In: Proceedings of the 2026 IEEE International Conference on Human-Machine Systems (ICHMS), 1-3 Jul 2026, Singapore, Singapore, pp. 77-82
DOI: 10.1109/ichms69701.2026.11602261

Recognising emotions in air-ground communications with deep learning

Cho Yin Yiu and Wen-Chin Li

***Abstract*—Decision-making in air traffic operations often requires a stable emotional state to ensure the decision quality. To ensure aeronautical decisions are free from emotional impacts, this research presents a novel dataset and deep learning model to identify the emotions of stakeholders, including pilots and air traffic controllers (ATCOs). We recorded 30 utterances in seven different emotions from 20 participants who are pilots or ATCOs. Features were extracted from the utterance recordings for emotion recognition. A long short-term memory (LSTM) model was constructed to perform emotion recognition. The proposed model yielded a test accuracy at 64.29%. The model demonstrates the potential in identifying emotions of pilots and ATCOs via their communications.**

I. Introduction

A. Research Background

Pilots are often required to maintain their situational awareness to make rapid and appropriate decisions in a short time, which directly or indirectly affects flight safety [1]. In particular, during landing, pilots need to process multidimensional information from multiple sources including weather, runway conditions, traffic density, radio communication, and unexpected scenarios, etc. Particularly, they need to communicate with air traffic controllers (ATCOs) effectively during the flight [2]. With the information gained from communication, they can collaborate with automation make the appropriate and accurate manoeuvre decisions [3], which is crucial to flight safety. However, these safety-critical decisions might also be affected by the operator's emotion [4]. Hence, it is essential to ensure pilots to maintain a stable emotion for effective decision-making to prevent accidents.

B. Emotion and human performance

Emotion plays a critical role in human cognition, behaviour and capability of problem-solving, particularly in high-stakes and high-pressure environments such as aviation [5, 6]. These can be revealed through subtle signs of unusual intonation, speech rate change, or cognitive overload in the crew's speech before critical errors occur [7]. As a result, maintaining emotions for higher cognitive flexibility is essential in proper decision-making. It ensures that decisions are not biased nor influenced by emotions of the operators. Indeed, Causse, et al. [8] revealed that emotions, particularly negative emotions, can significantly impair a person's ability to make rational judgements and decisions. Past research indicates that emotional instability contributes to aviation incidents or accidents since emotionally unstable pilots may delay their response or make inappropriate decisions in times of crisis [9]. Unstable or negative emotions, when coupled with a high pressure, may trigger a series of negative reactions, which in turn affect the quality of pilots' decision-making [10]. Hence, in aviation, emotional management is considered an essential pillar for flight safety.

To achieve effective emotional control, it is essential for pilots to regulate their emotions, maintain a calm mental state, and avoid excessive emotional fluctuations that could affect their professional judgement and operational abilities. During cadet pilot training, pilots are trained to maintain the emotion of "neutral". In general, emotionally stable pilots are able to respond more quickly to emergencies, leading to incidents or accidents. Therefore, it is critical to develop advanced methods to detect the nuances or signs of emotional instability promptly for preventing further incidents or accidents, which improve aviation safety.

C. Deep learning and affective computing

Affective computing aims to perform objective identification of one's emotion. Among many approaches, speech emotion recognition is one of the evolving fields to achieve affective computing. However, speech is a type of sequential data and temporally dynamic. Hence, advanced deep learning architectures, such as recurrent neural networks (RNN), are essential to model temporal dependencies in speech emotion recognition [11].

Since then, deep learning is widely used to capture temporal dynamics speech signals. For instance, Ye, et al. [12] proposed a temporal-aware bi-direction multi-scale network (TIM-Net) to learn the contextual affective representations from temporal perspective. This model is specifically applied generic speech emotion recognition and gained superior performance on several corpus.

