Aalborg Universitet

AALBORG UNIVERSITY

DENMARK

Fearless Steps APOLLO: Identifying Conversational Mission-Critical Topics in NASA Apollo Missions Audio Based on Keyword Spotting

Joglekar, Aditya; Espejo, Ivan Lopez; Hansen, John H. L.

Link to publication from Aalborg University

FEARLESS STEPS APOLLO: IDENTIFYING CONVERSATIONAL MISSION-CRITICAL TOPICS IN NASA APOLLO MISSIONS AUDIO BASED ON KEYWORD SPOTTING

Aditya Joglekar[1], Iván López-Espejo[1,2], John H. L. Hansen[1]
[1]Center for Robust Speech Systems (CRSS), University of Texas at Dallas, Texas - 75080, USA
[2]Department of Electronic Systems, Aalborg University, Aalborg - 9220, Denmark
{aditya.joglekar, john.hansen}@utdallas.edu, ivl@es.aau.dk

INTRODUCTION

Conversational topic detection/identification is the process of determining the primary subject being discussed within a diarized segment from an input audio stream. The goal is to categorize spoken content into specific topics, making it easier to document, navigate, and analyze much larger volumes of conversational data. The Fearless Steps APOLLO (FS) community resource, sponsored by NSF, is an audio and meta-data collection established by CRSS-UTDallas comprised of naturalistic communications between NASA-MCC and backroom staff, and Astronauts across the manned Apollo Missions. This massive audio resource will house up to +150,000hrs of organized audio along with metadata, and is expected to benefit several research and education communities. Supplementing audio with meta-data developed using robust automated diarization and speech recognition to transcribe and highlight naturalistic communications can facilitate new research opportunities for SLT, speech sciences, education, and historical archival communities. In this study, we focus on customizing keyword spotting (KWS) and topic detection mechanisms, laying foundational groundwork for advanced conversational comprehension.

METHOD

Extensive research on automatic speech recognition (ASR), speech activity detection, and speaker diarization using sentence-level manual annotations from a 100hr subset of the Apollo 11 Mission audio has demonstrated the need for robust domain-specific speech model development [1–4]. A major challenge in training KWS systems and topic detection models is the availability of word-level annotations. Forced alignment schemes evaluated using state-of-the-art ASR show significant degradation in segmentation performance for naturalistic data. This study explores advanced techniques for accurately extracting keyword segments from existing sentence-level transcriptions and proposes domain-specific KWS-based solutions to detect conversational topics in audio streams [4,5]. Subsequently, using speech-to-text processing systems [3,4] to automatically generate transcripts from audio, the proposed study aims to extract keywords from the entire 11,000hr Apollo-11 Mission audio. We discuss a novel topic identification scheme that utilizes extracted keyword sequences in conversations to infer topic labels using advanced Deep Learning (DL) methods.

IMPACT

Identifying conversational topics in continuous streams of natural speech (e.g., non-prompted) can have a far-reaching impact for speech and language technology, historical archival, and information retrieval domains. The meta-data and deep learning models developed for both tasks provide opportunities for applications in education and outreach. These datasets will be made freely available worldwide for researchers to advance their research efforts, which include novel topics, thus allowing for improvements in natural language understanding and processing.

ILLUSTRATION

We will demonstrate baseline and advanced systems and transcriptions created for Keyword Spotting and Topic Identification. This will include the research conducted till date on the FS APOLLO corpus to extract meaningful communication tasks and performance analysis, in addition to techniques used to adapt and improve conventional speech-related tasks.

REFERENCES

[1] Hansen J.H.L. *et al.*: Interspeech 2018, 2758-2762. [2] Hansen J.H.L. *et al.*: Interspeech 2019, 1851-1855. [3] Joglekar A. *et al.*: Interspeech 2019, 2617-2621. [4] Joglekar A. *et al.*: Interspeech 2021, 986-990. [5] Joglekar A., López-Espejo I., Hansen J.H.L.: *The Journal of the Acoustical Society of America* 153 (2023): A173-A173.