were played multiple times according to the following protocol. Each actor played each scenario 3 times, every time with a different aggression level evolution and every time with a different non-actor. Each non-actor participated in each of the three scenarios once, every time with a different actor, and every time with a different aggression level variation. Overall, the scheme led to 36 recorded interactions of approximately 3 minutes each. The recordings were manually segmented into utterances based on turn-taking and by splitting longer utterances at pauses, which resulted in 2417 observations.

Annotation (file=labels.csv): Degree of aggression was annotated by 3 raters on a 3 point scale (*no aggression*, *medium level of aggression*, and *high level aggression*) and selected for the person speaking.

Features (file=features.csv): The features extracted at the utterance level to explore suprasegmental speech traits. According to findings on the effects of emotion on vocal parameters¹, we included features related to pitch, intensity, temporal aspects and voice quality, namely: speech duration (without silences), fundamental frequency F0 (mean, standard deviation, max, mean slope with and without octave jumps), intensity (mean, standard deviation, max, slope and range), first four formants (F1-F4) (mean frequency and bandwidth), jitter (F0 perturbation), shimmer (intensity perturbation). In addition we extracted features related to energy distribution in the spectrum: mean and slope of the spectrum, center of gravity and skewness of the long term averaged spectrum, the Hammarberg index which measures the difference between the energy in the low-frequency area and the energy in the high-frequency area, and the ratio between energy in the higher frequency bands to lower frequency bands (HF500) (HF1000). Lastly, mean and standard deviation of harmonics to noise ratio (HNR) were added to the feature set. Table 1 below shows the description of features in the data.

¹ Scherer, K.R., Johnstone, T. and Klasmeyer, G., 2003. Vocal expression of emotion. Handbook of affective sciences, pp.433-456.

Table 1. Description of features in the data

Feature name	Variable description
Sample_ID	Sample identifier
durSound	Speech duration (without silences)
meanPitch	Mean fundamental frequency F0
maxPitch	Maximum fundamental frequency F0
stdPitch	Standard deviation of fundamental frequency F0
rangePitch	Range of fundamental frequency F0
meanAbsSlopePitch	Mean absolute slope of fundamental frequency F0
meanSlopeOctJumps	Mean slope of fundamental frequency F0 with octave jumps
meanIntensity	Mean intensity
maxIntensity	Maximum intensity
stdIntensity	Standard deviation of intensity
rangeIntensity	Range of intensity
meanSlopeIntensity	Mean slope of intensity
meanF1	Mean of the first formant
meanF2	Mean of the second formant
meanF3	Mean of the third formant
meanF4	Mean of the forth formant
bandwidthF1	Bandwidth of the first formant
bandwidthF2	Bandwidth of the second formant
bandwidthF3	Bandwidth of the third formant
bandwidthF4	Bandwidth of the fourth formant
jitter	F0 perturbation
schimmer	Intensity perturbation
meanEnergy	Mean energy distribution in the spectrum
slopeLTASoc	Slope of energy distribution in the spectrum
HammarbegIndex	The Hammarberg index (the difference between the energy in the low-frequency area and the energy in the high-frequency area)
cogSpectrum	Centre of gravity of the long term averaged spectrum
skewSpectrum	Skewness of the long term averaged spectrum
HF500	Ratio between energy in the higher frequency bands
HF1000	Ratio between energy in the lower frequency bands
HNR	Mean harmonics to noise ratio (HNR)
stdHNR	Standard deviation of harmonics to noise ratio (HNR)