Meanwhile, among different types of RNN, Long Short-term Memory (LSTM) is a type of RNN that selectively retain or discard information over long sequences, which makes LSTM a dominant architecture for processing cross-temporal speech features such as intonation transitions, rhythmic pauses, and speech rate changes [13]. Yang, et al. [14] adopted speech rate to restructured speech signals and trained a LSTM model for speech emotion recognition. The model achieved a high accuracy at up to 98.47%, which demonstrated the effectiveness of LSTM in speech emotion recognition. Hence, LSTM is considered as a proper model to perform speech emotion recognition.

Cho Yin Yiu is with Human Factors and Ergonomics Laboratory, Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China and Safety and Accident Investigation Centre, Cranfield University, Cranfield, Bedfordshire, United Kingdom (e-mail: james-cho-yin.yiu@connect.polyu.hk).

Wen-Chin Li is with Safety and Accident Investigation Centre, Cranfield University, Cranfield, Bedfordshire, United Kingdom (e-mail: wenchin.li@cranfield.ac.uk).

Published by IEEE. This is the Author Accepted Manuscript issued with: Creative Commons Attribution License (CC:BY 4.0).
The final published version (version of record) is available online at DOI:10.1109/ichms69701.2026.11602261. Please refer to any applicable publisher terms of use.

D. Research gap and objectives

While deep learning is proven effective in speech emotion recognition, the uniqueness of aviation phraseology made it challenging to directly adapt existing speech emotion recognition model in aviation usage. In particular, aviation language has its own set of sequence, phraseology, and word choices. A dedicated model for capturing aviation communications shall be developed to enhance the accuracy of emotion recognition in this specific context to mitigate the effects of emotion on aeronautical decision-making.

Hence, given the promise of deep learning in speech emotion recognition, this research aims to develop a speech-based emotion recognition model with long short-term memory (LSTM) to identify the emotions from the air-ground communication utterances. By identifying signs of different emotions through deep learning, proactive management of emotional states can be designed to help prevent impaired judgment and maintain the cognitive performance required for safety.

The paper is organised as follows: Section II presents the dataset collected to develop the deep learning model. Section III describes the proposed LSTM model for speech emotion recognition. Section IV reports the results of the emotion recognition. Finally, Section V concludes the work, discusses its limitation, and the potential for future enhancement.

II. DATASET

Given the domain-specificity in aviation phraseology, the accents and wordings adopted are known to be different from the other domains. Hence, using open-source emotional datasets like Toronto Emotional Speech Set (TESS) [15] and Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [16], which are recorded in controlled environments with general phrases, is not a promising choice for applications in aviation. Therefore, rather than using existing emotion recognition datasets, we developed our own dataset for speech emotion recognition in aeronautical context.

A. Design of utterances

With the unique linguistic and contextual properties of air-ground communications, we designed 30 utterances for each role in air-ground communications based on standard aviation phraseology and realistic operation contexts to simulate realistic communications between pilots and ATCOs. These utterances included a variety of air-ground communications such as routine calls, conflict resolution, weather diversions, clearance requests, status or problems report, emergency calls, etc.

To ensure the authenticity of the utterances, they were reviewed by three subject matter experts, including active flying pilots, tower controllers, and area controllers.

B. Definition of different emotions

To collect labelled data of emotion, it is essential to define the target emotions to be classified in order to collect the corresponding data. In this study, seven common emotions are selected for evaluation, including anger, disgust, fear, happiness, sadness, surprised, and neutral. TABLE I shows the definition of the target emotions.

TABLE I. DEFINITION OF TARGET EMOTIONS

Emotion	Definition
Anger	An emotion characterised by the feelings of displeasure, hostility, and antagonism, triggered by perceived threats, injustices, or frustrations including annoyance, rage, indignation, vengeance, etc [17].
Disgust	An emotion characterised by feelings of revulsion and aversion towards something perceived as offensive, which motivates individuals to avoid harmful situations [18].
Fear	An emotion characterised by unpleasant reaction to actual or perceived threats or danger, which can be triggered by concrete external stimuli, such as loud noise, or internal psychological threats, such as anticipations of pain [19].
Happiness	An emotion characterised by a state of positive wellbeing, contentment, and pleasure, and can be triggered by the experience of achieving goals, receiving rewards, or being in pleasant or fulfilling situations [20]
Sadness	An emotion characterised by feelings of loss, disappointment, helplessness, and withdrawal. It often arises in response to negative experiences such as failure or unmet needs [17].
Surprised	An emotion characterised by a brief and intense response to the unexpected event, which can be triggered by a sudden, unforeseen event that mismatches the prediction [17].
Neutral	An emotion characterised by emotional stability and the absence of strong positive or negative affective responses [21].

