A03 Voice Tech in the Multiverse Creative Challenge

Jiri Musil, Martin Demel

Department of Science, Technology, Engineering & Math,
Houston Community College

Natural Language Processing (NLP) - ITAI 2373

Patricia McManus

June 16th, 2025

Introduction

We chose the Space Station scenario because we are space enthusiasts, and our long-standing membership with the Johnson Space Center, just 45 minutes away, constantly motivates and inspires us about what is possible. Creating a reliable voice communication system for an international space station involves a distinctive set of challenges and opportunities. In a microgravity setting, metallic modules create prolonged reverberation, while the constant hum of life-support fans and the hiss from pressurized airlocks add to both consistent and variable noise sources. Spacesuit helmet microphones limit bandwidth and clarity, and the variety of accents plus stress-related vocal changes among an international crew add further complexity. This report explores acoustic and environmental factors in detail, proposing a tailored preprocessing pipeline, feature-extraction method, and ASR/TTS adjustments that together ensure clear and reliable communication, even during emergencies.

The proposed system guarantees clear speech across different station modules by combining adaptive noise cancellation, spectral subtraction methods, and helmet-transfer compensation filters. Accent-resistant language models, along with a hybrid DNN-based keyword-spotting fallback, ensure vital phrase recognition even as the signal-to-noise ratio drops significantly. Ultimately, our creative demo scenario demonstrates both normal operations and a crucial hull-breach event, showing how the system smoothly switches modes to protect lives. The executive pitch further develops these technical innovations into a compelling value proposition for stakeholders, emphasizing the competitive edge of our solution and branding.

Part 1: World Analysis

The space station's metallic modules and life-support systems generate a complex acoustic environment. In microgravity, sound bounces off solid surfaces, causing extended reverberation, while the ongoing vibrations from pumps and fans produce a low-frequency rumble. Pressurized airlock seals occasionally emit hissing and clanking noises from the valves. These factors, including decreased internal pressure that lowers high frequencies and electromagnetic interference from power converters, require specialized noise reduction and equalization methods.

Crew members communicate through helmet-mounted microphones that limit bandwidth to roughly 300 Hz–3 kHz and muffle high frequencies. An international team highlights notable phonemic variability, with stress or fatigue potentially further affecting vocal consistency. Noise assessments in key zones like life-support modules (55–65 dB), exercise areas (50–60 dB), and airlock corridors (60–70 dB) guide the development of adaptive filters. The on-board AI assistant, called "HERA," employs a synthetic vocal-tract model pre-adjusted for helmet acoustics, demonstrating how non-human voices can also benefit from particular compensatory modifications.

Part 2: Technical Solutions Design

Building on the world analysis, the proposed processing architecture begins with raw audio from helmet microphones. It passes through a multi-stage pipeline that reduces noise, corrects spectral distortions, and extracts dependable speech features.

A band-pass filter covering 300 Hz to 3.4 kHz eliminates out-of-band hum and high-frequency hiss. Adaptive noise cancellation (ANC) then focuses on reducing nonstationary fan and pump noise by constantly updating its reference filters. Spectral subtraction is used to reduce any remaining stationary noise floor. A pre-emphasis stage amplifies high-frequency speech components to counteract the low-pass characteristics of the helmet shell.

Once windowing and framing are completed, Mel-frequency cepstral coefficients (MFCCs), along with their first and second derivatives (Δ, ΔΔ), are calculated to represent both static and dynamic spectral characteristics. To address helmet-induced spectral tilt, an inverse-spectrum-response (ISR) filter, precomputed from measured transfer functions, is used to equalize resonant peaks prior to feature extraction. Finally, data normalization and augmentation (including simulated noise injection) prepare the feature vectors for the acoustic model, which supports both automatic speech recognition (ASR) and text-to-speech (TTS) synthesis.

Standard MFCC methods work well within the filtered band but need minor adjustments to the filterbank center frequencies to account for changes in speech spectra. Spectral subtraction effectively removes the station's steady fan hum but needs to be combined with ANC to manage transient valve hisses. The acoustics model is made resilient to low signal-to-noise ratios by training it on artificially created noise profiles that replicate actual station recordings. Accent adaptation is achieved through a multi-accent lexicon and specific language-model fine-tuning, guaranteeing dependable performance with international crews.

For ASR, confidence-weighted decoding includes an out-of-vocabulary detection mechanism, which is essential during emergency situations when new terms (e.g., panel IDs) need to be recognized. On the TTS front, a mix of waveform and parametric methods strikes a balance between sounding natural and being understandable. It dynamically tweaks prosody to show urgency, while avoiding distortion on radio and intercom links.

