Identifying Emotion-Sensitive Acoustic Features in Italian Speech: A Comparative Analysis of Anger and Neutral Expression

Michele Pizzi[1], Gražina Korvel[1], Bożena Kostek[2], Povilas Treigys[1]

[1]Institute of Data Science and Digital Technologies, Vilnius University, Lithuania
[2]Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technologies, Poland

Background

Understanding how emotions shape acoustic speech characteristics is essential for developing interpretable and robust speech emotion analysis methods. Emotional speech synthesis and cross-linguistic emotion research require descriptors that reliably capture emotion-related acoustic variation.

The Emozionalmente corpus [1] provides matched emotional utterances, later integrated into the CAMEO collection [2] for cross-linguistic emotion research.

Objective

To identify the acoustic features most strongly associated with anger in Italian speech and evaluate their consistency across speakers, with the goal of establishing interpretable acoustic markers for emotion analysis and emotional speech synthesis.

This work represents the first step towards identifying interpretable emotion-sensitive acoustic features that can support multilingual and controllable emotional speech synthesis.

Method

Dataset

Subset of the Italian Emozionalmente emotional-speech corpus [1], as included in the multilingual CAMEO framework [2].

Sample

- 28 speakers (20 female, 8 male)
- 5 matched sentences per speaker
- 280 recordings (anger + neutral)

Acoustic Features

- 388 descriptors: eGeMAPS [3] + MFCC 1–20
- MFCC Δ (delta) and ΔΔ (delta-delta) coefficients.

Analysis

- Automated extraction of all acoustic descriptors;
- Neutral-ratio normalization applied **only for visualization** (Figure 1);
- Per-speaker computation of mean and standard deviation;
- Pearson correlations with binary emotion labels (**global and per-speaker**);
- Ranking and selection of the **top 50 features correlated with anger**;
- Exploratory discovery of emotion-sensitive features.

Listening Analysis

- Critical listening to interpret feature patterns and speaker variability.

Results

- **Anger-related acoustic features**: Anger was most strongly associated with changes in spectral dynamics and vocal effort. Spectral Flux showed the strongest global association with anger (r = 0.49), followed by several loudness-related measures. Pitch-related features showed weaker associations overall.

- **Spectral features were more consistent across speakers:** Features related to spectral dynamics and shape showed relatively stable associations with anger, whereas F0-related features varied more across speakers.

- **Multiple strategies for expressing anger:** Some speakers relief more on pitch and intensity changes, whereas others showed greater articulation-related and spectral changes.