C. Participants and data collection

We collected voice recordings from 20 participants, including pilots (n = 16), tower controllers (n = 2), and area controllers (n = 2). The utterances were recorded in the *Safety and Accident Investigation Centre, Cranfield University*. All participants were fully informed about the purpose of the study through verbal explanation on the intended use of the recordings in this research. Participants were explicitly informed of their right to withdraw from the study at any time without any consequence. Ethical approval had been granted by the Cranfield University Research Ethics Systems (CURES). Pre-experiment informed consent was obtained in written form for all participants.

It is realised that the regional accents of English may affect the data quality. Hence, we collected data from multiple regions. Native languages covered included Mandarin (n = 8), Indonesian (n = 10), and Arabic (n = 2), covering a variety of regional English accents to ensure the robustness of the model. Each participant was asked to read aloud the 30 utterances in seven different emotions, including anger, disgust, fear, happiness, sadness, surprised, and neutral. Therefore, we have 4,200 recordings in total.

D. Feature extraction

It is noteworthy that a direct deployment of recordings in deep learning model training usually yields extremely low accuracy that could be equivalent to random chance. Instead, properly extracting relevant and high-quality features from the raw data boosts the accuracy of the model. Among many auditory features that can be extracted from voice recordings, Mel Frequency Cepstral Coefficients (MFCCs) are the most widely used features for emotion recognition due to their ability to represent the short-term power spectrum of speech

signals in a manner that closely aligns with human auditory perception [22, 23]. In particular, its ability to describe the spectral characteristics of speech has made it a powerful tool for emotion recognition. Hence, in this study, we extracted the Mel-Frequency Cepstral Coefficients (MFCC), a set of features that are highly relevant to emotions, from the raw recordings.

Past research indicated that 13 coefficients per speech frame are sufficient to capture the most perceptually relevant features for speech and audio analysis [24, 25]. The 13 coefficients include 12 cepstral coefficients with an energy coefficient. The energy term indicates the overall log-energy of the frame, which captures the overall loudness. This is important for tasks like speech recognition and emotion classification since loudness often correlates with certain emotions. The 12 cepstral coefficients describe the spectral envelope, indicating the speech spectrum and reflects the vocal tract configuration, resonance, and formant structure. It is ordered by their importance in capturing the spectral structure.

MFCCs are often found related to different emotions. For instance, anger is often characterised by increased speech rate, higher volume, and elevated pitch, resulting in higher energy level in the energy coefficient, shifted or broadened formants, and greater spectral envelope variations, leading to greater fluctuations in MFCC coefficients [26]. On the other hand, sadness tends to involve slower speech, lower pitch, energy concentrated in the lower frequency range [27], which reflects decreased sound energy, narrower spectral bandwidth in MFCC patterns [28]. These acoustic variations significantly alter the spectral shape of speech, which can be effectively captured by MFCC features.

The extraction process involves pre-emphasis, framing, windowing, Short-Time Fourier Transform (SFFT), mapping the powers of the spectrum to the mel scale using a filter bank, taking the logarithm of the powers, and finally applying the discrete cosine transform (DCT) to decorrelate the coefficients and reduce dimensionality. To obtain MFCCs, each audio recording is first divided into short, overlapping frames to capture the local temporal and spectral properties of the speech. The wave form is segmented into frames of 25 milliseconds, with a frame shift of 10 milliseconds. In this study, we employed Librosa 0.11.0 with Python 3.10.12 to extract the MFCCs. The number of MFCCs extracted is 13, with a sampling rate of 22,050 Hz.

