ELEC5305 Project Proposal

Perceptually Enhanced Speech Recognition Using Cochlear Filterbank and ICA-based Speech Separation

1. Introduction

Automatic Speech Recognition (ASR) has become one of the most active areas in modern audio signal processing. Despite significant advances, conventional systems often suffer from reduced accuracy under noisy or reverberant environments. Most existing front-end feature extraction pipelines (e.g., MFCC or log-Mel features) are based on simplified spectral models that do not fully reflect the human auditory system.

This project aims to design a perceptually enhanced speech recognition system that integrates two biologically inspired and signal processing–based modules derived from ELEC5305 laboratory content:
1. A Cochlear Filterbank Front-End to simulate the non-linear frequency selectivity of the human ear;
2. An Independent Component Analysis (ICA)-based separation stage to improve noise robustness before feature extraction.

The final system will demonstrate improved recognition accuracy and perceptual quality compared to a standard MFCC-based baseline.

2. Objectives

1. Develop a complete speech recognition pipeline derived from ELEC5305 lab topics, including STFT, feature extraction, and signal separation.
2. Implement and analyze a Cochlear-inspired front-end for perceptual feature extraction.
3. Apply ICA-based source separation as a speech enhancement module in noisy mixtures.
4. Compare baseline MFCC and proposed cochlear–ICA features in recognition accuracy and perceptual robustness.
5. Provide sound demonstrations and spectrogram visualizations to illustrate improvements.

3. Methodology

The proposed system integrates multiple modules inspired by ELEC5305 lab experiments. Each module is associated with both its origin and innovation aspect.

Module	Derived from Lab	Function	Innovation
Signal Analysis	BasicSignalAnalysis.mlx	Frame blocking, pre-emphasis,	Adaptive frame selection for

		normalization	speech regions
Time–Frequency Representation	STFTExample.mlx	Short-Time Fourier Transform (STFT)	Dynamic window adjustment for energy stability
Feature Extraction	AudioFeatures.mlx	MFCC, spectral centroid, zero-crossing rate	Fusion of cochlear features and MFCC
Perceptual Modeling	CochlearFilterBank2025.mlx / CochlearMaskExample2025.mlx	Simulate auditory frequency selectivity and masking	Replace Mel filterbank with cochlear filter responses
Speech Separation	ICAExample.mlx	Independent Component Analysis for noise and speaker separation	Front-end speech enhancement
Evaluation and Visualization	StereoDecomposition & SpatialMixing	Energy distribution and localization visualization	Visualize perceptual improvement in spectrograms

Pipeline Overview:
Input Speech → ICA Separation → Cochlear Filterbank → Feature Extraction (MFCC + Spectral Features)
↓
Feature Fusion and Normalization
↓
Classifier (GMM / DNN / Template Matching)
↓
Recognition Results and Evaluation Metrics

4. Evaluation Plan

Metric	Description
Recognition Accuracy (%)	Measured on test dataset comparing

	baseline vs. proposed features
ΔSNR (dB)	Improvement from ICA-based separation front-end
Perceptual Mean Opinion Score (MOS)	Subjective evaluation of clarity (1–5 scale)
Confusion Matrix Visualization	Graphical comparison of misclassification patterns
Spectrogram Comparison	Visual analysis of noise reduction and feature clarity

5. Expected Outcomes and Contributions

- A working prototype of a perceptually motivated speech recognizer.
- Objective evidence showing improved robustness in noisy or reverberant conditions.
- Demonstrations illustrating perceptual benefits of cochlear modeling and ICA separation.
- A comprehensive 10-page final report and 5-minute presentation video explaining the algorithm and showcasing sound examples.

6. Project Timeline

Week	Task	Deliverable
Week 8–9	Literature review on cochlear filterbanks and ICA for ASR	Brief summary & initial sketches
Week 10	Implement baseline MFCC-based recognizer	Baseline accuracy report
Week 11	Implement ICA speech separation and cochlear filterbank modules	Enhanced feature dataset
Week 12	Train and evaluate combined system (ICA + Cochlear + GMM)	Evaluation tables & confusion matrix
Week 13	Final analysis, discussion, and improvements	Draft final report
Week 14	Prepare final report and presentation video	Complete deliverables submission

7. References (Indicative)

- Lyon, R. F. (2017). Human and Machine Hearing: Extracting Meaning from Sound. Cambridge University Press.
- Hyvärinen, A., & Oja, E. (2000). Independent Component Analysis: Algorithms and Applications. Neural Networks.
- Davis, S., & Mermelstein, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE TASSP.
- Smith, J. O. (2011). Spectral Audio Signal Processing. W3K Publishing.