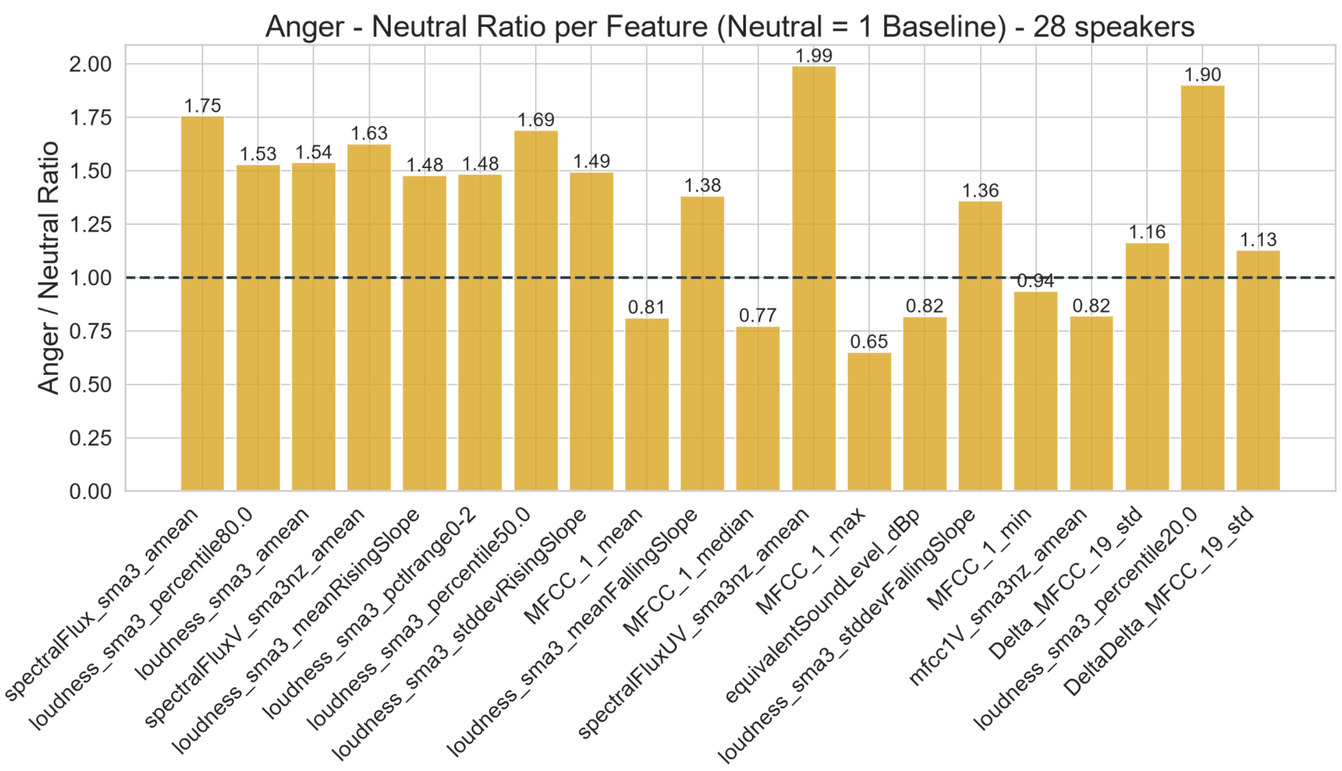

Figure 1. Top 20 anger-sensitive features ranked by global correlation with emotion labels. Ratios are shown relative to neutral speech (baseline = 1).

