*Applied Human Factors and Ergonomics (AHFE2025),* Vol. 199, 2025, 1067–1074
https://doi.org/10.54941/ahfe1006919

# Can Voices Predict Emergency Severity? An Exploratory Analysis of EMS Calls

**Kanji Okazaki[1] and Keiichi Watanuki[1,2]**

[1]Graduate School of Science and Engineering, Saitama University 255 Shimo-okubo, Sakura-ku, Saitama-shi, Saitama 338–8570 Japan
[2]Advanced Institute of Innovative Technology, Saitama University 255 Shimo-okubo, Sakura-ku, Saitama-shi, Saitama 338–8570 Japan

## ABSTRACT

This study explores whether voices recorded in emergency medical service (EMS) calls can be used to predict the severity of reported cases. Emergency call recordings provided by the Tokyo Fire Department were categorized into serious and minor groups, each consisting of approximately 100 anonymized segments. After resampling to 16 kHz and normalization, acoustic features were extracted, including fundamental frequency (F0), Mel-frequency cepstral coefficients (MFCCs), Mel-spectrogram energies, and spectral descriptors. Perturbation measures such as jitter and shimmer were attempted but proved unreliable under noisy conditions and were excluded from further analysis. Univariate statistical comparisons using Mann–Whitney U tests with false discovery rate (FDR) correction revealed significant differences between serious and minor calls, particularly in F0 variability, MFCCs, and spectral energy features. To complement these findings, exploratory multivariate models (logistic regression, support vector machines, and random forests) were trained with stratified cross-validation, showing promising discriminative capability as illustrated by ROC curves. While specific performance metrics were not emphasized due to the exploratory nature of the study, the results demonstrate that selected acoustic features carry informative patterns for severity assessment. These findings provide an empirical basis for developing robust, noise-tolerant decision-support systems to aid EMS dispatchers in rapid and reliable triage.

**Keywords:** Emergency medical services (EMS), Emergency call analysis, Acoustic features, Mfcc, Fundamental frequency (F0), Severity classification, Machine learning, Statistical analysis

## INTRODUCTION

Emergency medical services (EMS) rely heavily on information conveyed during emergency calls. Accurately assessing the severity of a case at the time of the call is essential for effective triage and rapid dispatch of appropriate resources. Traditional triage methods primarily depend on dispatcher judgment, which can be influenced by stress, caller behaviour, and background noise. As a result, automated methods for analysing emergency calls have attracted increasing research attention.

Recent studies have explored the use of acoustic features to support health monitoring, emotion recognition, and speech disorder assessment. In emergency call contexts, however, challenges remain due to the variability of speech, background noise, and the urgent, high-stakes environment in which data are recorded. Previous work has indicated that acoustic features such as fundamental frequency (F0), Mel-frequency cepstral coefficients (MFCCs), and spectral descriptors may contain valuable information for distinguishing between severe and non-severe cases. Perturbation measures such as jitter and shimmer have also been investigated but are often unreliable in real-world conditions.

Building on this background, the present study conducts an exploratory analysis of anonymized emergency call recordings provided by the Tokyo Fire Department. The objective is to identify acoustic features that differ significantly between serious and minor cases and to assess their potential discriminative power. In addition, we conduct exploratory multivariate modeling to examine whether combinations of features can provide robust classification performance. These findings are intended to establish an empirical foundation for future development of decision-support systems in EMS triage.

## MATERIAL AND METHODS

### Data Collection

Emergency call recordings were provided by the Tokyo Fire Department. Two classes of cases were defined: serious (requiring immediate advanced medical intervention, e.g., cardiac arrest or severe trauma) and minor (conditions judged to be non-life-threatening at dispatch). Approximately 100 anonymized segments were included in each group. All recordings were anonymized to remove personal identifiers, and ethical approval was obtained in accordance with institutional guidelines. Recordings were originally sampled at 48 kHz/16-bit PCM and were resampled to 16 kHz for analysis.

