FEARLESS STEPS CORPUS

UTDALLAS · UTD CRSS · RSTL · NASA · NSF

PRESENTED BY: CENTER FOR ROBUST SPEECH SYSTEMS (CRSS)
ROBUST SPEECH TECHNOLOGIES LAB (RSTL-CRSS),
THE UNIVERSITY OF TEXAS AT DALLAS

An Audio Corpus of the entire Apollo-11 Mission

TO BE RELEASED AT : INTERSPEECH 2018

19000 HOURS OF MASSIVE NATURALISTIC MULTI-CHANNEL DATA

NASA's Apollo program stands as one of mankind's greatest achievements in the 20th century. The CRSS Lab successfully digitized the mission audio tapes, and are now making the data publicly available, Intended to advance the Speech and Language Research Community

About the Corpus :

Full Corpus [19k hours]: Mission Control, Air-To-Ground, all Back-Room communications, IRIG timecode for 30 synch. channels

A 100 Hour Sub-Challenge Corpus will be Released with 5 time-synchronized Challenge Tasks established

Pipeline Diarization Transcripts will be provided in addition to the VAD and ASR Transcripts for the Corpus

The Challenge Corpus will be Open Source, Freely Distributed (with a small fee for Complete Corpus Delivery on Hard-Drive)

The Challenge Tasks :

1. Speech Activity Detection (SAD)
2. Speaker Diarization
3. Speaker Identification (SID)
4. Automatic Speech Recognition (ASR)
5. Keyword Spotting for Joint Topic and Sentiment Detection