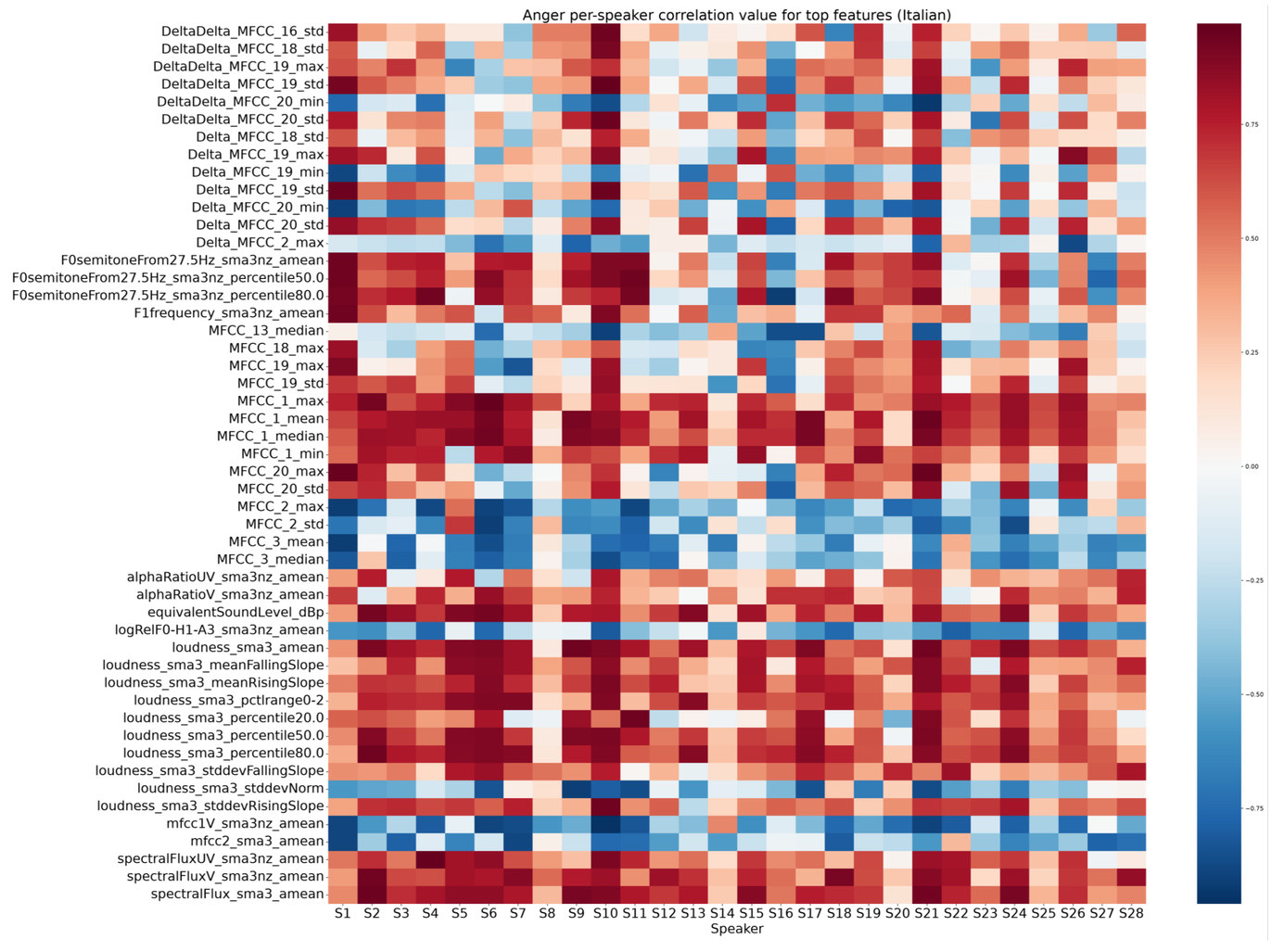

Figure 2. Per-speaker correlations for the 50 highest-ranked anger-related acoustic features. Features related to spectral dynamics and vocal effort show greater consistency across speakers than pitch-related descriptors.

Discussion

In the analysed corpus, anger was most consistently associated with increased spectral activity (Spectral Flux), vocal effort, and MFCC-derived descriptors, whereas pitch-related features showed greater inter-speaker variability. Spectral-dynamic and loudness-related features showed the most consistent associations with anger across speakers, suggesting they may provide relatively stable and interpretable cues for emotion analysis. The observed speaker-level variation also suggests that anger may be expressed acoustically through different speech production strategies.

Study Scope and Considerations

- The corpus includes recordings collected in diverse acoustic environments.
- The analysed subset included more female (20) than male (8) speakers.
- The present analysis focused on anger and neutral speech in Italian and provides an initial basis for future comparisons across datasets, emotions, and languages.

Conclusions

Within the analysed corpus, Spectral Flux, loudness-related descriptors, and MFCC-derived features showed the most consistent associations with anger, whereas pitch-related descriptors exhibited greater inter-speaker variability. These findings suggest that speakers may express anger acoustically in different ways. Future work will assess whether these patterns extend across additional emotions and languages, aiming to identify shared and language-specific characteristics of emotional speech expression.

References

[1] F. Catania, J. W. Wilke and F. Garzotto, "Emozionalmente: A Crowdsourced Corpus of Simulated Emotional Speech in Italian," in IEEE Transactions on Audio, Speech and Language Processing, vol. 33, pp. 1142-1155, 2025, doi: 10.1109/TASLPRO.2025.3540662.

[2] I. Christop and M. Czajka, "CAMEO: Collection of Multilingual Emotional Speech Corpora," ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2026, pp. 19457-19461, doi: 10.1109/ICASSP55912.2026.11461331.

[3] F. Eyben et al., "The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing," in IEEE Transactions on Affective Computing, vol. 7, no. 2, pp. 190-202, 1 April-June 2016, doi: 10.1109/TAFFC.2015.2457417.

Acknowledgments

Research is funded by the Research Council of Lithuania under the Programme “University Excellence Initiatives” of the Ministry of Education, Science and Sports of the Republic of Lithuania (Measure No. 12-001-01-01-01 “Improving the Research and Study Environment”). Project No.: S-A-UEI-23-11