### Preprocessing

Audio files were processed using Librosa. If the original sampling rate differed, they were resampled to 16 kHz with Kaiser-windowed sinc interpolation. Stereo recordings were averaged to mono. Each signal was amplitude-normalized to the range [–1, 1] to prevent clipping. For feature extraction, a 25 ms window (400 samples), 10 ms hop (160 samples), and 512-point FFT were applied, with Hamming or Hann windows as appropriate. A 40-band Mel filterbank was used for MFCC and Mel-spectrogram analyses.

### Feature Extraction

Acoustic features were extracted using Librosa and Praat (via Parselmouth):

- Fundamental frequency (F0): Estimated using the probabilistic YIN (pyin) algorithm (fmin = 60 Hz, fmax = 400 Hz). Derived statistics included mean, standard deviation, interquartile range, minimum, maximum, range, voiced frame ratio (f0_valid_ratio), number of valid frames (f0_n), and variability of frame-to-frame differences (f0_delta_sd).

- Perturbation measures (jitter and shimmer): Calculated using Praat's PointProcess method. However, due to noise and overlapping speech, values were unstable and frequently missing; therefore, they were excluded from statistical analysis.
- Cepstral features (MFCCs): 13 Mel-frequency cepstral coefficients (20–8000 Hz, HTK Mel scale, 40 Mel filters). For each coefficient, summary statistics (mean, SD, IQR, min, max, range) were computed.
- Mel-spectrogram features: 40 log-scaled Mel-band energies averaged across time.
- Spectral features: Spectral centroid (mean, SD), spectral bandwidth (mean, SD), spectral roll-off (95%), zero-crossing rate (mean), and root mean square (RMS) energy (mean, 90th percentile).

Each file was thus represented by a structured feature vector comprising F0, MFCC, Mel, and spectral descriptors.

### Statistical Analysis

Numerical features were compared between serious and minor groups using the Mann–Whitney U test (two-sided, non-parametric). Effect sizes were calculated to evaluate practical significance: rank-biserial correlation, Cliff's delta, and Cohen's d. Multiple comparisons were adjusted with the Benjamini–Hochberg false discovery rate (FDR) procedure, with adjusted $p < 0.05$ considered significant. Features were visualized with boxplots and ranking plots (–log10 FDR vs. effect size).

### Additional Exploratory Modelling

To examine the multivariate discriminative potential, exploratory classification was performed. Non-numeric metadata were removed, labels were binarized (serious = 1, minor = 0), missing values were imputed with feature-wise means, and all features were standardized (zero mean, unit variance).

Three classifiers were tested: Logistic Regression (L2 regularization), Support Vector Machine (RBF kernel), and Random Forest (100 trees). Stratified 5-fold cross-validation was applied. Metrics included accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Feature selection was explored with Recursive Feature Elimination (RFE) and dimensionality reduction with Principal Component Analysis (PCA). ROC curves were plotted for each model to illustrate trade-offs between sensitivity and specificity.

## RESULTS

### Feature Extraction and Dataset

Acoustic features were successfully extracted from all emergency call segments (serious ≈100, minor ≈100). Each file yielded F0, MFCCs, Mel-spectrogram, and spectral descriptors. Jitter and shimmer were attempted but proved unstable under noisy, overlapping speech conditions and were excluded from subsequent analysis.

## Univariate Analysis

Mann–Whitney U tests with FDR correction identified several features that significantly differed between serious and minor calls.

- F0 variability (f0_delta_sd): higher in serious cases.
- MFCCs (2–13): consistent shifts in central tendency between groups.
- Spectral descriptors (centroid, RMS energy): significant contrasts across conditions.

Ranking plots (Fig. 1) highlighted MFCCs and F0 variability as top discriminators. Boxplots (Fig. 2) illustrated clear distributional differences between serious and minor calls.