2.1 Custom preprocessing pipeline flowchart

1. Raw Helmet Microphone Input: Audio recorded inside the helmet.

2. Band-Pass Filter (300–3400 Hz): Eliminates out-of-band noise, such as low hum and high hiss.

3. Adaptive Noise Cancellation (ANC): Continuously reduces non-stationary fan and pump noises.

4. Spectral Subtraction: Additional reduction of stationary noise floor.

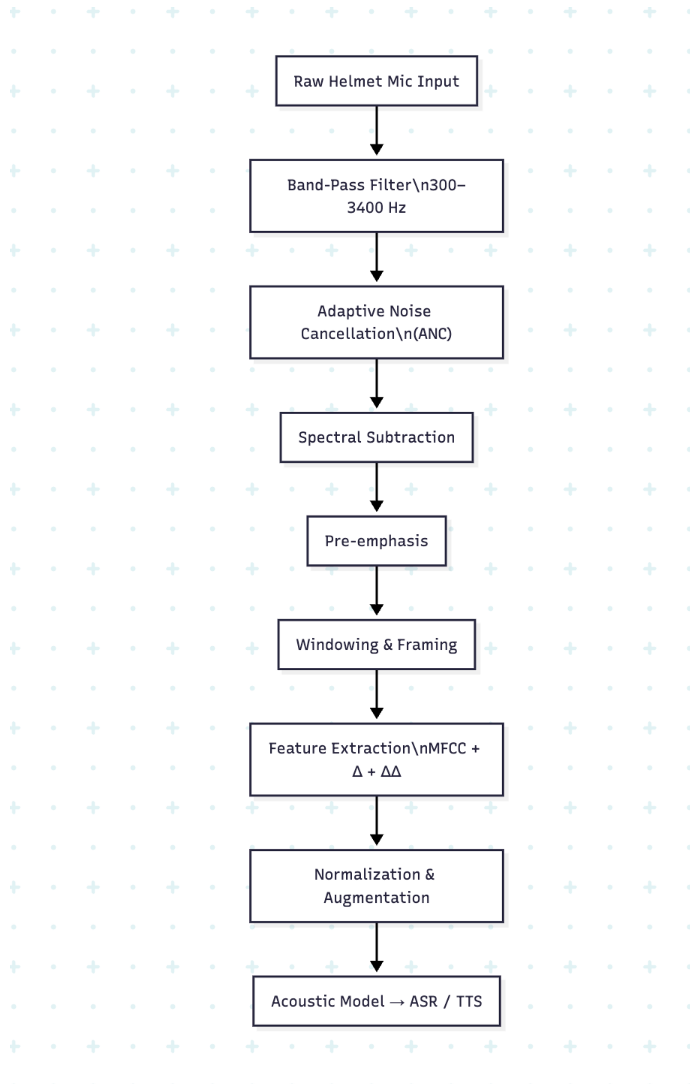
5. Pre-emphasis: Enhances high-frequency components lost to the helmet shell.

6. Windowing & Framing: Divides the signal into segments for short-time analysis.

7. Feature Extraction (MFCC + Δ + ΔΔ): Calculates static and dynamic cepstral features, with helmet-transfer compensation applied beforehand.

