FEARLESS STEPS APOLLO: TOWARDS COMMUNITY RESOURCE DEVELOPMENT FOR SCIENCE, TECHNOLOGY, EDUCATION, AND HISTORICAL PRESERVATION

John H.L. Hansen, Aditya Joglekar, Meena M. C. Shekar, Szu-Jui Chen, Xi Liu

Center for Robust Speech Systems (CRSS), Eric Jonsson School of Engineering,
The University of Texas at Dallas (UTD), Richardson, Texas, USA

ABSTRACT

The Fearless Steps Apollo (FS-APOLLO) resource is a collection of over 150,000 hours of audio, associated meta-data, and supplemental technological toolkit intended to benefit the (i) speech processing technology, (ii) communication science, team-based psychology, and history, and (iii) education/STEM, preservation/archival communities. The FS-APOLLO initiative which started in 2014 has since resulted in the preservation of over 75,000 hours of NASA Apollo Missions audio. Systems created for this audio collection have led to the emergence of several new Speech and Language Technologies (SLT). This paper seeks to provide an overview of the latest advancements in the FS-Apollo effort and explore upcoming strategies in big-data deployment, outreach, and novel avenues of K-12 and STEM education facilitated through this resource.

Index Terms— Massive Naturalistic Community Resource, NASA Apollo Missions, Fearless Steps (FS-APOLLO)

1. INTRODUCTION

The ability for teams to work collaboratively to learn, engage, and solve complex problems requires effective communication. This relies on vocal communications to identify common knowledge, member expertise, solution strategies, and a decisive implementation plan. As human-machine interaction continues to evolve, Speech & Language Technologies (SLT) become crucial for automatically & effectively extracting information from digital recordings of speech communications. SLT systems have historically focused on 1-way or 2-way voice communications (e.g., commands, speeches, broadcast news, telephone), typically face-to-face or over voice channels/cell phones (2-way). The primary obstacle in analyzing communications from naturalistic settings is the limited public access to such audio resources.

The Fearless Steps Apollo (FS-APOLLO) initiative bridges this gap by establishing a robust, open-access research corpus, the largest of its kind. This community-based project is focused on advancing science & technological knowledge to improve our understanding of team-based communications for solving complex problems [1, 2, 3, 4, 5]. Exploring this challenge through voice has direct benefits for three related communities: Speech & Language processing technology (SpchTech), Communication Science, Psychology/Small Group Teams (CommSciPsych), & Education/STEM, Preservation/History/Archiving, Community-use (EducHistArch).

Next-generation knowledge-based SLT systems will need to collectively analyze both individual and team-based communications. FS-APOLLO seeks to develop the largest naturalistic open-access multi-speaker audio research corpus for such SLT development, catering specifically to the SpchTech, CommSciPsych, and EducHistArch communities. The FS-APOLLO Community Resource establishes a sustainable multi-speaker task-based corpora generation process by advancing and deploying extensive machine learning technologies. The project aims to recover and distribute the entire team engagement communications (+150,000hrs of audio) from one of mankind's greatest scientific accomplishments over the past century – the NASA Apollo missions.

Restoring 50-year-old Apollo audio contributes to historical preservation. However, for it to be beneficial for the aforementioned communities, supplementing the audio with comprehensive meta-data is essential. Consequently, CRSS-UTDallas has focused on bolstering the necessary speech diarization infrastructure, unlocking new research & educational avenues across the three communities. Through multifaceted outreach initiatives, this resource has encouraged scientists, educators, & technologists to explore novel challenges in team communication [6, 7, 8, 9, 10]. Over the past decade, CRSS-UTDallas-led efforts have resulted in multiple state-of-the-art (SOTA) SLT solutions developed by us, and researchers worldwide through outreach efforts.

Massive audio (big-data) resource and limited human annotations largely fulfills the SpchTech community's demands. However, tailored efforts are essential to craft equivalent resources for the CommSciPsych and EducHistArch communities, a pivotal challenge being availability of multi-team

This project was supported by NSF-CISE Community Resource Project 2016725, and partially by the University of Texas at Dallas from the Distinguished University Chair in Telecommunications Engineering held by J.H.L. Hansen. We would like to acknowledge the immense effort of our previous FS-APOLLO team members A. Sangwan, C. Yu, L. Kaushik, C. Belitz, & M. Yousefi, and thank them for their contribution. We would also like to convey our special thanks to T. Nguyen, K. King, K. Foxworth, J. Chiu, and F. Carmona (CRSS-UTDallas Digitization & Transcription Team) for their digitization and meta-data development efforts for the FS-APOLLO Corpora.