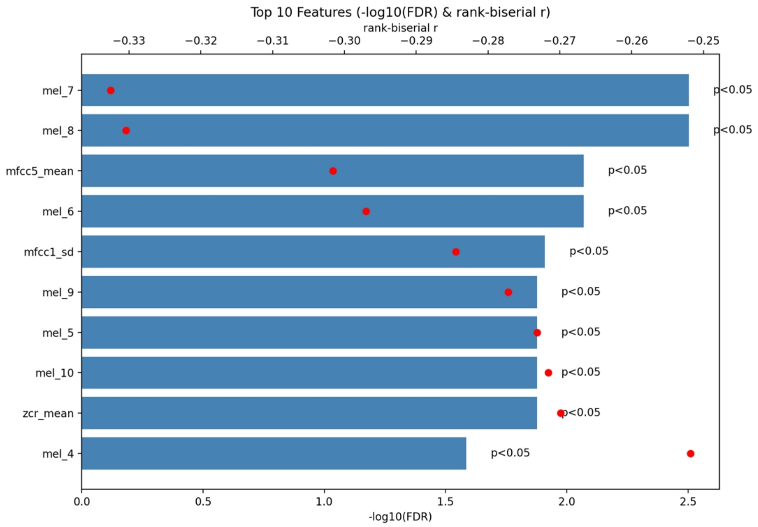


**Figure 1**: Ranking of acoustic features based on univariate analysis. Bars represent –log10(FDR-adjusted p-values), with red markers indicating rank-biserial effect sizes. Features with adjusted p <. 0.05 were considered significant.

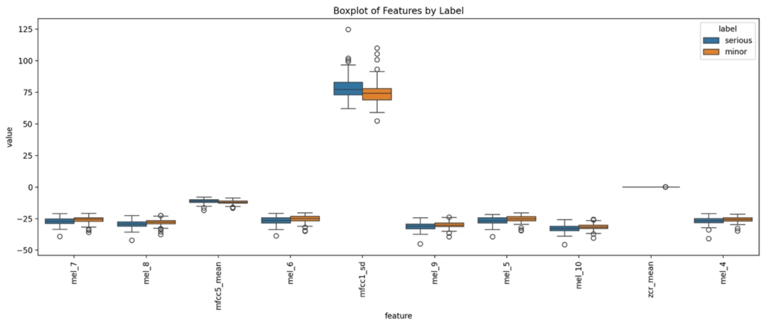


**Figure 2**: Boxplots of selected acoustic features (top 10). Distributions are shown separately for serious and minor cases, highlighting significant differences in MFCCs, F0 variability, and spectral energy descriptors.

### Exploratory Multivariate Modelling

Exploratory classification with logistic regression, support vector machines, and random forests was performed using stratified 5-fold cross-validation.

- All models achieved above-chance level discrimination.
- SVM and Random Forest demonstrated higher sensitivity than Logistic Regression.
- ROC curves (Fig. 3) indicated the potential of multivariate acoustic features to support severity detection.

### Summary of Findings

These results suggest that MFCCs, F0 variability, and spectral energy features are robust indicators of severity in emergency calls. Jitter and shimmer were unsuitable for this dataset. Exploratory multivariate models confirmed that combinations of acoustic features could provide meaningful discriminative performance in EMS triage.

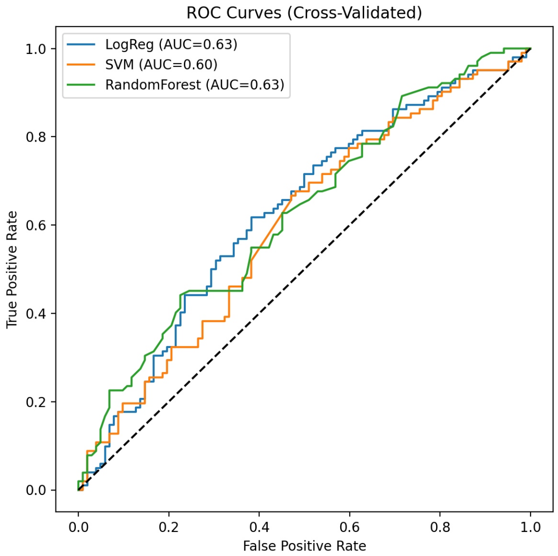


**Figure 3**: Receiver operating characteristic (ROC) curves for exploratory classification models (logistic regression, support vector machine, random forest). Curves illustrate the trade-off between sensitivity and specificity, demonstrating the potential of multivariate acoustic features for severity detection.