III. Problem Description

The speech emotion recognition model is formulated as a classification problem with seven classes. Each class corresponds to a target emotion to be classified from the voice-recorded utterances. In this section, we describe the proposed speech emotion recognition model developed for air-ground communications.

A. Long short-term memory layer

In this speech emotion recognition problem, we propose the adoption of LSTM, which is a type of recurrent neural network (RNN), to solve this classification problem. LSTM is designed to effectively learn and model long-term dependencies in sequential data. Unlike traditional RNNs, LSTM addresses the vanishing gradient problem with a gating mechanism that controls the flow of information through the network. This allows LSTM to retain important information over extended sequences while ignoring less relevant data.

As LSTM is particularly superior in capturing temporal relationships, it aligns with our original data captured as time-series audio recordings, which features are sequential and context-dependent. With LSTM, the short-term changes (e.g., specific phonemes or syllables) and long-term dependencies (e.g., the emotional tone of an entire sentence) can be captured, making them highly effective for tasks involving temporal patterns of emotion in speech.

B. Hyperparameters tuning

Deep learning models can be enhanced through hyperparameters and network architecture tuning [29]. Hence, a proper tuning of these parameters is required to achieve the optimal results. We deployed KerasTuner to seek the optimal classification performance through a grid search. In the grid search, we seek the optimal hyperparameters, including learning rate and batch size, as well as the network structure, through an iteration of different parameter configurations and yield the best configuration.

The search is conducted with the objective function of maximising validation accuracy. For hyperparameters, we evaluated a batch size of 32, 64, and 128, with a learning rate of 0.001 and 0.0001. The LSTM layers are evaluated with (512), (256), (512, 256), (512, 256, 128). Two fixed dense layers of (128, 64) are added before the final dense layer of 7 units for classification.

The grid search is conducted in the Windows 11 Enterprise 64-bit operating environment with AMD Ryzen 9 9950X3D 4.30 GHz CPU, 96 GB RAM, and NVIDIA GeForce RTX 5090 32 GB with Python 3.8.18. Our results indicates that the optimal network structure is reached at one LSTM layer only with 256 units, with the optimal learning rate of 0.001 and the optimal batch size of 32.

C. Model configuration

The final configuration of the model consists of one LSTM module with 256 units, followed by two Dense modules of 128 and 64 units, respectively. Each module accompanies with a BatchNormalisation layer and a dropout layer with rate 0.3 to prevent overfitting. For each LSTM and Dense layer, we also applied an L2 regularisation for generalising the model. We adopted a stratified sampling method by participant and emotion class. Hence, the sample for each emotion class is equal. For simplicity, we split the 10% data for testing. The remaining 90% dataset is separated for training and testing on a ratio of 8:1.

All models are coded with TensorFlow Keras 2.10.0 in Python 3.8.18. The maximum epoch for training is set as 200. If there are no improvement in validation accuracy for ten epochs, the training process will terminate earlier. A learning rate of 0.001 and a batch size of 32 is adopted on an Adam optimiser. The model is trained using the same platform as the grid search.

IV. Results

The model training terminated at Our model achieved an average test accuracy of 64.29% and an average class-weighted F1-score at 61.95%. Table II lists the precision, recall, and the F1-score of each class.

TABLE II. Precision, Recall, and F1-Score by Class

Emotion	Precision	Recall	F1-score
Anger	0.6379	0.6167	0.6271
Disgust	0.6604	0.5833	0.6195
Fear	0.6271	0.6167	0.6218
Happiness	0.6949	0.6833	0.6891
Sadness	0.5556	0.5833	0.5691
Surprised	0.5625	0.6000	0.5806
Neutral	0.6094	0.6500	0.6290

It is noteworthy that a seven-class classification is particularly challenging: a random chance could only achieve an accuracy of 0.1429. The accuracy achieved by our proposed model have significantly improved from a random prediction. Apart from the overall accuracy, we also evaluated the precision, recall, and F1-score by each class. From Table II, the highest F1-score is attained for happiness at 0.6891, while the lowest F1-score is attained for sadness at 0.5691. The average F1-score across class in 0.6195, indicating a satisfactory level of accuracy for a seven-class classification problem.