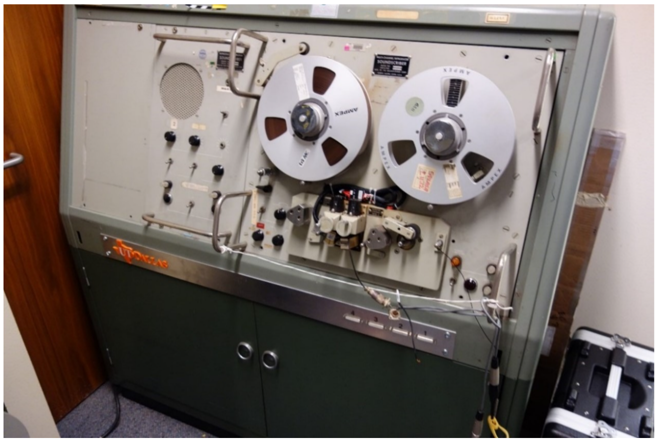

Fig. 1. SoundScriber system for processing analog tapes.

task-based audio that is not simulated. For this, FS-APOLLO aims at recovering all Apollo missions (all air-to-ground CAPCOM, Mission Control, backrooms) from 30-track analog tapes. Presently, 75,000hrs have been digitized from the analog tapes, and ongoing pipeline diarization process has generated transcripts for 19,000hrs of audio.

The subsequent sections of this paper are structured as follows: (i) Apollo Tape Recovery (Sec. 2), (ii) Overview of the FS-APOLLO audio characteristics (Sec. 3) (iii) Diarization Pipeline Processing / Meta-Data Creation (Sec. 4), and (iv) Community Engagement / Resources (Sec. 5).

2. DATA PROCESSING

The distinctive audio data from Apollo missions was continually captured using two 30-track analog reel-to-reel historical recording machines, which were known as Historical Recorders 1 (HR1) and 2 (HR2) [11]. The recordings encompassed entire communication loops between the astronauts and NASA Mission Control Center (MCC), ensuring comprehensive coverage. To maintain continuity, tapes were periodically changed throughout the process.

2.1. Data Digitization

Digitizing the original analog tapes can be accomplished directly through the use of the SoundScriber player, as depicted in Fig. 1, which is housed within the US National Archives & Records Administration (NARA) [12]. To preserve data synchronicity and enable the storage of each channel individually, a customized read-head for SoundScriber was designed. This innovation allows for the simultaneous digitization of all 30 channels [11]. Notably, Channel-1 on each tape is reserved for encoding Mission Elapsed Time (MET), serving as a crucial reference for synchronizing all channels.

2.2. Data pre-processing

The digitized data initially has a sampling rate of 44.1kHz and often spans more than 10 hours, which may not be ideal for speech analysis tasks due to its duration. Additionally, we've observed occasional waveform clipping in the digitized data,

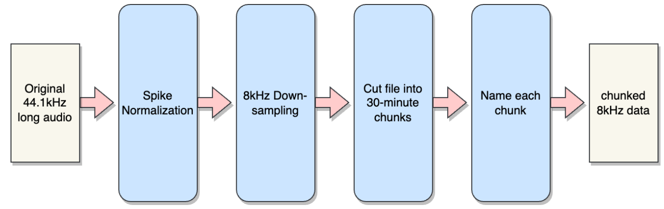

Fig. 2. Overview of the data pre-processing pipeline.

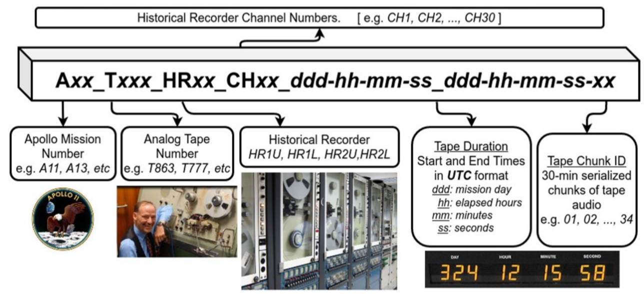

Fig. 3. Naming convention for each 30-minute chunk.

which likely occurs during the activation of the start/stop on the playback system. Thus, to enhance the usability and audibility of the audio data, we developed an automated pre-processing pipeline to further refine the original digitized recordings. An overview of this data pre-processing pipeline is presented in Fig. 2. To facilitate the identification of individual segments, we label the segmented 8kHz data with comprehensive tape details. This labeling creates a distinct identity for each segment, as elaborated in Fig. 3 [5, 13]. This approach enables individuals to easily discern historical tape information associated with specific segments.