## DISCUSSION

This exploratory analysis demonstrates that acoustic features extracted from emergency call recordings can provide valuable information for severity assessment. In particular, MFCCs, F0 variability, and spectral energy measures showed consistent and statistically significant differences between serious and minor cases. These findings align with prior research suggesting that cepstral and spectral descriptors capture key aspects of vocal effort, stress, and urgency, which are relevant in emergency communication.

By contrast, perturbation measures such as jitter and shimmer proved unreliable under real-world recording conditions. Emergency calls often contain overlapping speech, background noise, and irregular phonation, which hinder stable estimation of fine-grained perturbation metrics. This limitation underscores the importance of selecting robust acoustic features when working with spontaneous and noisy speech data.

The exploratory multivariate modelling further confirmed that combinations of acoustic features could discriminate between serious and minor cases. Although specific numerical results were not emphasized, support vector machines and random forests demonstrated stronger sensitivity than logistic regression. Importantly, recall-oriented performance is particularly relevant in EMS contexts, where missing severe cases has critical consequences.

From a practical perspective, these results suggest that acoustic features could complement dispatcher assessments in real time. Integration into decision-support systems may enhance the consistency and speed of triage, thereby improving resource allocation in emergency medical services. However, this study remains exploratory. The dataset was relatively small, and further validation with larger and more diverse samples is required. Moreover, future work should investigate robustness through data augmentation, cross-dataset validation, and integration with linguistic features to enhance performance in operational settings.

## CONCLUSION

This study provided an exploratory analysis of acoustic features in emergency medical service (EMS) calls, demonstrating that measures such as MFCCs, F0 variability, and spectral energy can distinguish between serious and minor cases. Perturbation measures (jitter and shimmer) were found to be unsuitable under real-world conditions, emphasizing the importance of selecting robust and noise-tolerant features. Exploratory modelling further indicated that combinations of acoustic features have the potential to support automated triage systems.

While these findings establish a promising foundation, they remain preliminary. Future work must address robustness and generalizability through larger-scale datasets, data augmentation, and advanced modelling approaches. In particular, recall-oriented frameworks should be prioritized, as the cost of overlooking severe cases in EMS triage is critical. Building on the present exploratory results, our next step is to develop and evaluate robust machine learning models designed for real-world system integration. This transition will form the basis of our subsequent study, which emphasizes recall-focused evaluation and operational deployment in EMS decision support.

## ACKNOWLEDGMENT

The authors would like to acknowledge the Tokyo Fire Department for providing access to anonymized emergency call recordings used in this study. We also thank our colleagues and collaborators for their valuable technical advice and discussions.

## REFERENCES

Abi Kanaan, M. (2023) 'A methodology for emergency calls severity prediction', [Online]. Available at: https://publiweb.femto-st.fr/tntnet/entries/19812/documents/author/data (Accessed: 24 August 2025).

Ahmed, M., et al. (2024) 'A systematic literature review on acoustic features in cognitive decline evaluation', Journal of Speech, Language, and Hearing Disorders, [Online]. Available at: https://jslhd.org/xml/39916/39916.pdf (Accessed: 24 August 2025).

Deschamps-Berger, T., Lamel, L. and Devillers, L. (2021) 'End-to-End Speech Emotion Recognition: Challenges of Real-Life Emergency Call Centers Data Recordings', arXiv preprint, arXiv:2110.14957.

Fagherazzi, G., et al. (2021) 'The use of vocal biomarkers from research to clinical practice', NPJ Digital Medicine, 4, 12.

Figueroa, C., et al. (2024) 'Comparative study of acoustic parameters of voice and speech in adolescents with and without suicidal risk', Journal of Affective Disorders, 354, pp. 154–163.