We then further inspect the confusion matrix for classes with a F1-score less than 60%, i.e., sadness and surprised. From our inspection, we realise that the model mainly confuses sadness with anger (n = 5), disgust (n = 5), and surprised (n = 5). Meanwhile, surprised is mainly confused fear (n = 8) and neutral (n = 5). Indeed, certain emotions can be seen as groups of emotions given their linkage. Sometimes, anger can be the precursor of sadness and disgust, which might affect the model's prediction quality. On the other hand, fear is often caused by unexpected negative consequences, which can be seen as a result of surprise. The transition and similarity between two emotions could be one of the key challenges for the model to distinguish them clearly. Nevertheless, the results demonstrated the ability and potential of the model in recognising emotions in air traffic operations.

V. Conclusion

Overall, this study adopted a self-collected dataset with all audio data recorded by air traffic operators (pilots and ATCOs with different roles) for speech emotion recognition with utterances in air-ground communications. 13 MFCCs are extracted as input features. The LSTM model attained an overall test accuracy at 64.29%, which shows the potential of recognising emotions with aviation speech. In practice, this model can be applied to real-time detection of emotions and be used to design intelligent automations based on the emotional state of pilots.

However, this study also subjected to several limitations. First, the sample size of the current dataset is limited to 4,200 recordings from 20 participants. Although the sample size is significantly larger than other datasets like TESS (2,800 recordings from two participants), the dataset can be extended in the future. Particularly, majority of our participants are pilots while only four of them (20%) are ATCOs. The dataset can thus be balanced between different types of participants for a higher representativeness. While we performed a grid search on the current dataset, it is noteworthy that there are infinite numbers of combination for hyperparameters and network architecture. Hence, it is impossible to cover all combinations. Future work may consider evaluating other deep learning architectures, or even a hybrid model with convolutional neural networks, which may facilitate the prediction of emotional states of pilots and ATCOs in aviation communications for enhanced aviation safety. The use of cross-dataset tests or transfer learning experiments can also ensure the results to be more robust in the future for other emotion recognition tasks.

Furthermore, data augmentation can also be considered to expand the dataset to enhance the model training quality. In particular, the diversity of participants can be enhanced through increasing the number of ATCOs and with a wider distribution in terms of age and regional accents. Lastly, the explainability of deep learning-driven speech emotion recognition model is a major challenge. In high-stakes industry like aviation, one of the key hurdles of applying artificial intelligence (AI) tools is the trustworthiness of AI. Hence, future studies can leverage explainable AI, such as the Shapley Additive Explanations (SHAP) or other methods to enhance the explainability, transparency, and trustworthiness of AI.

Acknowledgment

The research is supported by Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong SAR (RLPA, YWFR) and Safety and Accident Investigation Centre, Cranfield University, United Kingdom. Cho Yin Yiu is a recipient of the Hong Kong PhD Fellowship (Reference number: PF21-62058). This study has been granted human ethics approval from the Cranfield University Research Ethics Systems.

References

[1] C. Y. Yiu, Q. Li, K. K. H. Ng, and X. Yuan, "Enhancing Pilot-Autopilot Collaboration in Cruising Operations Through Task Allocation: A Psychophysiological Approach," *International Journal of Human–Computer Interaction,* pp. 1-25, 2025, doi: 10.1080/10447318.2025.2608264.

[2] C. Y. Yiu, K. K. H. Ng, Q. Li, and X. Yuan, "Gaze behaviours, situation awareness and cognitive workload of air traffic controllers in radar screen monitoring tasks with varying task complexity," *International Journal of Occupational Safety and Ergonomics,* vol. 31, no. 2, pp. 504-515, 2025, doi: 10.1080/10803548.2025.2453312.

