Perceptually Adaptive Speech Clarity System for Reverberant Environments

Author: Dr Iain McGregor
Affiliation: Edinburgh Napier University
ORCID: 0000-0002-0230-9007

Abstract

This document outlines a perceptually adaptive audio system designed to improve the intelligibility of spoken announcements in highly reverberant environments. Using speech buffering, pacing modification, spectral shaping, and real-time feedback, the system enhances clarity without architectural intervention. All processing is local and privacy-respecting. This record constitutes a timestamped public disclosure to support further research and development.

Background and Motivation

Reverberant public spaces—such as train stations, cathedrals, and large atria—often render announcements difficult to understand, particularly for individuals with hearing challenges or those in acoustically disadvantaged positions. Current solutions often require hardware upgrades or architectural changes, which are expensive, disruptive, or impractical.

This system proposes a non-invasive, low-cost method to enhance speech clarity by adapting announcements based on perceptual principles.

System Overview

The system operates between the sound source (microphone or media player) and the existing public address (PA) system. It enhances intelligibility by dynamically modifying temporal and spectral features of speech and continuously evaluating performance through real-time environmental monitoring.

Signal Flow

Speech Input: Live or pre-recorded announcements are received from standard microphones or audio devices.

Buffering and Pacing Adjustment: Short-term buffering allows the insertion of micro-pauses between words or syllables to improve intelligibility in reverberant spaces.

Spectral Shaping and Language-Aware Filtering: Frequencies most susceptible to masking are emphasised. The system also applies adaptive shaping depending on language features and detected speech characteristics.

Public Address Output: The optimised signal is transmitted through the existing loudspeaker system. No retrofitting is required.

Monitoring and Adaptation

Remote Microphone Monitoring: A secondary microphone (e.g. Bluetooth, wired, or smartphone-based) captures the in-room audio as experienced by listeners.

Speech Transmission Index (STI) Evaluation: The system computes the STI from the monitored signal and compares it to a clarity threshold (e.g. STI ≥ 0.6).

Calibration and Recalibration: If clarity is suboptimal, the system recalibrates processing parameters. Initial calibration can be crowd-sourced via smartphone recordings taken at various listener positions.

Error Handling: In the event of microphone failure or invalid STI measurements, the system reverts to the last known good calibration and may optionally alert the operator for manual review.

System Features

Edge Processing: All computation is performed locally. No data is uploaded or stored in the cloud unless a software update is requested.

Emergency Bypass: The system allows instant passthrough of speech in critical situations without processing delay.

Operator Control Panel: A tablet or desktop interface allows manual overrides, profile switching, and system diagnostics.

Low-Cost Implementation: Designed to run on widely available devices (e.g. tablets, embedded systems), minimising deployment cost.

Context-Aware Profiles: Settings may adjust automatically by time of day, room use, or ambient spectral conditions.

Clarity Compliance Support: The system can log metrics such as announcement repeat frequency or average STI to support accessibility and regulatory compliance.

Applications

- Transport terminals (rail, airport, bus)
- Heritage and cultural sites
- Museums and exhibition spaces
- Large classrooms and lecture theatres
- Emergency communication systems in complex indoor environments

Ethical and Accessibility Considerations

Perceptual Sovereignty: Enhancements are global, not personalised—maintaining fairness in shared spaces.

Privacy by Design: No recording or storage of user voices; all monitoring is anonymised and local.

Universal Benefit: Improves clarity for all listeners, including those with mild hearing loss, cognitive processing differences, or language barriers.

Future Extensions for Communication Systems

Although originally designed for announcement clarity in shared public spaces, the underlying principles can also enhance one-to-one and group communication in acoustically complex environments.

Voice-Adaptive Communication Systems: The system can dynamically adjust based on the characteristics of the speaker's voice (e.g. clarity, frequency content, rhythm) and the background environment. This could benefit speakers with quieter voices, non-native accents, or speech production differences.

Real-Time Speech Support: Using low latency buffering and pacing, the system could assist in real-time conversations in noisy environments (e.g. care homes, call centres, industrial settings), helping ensure critical spoken information is not lost.

Masking Detection and Compensation: By comparing voice input to the ambient spectrum, the system could reduce the risk of verbal masking—particularly in group settings where overlapping speech is common. This could enhance communication inclusivity in classrooms or multi-party calls.

Private, On-Device Use: The same engine could be embedded in communication aids or hearing-accessibility apps that adapt live conversations without recording, transcription, or cloud connectivity—preserving privacy while enhancing audibility.

Disclosure and Licensing

This document represents a timestamped public disclosure of the concept and design principles. No patents have been filed. This work is shared under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence and may be reused or extended with attribution.