# THE SHEFFIELD SEARCH AND RESCUE CORPUS

The University Of Sheffield

Saeid Mokaram, Roger K. Moore
{s.mokaram, r.k.moore}@sheffield.ac.uk

SP AND H

## Introduction

- As part of an ongoing research into extracting mission-critical information from Search and Rescue speech communications, a corpus of unscripted, goal-oriented, two-party spoken conversations has been designed and collected.
- The Sheffield Search and Rescue (SSAR) is a multi-speaker speech corpus with 96 two-party, goal-oriented, spoken conversations lasting between 6 to 8 minutes each (a total of ~12 hours of data), spoken by 24 native British English speakers (66.6% Male).
- Each conversation is about a collaborative task of exploring and estimating a simulated indoor environment. The task has carefully been designed to have a quantitative measure for the amount of exchanged information about the discourse subject.
- SSAR includes different layers of annotations which should be of interest to researchers in a wide range of human/human conversation understanding as well as automatic speech recognition.
- This corpus is being made available for research purposes (via LDC ).

## Conversation scenario

- An abstract speech communication model between First Responders (FR) and Trask leaders (TL) in a Search and Rescue context was used to design the underlying task for the SSAR conversations.

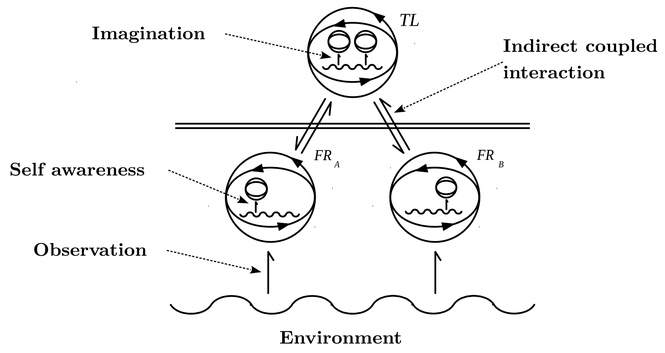



Fig. 1. Pictograph illustration of the abstract communication model within the SAR context.

- In this model, FRs' goal is to explore an environment and report their observations back to the control hub to update the TL's knowledge about the incident scene.
- The SSAR task involves two participants in the roles of an FR and a TL.

## Simulated environment maps design

- Inspired by the simulation training systems which are being used by some fire departments to practice their communication performance and decision making, a simulated indoor environment was designed and built in Unity 3D game engine.

Fig. 2. A user-view of the designed simulation system.

## Maps design

- The graph entropy, which is commonly used as the structural information content and the complexity of a graph is used to design four different map settings with a range of complexity.

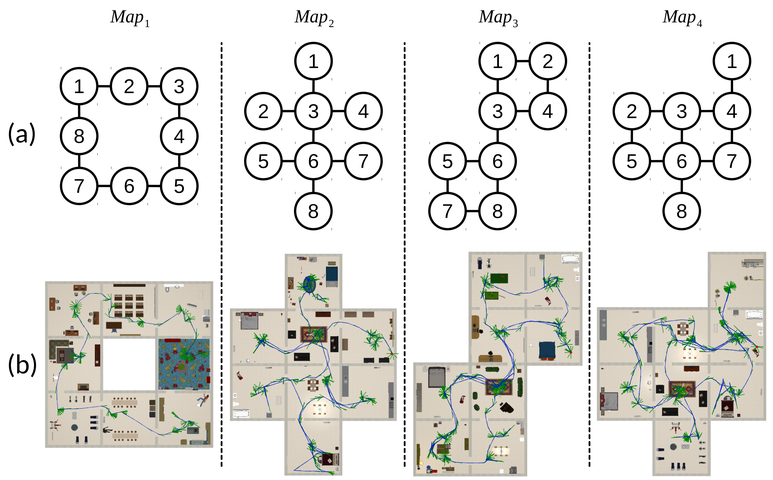
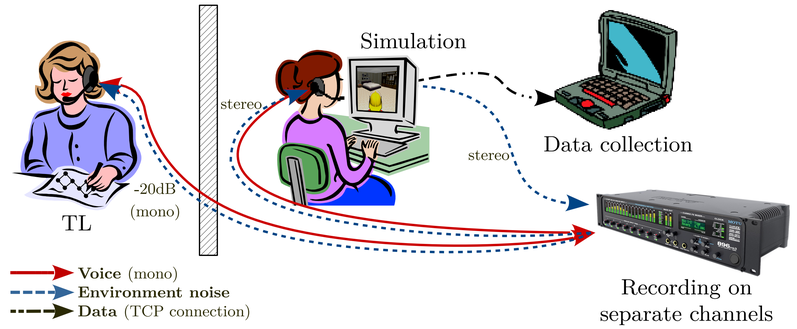



Fig. 3. (a) The topological structure of four different map settings; (b) corresponding top-view image of each map which are overlaid with the motion trajectory of a participant and her viewing directions (small arrows) at each time.

## Corpus recording

- Recordings were performed in two separate quiet rooms for avoiding external acoustic disturbances and crosstalk between the two speakers' voice.
- Each speaker's clean speech and the environment noise are available on separate channels.



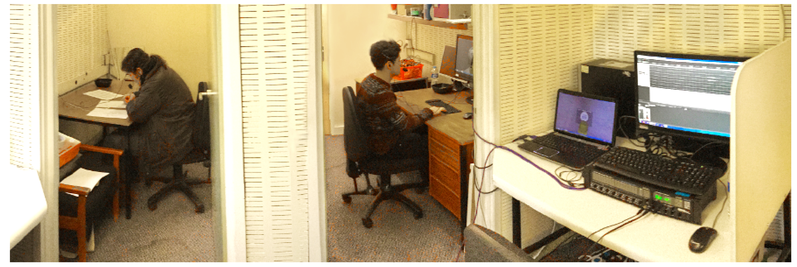



Fig. 4. top: the recording scenario, bottom: the recording set-up in two separate quiet rooms.

## Transcription and annotation

- The corpus totals about 80K words of manual transcription with ~16K vocabulary size, ~11K utterances and ~1K dialogue turns.
- Aligned with word level annotations, other information about the participants' locations, actions and objects in their field of view in the environment are available on computer readable log-files providing a form of conceptual annotation for the conversations.

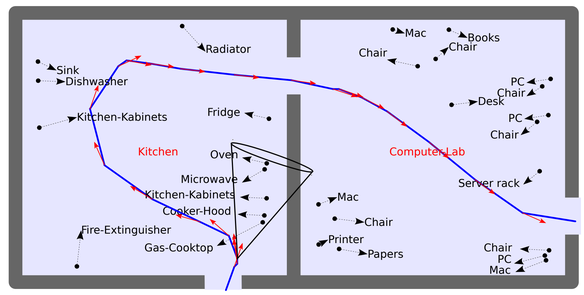



Fig. 5. An example of motion trajectory information plotted over the environment map, an instance of a participant's field of view and surrounding objects in the simulated environment.