[3] C. Y. Yiu, K. K. H. Ng, Q. Li, and X. Yuan, "How do the levels of automation in flight operations affect pilots' cognitive workload, reaction time, and

EEG brain waves in cruising flights?," in *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*, 15-17 May 2024, pp. 1-6, doi: 10.1109/ICHMS59971.2024.10555620

[4] W. T.-K. Chan, W.-C. Li, and G. Braithwaite, "Pilots' training backgrounds affecting the attribution of event causal factors and airline safety management," *Journal of Air Transport Management,* vol. 125, p. 102786, 2025, doi: 10.1016/j.jairtraman.2025.102786.

[5] H. Okon-Singer, T. Hendler, L. Pessoa, and A. J. Shackman, "The neurobiology of emotion–cognition interactions: fundamental questions and strategies for future research," *Frontiers in Human Neuroscience,* Review vol. Volume 9 - 2015, 2015, doi: 10.3389/fnhum.2015.00058.

[6] C. M. Tyng, H. U. Amin, M. N. M. Saad, and A. S. Malik, "The Influences of Emotion on Learning and Memory," *Frontiers in Psychology,* Review vol. Volume 8 - 2017, 2017, doi: 10.3389/fpsyg.2017.01454.

[7] T. Lu, Y. Li, C. Zhou, M. Tang, and X. You, "The Influence of Emotion Induced by Accidents and Incidents on Pilots' Situation Awareness," *Behavioral Sciences,* vol. 13, no. 3, p. 231, 2023, doi: 10.3390/bs13030231.

[8] M. Causse, F. Dehais, P. Péran, U. Sabatini, and J. Pastor, "The effects of emotion on pilot decision-making: A neuroergonomic approach to aviation safety," *Transportation Research Part C: Emerging Technologies,* vol. 33, pp. 272-281, 2013, doi: 10.1016/j.trc.2012.04.005.

[9] J. S. Lerner, Y. Li, P. Valdesolo, and K. S. Kassam, "Emotion and Decision Making," *Annual Review of Psychology,* vol. 66, no. Volume 66, 2015, pp. 799-823, 2015, doi: 10.1146/annurev-psych-010213-115043.

[10] C. Y. Yiu, K. K. H. Ng, Q. Li, and X. Yuan, "Keeping Pilots in the Loop: An Explainable Spatiotemporal EEG-Driven Deep Learning Framework for Adaptive Automation in Cruising Flight Phase," *IEEE Transactions on Intelligent Transportation Systems,* vol. 26, no. 7, pp. 9838-9851, 2025, doi: 10.1109/TITS.2025.3567987.

[11] M. Khan, W. Gueaieb, A. El Saddik, and S. Kwon, "MSER: Multimodal speech emotion recognition using cross-attention with deep fusion," *Expert Systems with Applications,* vol. 245, p. 122946, 2024, doi: 10.1016/j.eswa.2023.122946.

[12] J. Ye, X. C. Wen, Y. Wei, Y. Xu, K. Liu, and H. Shan, "Temporal Modeling Matters: A Novel Temporal Emotional Modeling Approach for Speech Emotion Recognition," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4-10 June 2023 2023, pp. 1-5, doi: 10.1109/ICASSP49357.2023.10096370

[13] C. Beling and G. Wild, "The association between emotional intelligence and decision making for pilots," *Journal of Air Transport Management,* vol. 114, p. 102506, 2024, doi: 10.1016/j.jairtraman.2023.102506.

[14] Z. Yang, Z. Li, S. Zhou, L. Zhang, and S. Serikawa, "Speech emotion recognition based on multi-feature speed rate and LSTM," *Neurocomputing,* vol. 601, p. 128177, 2024, doi: 10.1016/j.neucom.2024.128177.

[15] M. K. Pichora-Fuller and K. Dupuis. *Toronto emotional speech set (TESS)*.