Harada, S., Saito, Y. and Saruwatari, H. (2022) 'Analysis of urgency in evacuation voice messages and its application to speech synthesis', Proceedings of the Acoustical Society of Japan.

Ishihara, I. (2022) 'Detection of pathological voices using the cepstrum method', Journal of the Acoustical Society of Japan, 78(9), pp. 496–499.

Kadiri, S. R., Kodali, M. and Alku, P. (2023) 'Severity classification of Parkinson's disease from speech using single frequency filtering-based features', arXiv preprint, arXiv:2308.09042.

Kamble, M. R., Patiño, J., et al. (2022) 'Exploring auditory acoustic features for the diagnosis of COVID-19', arXiv preprint, arXiv:2201.09110.

Kim, S., et al. (2023) 'COVID-19 detection model with acoustic features from cough sounds', Applied Sciences, 13(4), 2378.

Kobayashi, M. (2022) 'Acoustic features correlated to perceived urgency in speech', Journal of the Acoustical Society, [Abstract].

Li, G., Hou, Q., Zhang, C., et al. (2021) 'Acoustic parameters for the evaluation of voice quality in patients with voice disorders', Annals of Palliative Medicine, 10(1), pp. 104–112.

Li, X. (2007) 'Stress and emotion classification using jitter and shimmer', Doctoral dissertation, Marquette University. Available at: https://epublications.marquette.edu/data_drdolittle/9/.

Min, S., et al. (2023) 'Acoustic analysis of speech for screening for suicide risk', Journal of Medical Internet Research, 25, e45456.

Mori, H. (2023) 'Acoustic feature analysis of speech for estimating depression levels', Proceedings of the Annual Conference of the Japanese Society for Artificial Intelligence (JSAI2023), 1O3GS705.

Ota, K. (2023) 'Design of acoustic features for detecting abnormal industrial sounds using amplitude-enhanced shimmer', JAIST Repository. Available at: https://dspace.jaist.ac.jp/dspace/bitstream/10119/18821/2/paper.pdf.

Rashid, M., Alman, K. A., Hasan, K., Hansen, J. H. L. and Hasan, T. (2020) 'Respiratory distress detection from telephone speech using acoustic and prosodic features', arXiv preprint, arXiv:2011.09270.

Shams, M. Y., et al. (2024) 'Acoustic data detection in large-scale emergency vehicle dispatch', Expert Systems with Applications.

Shiotani, K., Kimura, K., Kitakoya, K. and Komatani, T. (2023) 'Analysis of speech behavior in 119 emergency calls: Characterizing information transmission structures', Bulletin of Kansai University Society of Human Health, 14, pp. 45–56.

Srivastava, T., Chou, J.-C., Shroff, P., Livescu, K. and Graziul, C. (2024) 'Speech Recognition for Analysis of Police Radio Communication', arXiv preprint, arXiv:2409.10858.

Toyé, A. A. and Kompalli, S. (2021) 'Comparative study of speech analysis methods to predict Parkinson's disease', arXiv preprint, arXiv:2111.10207.

Tracey, B., Volfson, D., Glass, J., et al. (2023) 'Towards interpretable speech biomarkers: exploring MFCCs', Scientific Reports, 13, 22787.

Valizada, A. (2021) 'Development of speech recognition systems in emergency call centers', Symmetry, 13(4), 634.

Wertzner, H. F., Schreiber, S., Amaro, L., et al. (2015) 'Analysis of fundamental frequency, jitter, shimmer and intensity in children with phonological disorders', Brazilian Journal of Otorhinolaryngology, 81(6), pp. 581–588.

Yanuma, M. (2021) 'Communication characteristics in reports of cardiac arrest by elderly callers: Examination of delay factors in telephone-assisted CPR', Bulletin of Kokushikan University Faculty of Letters, 52(2), pp. 33–42.

Young, V. (2016) 'Exploratory analysis of real personal emergency response call conversations', PMC, [Online]. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5109662 (Accessed: 24 August 2025).