3. FS-APOLLO AUDIO CHARACTERISTICS

The lunar missions' triumph was a testament to the dedication of over 400,000 personnel including NASA scientists, astronauts, and support staff, who supported the program in diverse roles. Transcribing interactions among these individuals requires a thorough understanding of the recordings' acoustic characteristics. As a results, pre-processed FS-APOLLO audio from the Apollo-11 mission was chosen for an initial assessment. Analysis of pivotal mission phases—such as Lift-Off, Lunar Landing, and Lunar Walking—from the Apollo-11 audio underscored the challenges of extracting valuable knowledge from such naturalistic data sources [1, 2].

The FS-APOLLO data stands out with its unique challenges, including various noise types, signal degradation, and numerous overlapping instances across all channels. Many audio streams suffer from degradation due to factors like system noise, attenuated bandwidth, cosmic disturbances, and analog tape decay. Furthermore, speech density, noise conditions, and noise levels fluctuate across time and channels. These characteristics often result in degraded performance for acoustic modeling-based systems like SAD, keyword spotting (KWS) and speech recognition (ASR). Time-critical mission elements often led to swift speaker transitions, with some responses lasting 0.2sec. For context, during mission status updates, speaker durations averaged around 0.5sec, and it wasn't uncommon to have 15 consecutive speakers in a mere 10-second window. At times, over 20 speakers might converse non-stop for 30 minutes, followed by prolonged silences. Given these challenges and complexities, it's imperative to develop tailored solutions to maximize the utility of this resource. Such dynamics pose significant challenges for speaker recognition (SR) and diarization systems.

Domain-specific system design is necessary to overcome these challenges. In the following section, we detail the novel techniques developed to provide robust speech recognition and diarization solutions for this data.

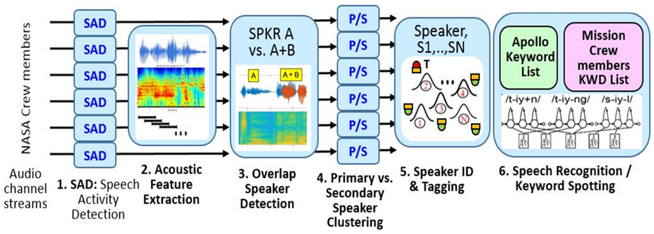

Fig. 4. Overview of the diarization pipeline.

4. SPEECH TECH PIPELINE

Our pipeline diarization system (Fig. 4) comprised of three key components, i.e. SAD, speaker clustering, and speech recognition, is introduced in this section.

4.1. Speech Activity Detection

Speech Activity Detection (SAD) serves as an essential precursor for subsequent speech and language processing tasks. DeepComboSAD [14], a supervised deep-learning (DL)-based SAD system with learnable spectro-temporal features was designed as the initial front-end system for the pipeline. The proposed system learns features in an end-to-end (E2E) manner by leveraging convolutional and recurrent neural networks combined with correlations in time and frequency domain. Our DeepComboSAD solution achieves 3.42% DCF (detection cost function) on the FS Challenge Phase-4[1] evaluation set, which is over 50% relative improvement from the previous SOTA systems developed for this data.

4.2. Finding Waldo and Speaker Clustering

Our goal is to (i) create usable meta-data for EducHistArch and CommSciPsych communities, and (ii) provide tags that LDC and NIST could use for future real-world speaker recognition evaluations (i.e. NIST SRE) [15]. We started by analyzing and assessing the baseline SR models on the Apollo-11 portion of the Fearless Steps Corpus. This provides a foundation to identify, analyze, and track key NASA mission specialists across missions. The concept of 'Finding Waldo' was used to track and tag five speakers-of-interest using the baseline SR system, for the three Apollo-11 mission phases mentioned in Sec. 3 [16, 17]. Duration for primary vs. secondary speakers was analyzed, as well as speaker turns according to speaker roles on Apollo-11. Here, the focus has been on developing an advanced speaker clustering solution that would allow grouping key speakers across full missions, or across multiple missions. Furthermore, a novel Graph Attention Network was proposed to overcome the varying utterance duration problem that occurs in FS-APOLLO corpus which achieves a classification accuracy of 79.86% for 140 speakers. 15 speakers of interest were identified from three phases of the mission [18]. This advancement aids in understanding the structure of the communication team protocols used by NASA personnel, which would allow future researchers to derive contextual hypotheses.