[16] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," *PLOS ONE,* vol. 13, no. 5, p. e0196391, 2018, doi: 10.1371/journal.pone.0196391.

[17] P. Ekman, "Facial expression and emotion," vol. 48, pp. 384-392, 1993, doi: 10.1037/0003-066X.48.4.384.

[18] P. Rozin, J. Haidt, and C. R. McCauley, "Disgust: The Body and Soul Emotion," in *Handbook of Cognition and Emotion*, 1999, pp. 429-445.

[19] A. Öhman, "Fear and anxiety: Overlaps and dissociations," in *Handbook of emotions, 3rd ed.* New York, NY, US: The Guilford Press, 2008, pp. 709-728.

[20] E. Diener, S. Oishi, and R. E. Lucas, "Personality, Culture, and Subjective Well-Being: Emotional and Cognitive Evaluations of Life," *Annual Review of Psychology,* vol. 54, no. Volume 54, 2003, pp. 403-425, 2003, doi: 10.1146/annurev.psych.54.101601.145056.

[21] F. Luciani, G. Veneziani, C. Ciacchella, G. Rocchi, M. Reho, A. Gennaro, and C. Lai, "Safety at high altitude: the importance of emotional dysregulation on pilots' risk attitudes during flight," *Frontiers in Psychology,* Original Research vol. Volume 13 - 2022, 2022, doi: 10.3389/fpsyg.2022.1042283.

[22] M. Gupta and S. Chandra, "Speech Emotion Recognition Using MFCC and Wide Residual Network," presented at the Proceedings of the 2021 Thirteenth International Conference on Contemporary Computing, Noida, India, 2021.

[23] S. Jena, S. Basak, H. Agrawal, B. Saini, S. Gite, K. Kotecha, and S. Alfarhood, "Developing a negative speech emotion recognition model for safety systems using deep learning," *Journal of Big Data,* vol. 12, no. 1, p. 54, 2025, doi: 10.1186/s40537-025-01090-0.

[24] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 28, no. 4, pp. 357-366, 1980, doi: 10.1109/TASSP.1980.1163420.

[25] J. de Lope and M. Graña, "An ongoing review of speech emotion recognition," *Neurocomputing,* vol. 528, pp. 1-11, 2023, doi: 10.1016/j.neucom.2023.01.002.

[26] D. Ververidis and C. Kotropoulos, "Emotional speech recognition: Resources, features, and methods," *Speech Communication,* vol. 48, no. 9, pp. 1162-1181, 2006, doi: 10.1016/j.specom.2006.04.003.

[27] P. Larrouy-Maestri, D. Poeppel, and M. D. Pell, "The Sound of Emotional Prosody: Nearly 3 Decades of Research and Future Directions," *Perspectives on Psychological Science,* vol. 20, no. 4, pp. 623-638, 2025, doi: 10.1177/17456916231217722.

[28] C. M. Lee and S. S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing,* vol. 13, no. 2, pp. 293-303, 2005, doi: 10.1109/TSA.2004.838534.

[29] C. Y. Yiu, K. K. H. Ng, Q. Li, and X. Yuan, "Identifying Human Out-of-the-Loop in Cruising Flights Using EEG Spectral Features with Deep Learning," in *HCI International 2025 Posters*, Cham, C. Stephanidis, M. Antona, S. Ntoa, and G. Salvendy, Eds., 2025: Springer Nature Switzerland, pp. 123-133.

https://dspace.org/

Recognising emotions in air-ground communications with deep learning

Yiu, Cho Yin; Li, Wen-Chin

2026-07-01
Institute of Electrical and Electronics Engineers (IEEE)
https://doi.org/10.1109/ichms69701.2026.11602261
Yiu CY, Li W-C. (2026) Recognising emotions in air-ground communications with deep learning. In: Proceedings of the 2026 IEEE International Conference on Human-Machine Systems (ICHMS), 1-3 Jul 2026, Singapore, Singapore, pp. 77-82

Downloaded from a DSpace repository. https://dspace.org/