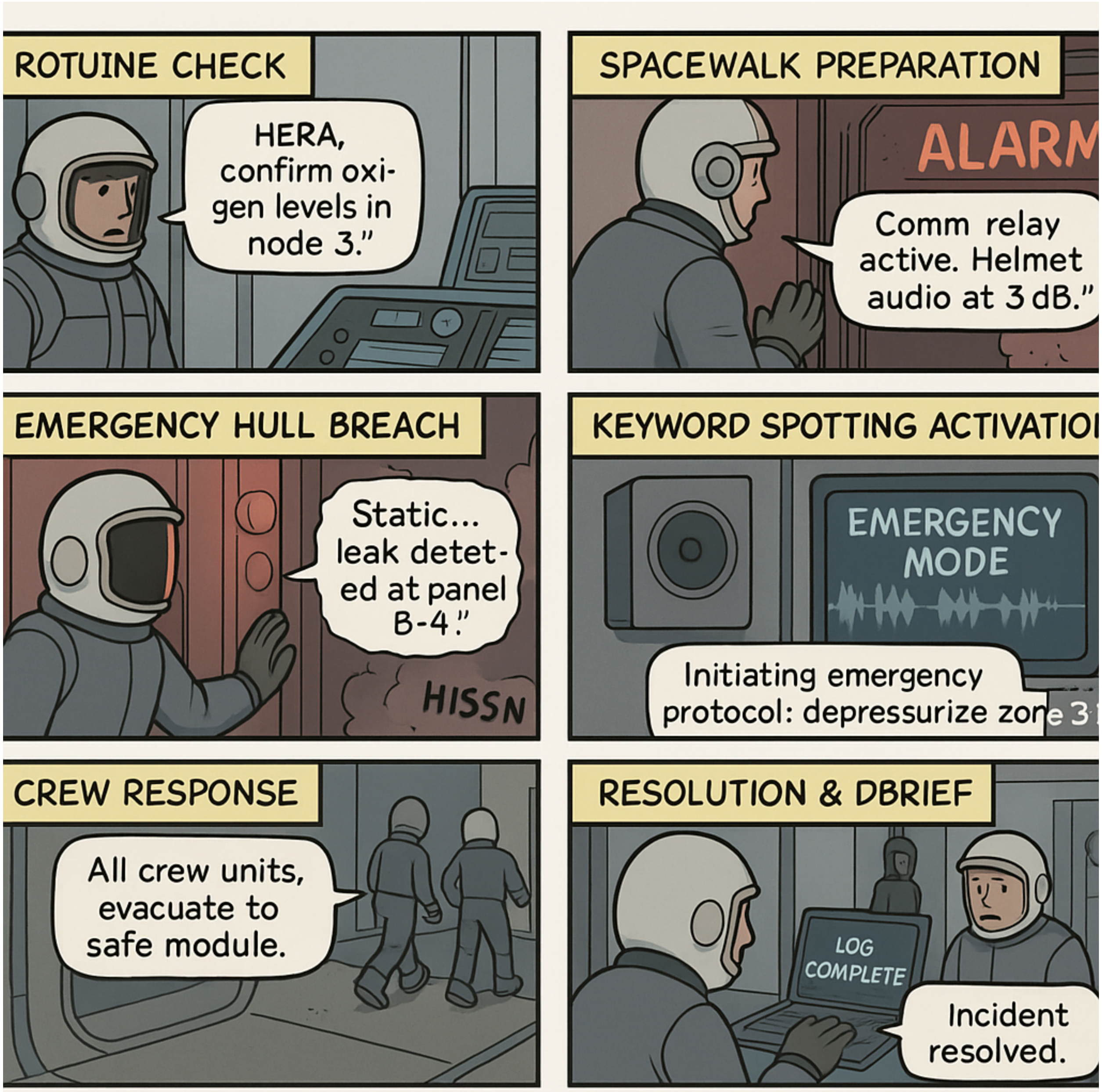
8. Normalization & Augmentation: Scales features and adds synthetic noise to enhance model robustness.

9. Acoustic Model → ASR / TTS: The final recognition or synthesis engine, which includes emergency-mode keyword spotting and urgency-aware TTS.



Part 3: Demo Scenario

A six-panel storyboard demonstrates the system's capabilities in both normal and emergency scenarios.



1. Routine Check: The commander says, "HERA, confirm oxygen levels in node 3." Adaptive noise cancellation lowers background pump noise, providing a clear voice signal to the AI assistant.
2. Spacewalk Preparation: An engineer in the airlock reports, "Comm relay active. Helmet audio at +3 dB." The band-pass filter automatically activates, removing low-frequency valve hiss.
3. Emergency Hull Breach (Failure Mode): A sudden alarm and hissing drown out the channel. Speech becomes garbled: "Static… leak detected at panel B-4!" As the SNR decreases, intelligibility drops below the ASR threshold.
4. Keyword Spotting Activation: The system switches to a DNN-based keyword-spotting model, capturing commands like, "Initiating emergency protocol: depressurize zone 3." This ensures necessary commands are detected even in difficult acoustic environments.
5. Crew Response: The TTS engine announces, "All crew units, evacuate to safe module," with increased prosody to convey urgency while maintaining clarity.

6. Resolution & Debrief: After the crew is safely regrouped, the system records raw and processed audio segments for post-incident review, verifying that both ASR and TTS modules worked correctly.

Part 4: Executive Pitch

Please see the attached file.

Resources

1. *The Hera Space Companion: new frontiers in science communication – News Center*. (2024, November 20). https://news.microsoft.com/de-at/the-hera-space-companion-new-frontiers-in-science-communication/
2. *International Space Station - NASA*. (June,2025). NASA. https://www.nasa.gov/reference/international-space-station/
3. NASA. (2015). *Environmental control & life support system (ECLSS) technical brief*. NASA. https://www.nasa.gov/wp-content/uploads/2023/07/eclss-technical-brief-ochmo.pdf
4. IBM. (June,2025). *What is speech recognition?*. IBM. https://www.ibm.com/think/topics/speech-recognition
5. *End-to-End speech AI pipelines*. (June,2025). NVIDIA. https://resources.nvidia.com/en-us-speech-ai-ebooks-gated/speech-ai-using-asr-and-tts?nvid=nv-int-tblg-974861#cid=dl20_nv-int-tblg_en-us