

# Robust Speech Processing for Immersive Interactions

## Mission Rehearsal Exercise Project

### Executive Summary

Recent advances in information and communication technologies have not been paralleled by similar advances that ensure ubiquitous and natural interaction with information by everyone, regardless of their skills, abilities and preferences. Spoken language is the most natural and powerful medium for human communication. Conversational interfaces hold the promise of providing natural and efficient interaction with machine agents. Our ultimate goal is to enable immersive conversations, where people can freely interact with one another and with the virtual agents without regard to the physical constraints of where they are. Through creation and implementation of robust speech recognition technology, we propose to advance the usability and usefulness of military training simulation environments.

The Mission Rehearsal Exercise project creates a framework for creating and implementing state-of-the-art in speech technology especially aimed at the challenging requirements of military training environments. The proposal targets three specific tasks: (1) *Speech algorithm development* : acoustic models for robust ASR performance, language models designed for military training scenarios, automatic emotion recognition capability and an integrated approach to ASR in conjunction with dialog and emotion models (2) *Testing, Evaluation and Refinement*: The models and algorithms will be tested with data obtained under realistic training scenario conditions and in turn utilized in continual refinement of various models. (3) *MRE Integration and porting*: Algorithms developed will be ported and evaluated in the MRE training prototype.

Algorithmically, a unified stochastic framework will be adopted to provide a tight integration across the speech modules. This framework will combine models for the speech acoustics, language patterns, emotional information from speech and user behavior in a mathematically unified way. Signal processing and speech/emotion recognition will explicitly integrate information from language understanding, dialog state and the user model information including the emotion state. The speech processing R&D will leverage the research capabilities of the USC Speech Analysis and Interpretation Laboratory that the PI directs.