4.3. Robust Speech Recognition

To provide searchable content for researchers, we developed automatic speech recognition (ASR) systems to convert audio into text as part of the meta-data. A hybrid ASR system, which is an expansion from our earlier formulated ASR engine tuned to Apollo-11 (developed in 2015), is first built as the new baseline [19]. We incorporate a scenario vector trained using the Triplet-loss function as a part of the input feature for the acoustic model. This results in a 26.16% word error rate (WER) on the development set of the FS Challenge Phase-2 corpus, representing a significant 52.4% absolute improvement compared to our earlier ASR engine. Later on, we explored the modern E2E architecture built using the popular self-supervised learning representations (SSLRs) as input features. Several feature combination strategies for these SSLRs are compared and discussed [20]. With the advancement in the ASR system, we can generate the next-generation transcripts for all the digitized Apollo missions.

5. COMMUNITY OUTREACH

5.1. Undergraduate CS Senior Design Teams

In an effort to build upon the existing capabilities of Explore-Apollo.org[2], multiple senior design teams across different semesters were supported. The UTD-CS teams in Fall'21 and Spring'22 were mainly responsible for the ExploreApollo.org website resources, LanguageARC[3], and the Apollo 8 content. They laid the foundation for audio and meta-data uploads to the website, particularly creating tasks on LanguageARC for crowd-sourcing meta-data [21, 22]. The involvement of UPenn's Linguistic Data Consortium (LDC), under the guidance of Dr. Chris Cieri, was pivotal in the success of the LanguageARC platform which is a critical feature for both researchers and general users [23, 5]. Currently, this consists of three tasks citizen-scientists can contribute to: audio quality analysis, speaker count identification, and audio transcription. Previous CS Senior Design Team in Fall'20 addressed fundamental concerns such as software compatibility, website dependencies, and updating Ruby for long-term

[1]https://fearless-steps.github.io/ChallengePhase4

[2]app.exploreapollo.org

[3]https://languagearc.com/projects/21

maintenance. They also introduced new informative tabs on the website and contributed to educational outreach by creating a "Lessons Plan" page aimed at K-12 teachers. The Spring 2021 team took the utility a notch higher by designing an audio player capable of handling and playing multiple audio streams simultaneously. They made both front-end and back-end changes to ensure compatibility across all Apollo missions, resolving prior limitations related to performance issues and transcript synchronization. These enhancements make it easier for users to interact with the website, providing a more comprehensive view of the Apollo missions.

5.2. MCC Console mock-up for K-12 Engagement.

Understanding the achievements of the Apollo missions and immersing oneself in the challenges faced and surmounted by the NASA MCC personnel offers a profound learning experience. This exposure aims to inspire K-12 students, fostering a deeper appreciation for the fields of science, technology, engineering, and mathematics. By connecting with this rich history, we hope to bolster their engagement in STEM disciplines, instilling a sense of curiosity and determination to tackle future challenges. Consequently, a 5-member ECE senior design team was recruited from University of Houston, Texas (UofH) to assist with a museum exhibit development. The UofH team devised a hardware display for time-synchronization data decoding, and crafted an interactive software for a scaled NASA MOCR control station replica (shown in Fig. 6), simulating the audio playback experience of the Mission Operation Control Room (MOCR). Using MATLAB, the UofH team produced a software solution and design blueprints for the physical model, laying the foundation for the next construction phase. Fig. 6 presents snapshots from their live demos. The next phase in this ongoing effort will focus on building a mock-up console using the blueprints, and installing the software developed by the UofH team. The concluding stage of this initiative includes setting up MCC mock-ups in prominent public science museums (such as Perot Museum Dallas & Space Center Houston) as interactive annotated-audio-playback consoles geared towards enhancing STEM involvement among K-12 students.

5.3. Fearless Steps (FS) Challenge Series

To promote research on naturalistic speech, CRSS-UTDallas hosted a series of SLT challenge tasks utilizing human annotated portions of the FS-APOLLO data. This effort was named the Fearless Steps Challenge (FSC), with annually organized events released as FSC Phases. With each FSC Phase (four have been organized so far), progressively complex 100hr datasets and associated annotations for six challenge tasks (including SAD, speaker diarization, ASR, & sentiment detection) were publicly released. The audio and meta-data were freely distributed to individuals interested in receiving the data for their personal use, or to participate in the challenge tasks. Over the FSC Phases (1-4), over 500 sites (from industry/academia/independent organizations) utilized this resource, leading to the development of multiple novel SLT systems and interactive educational websites [2, 3, 4, 21, 24].

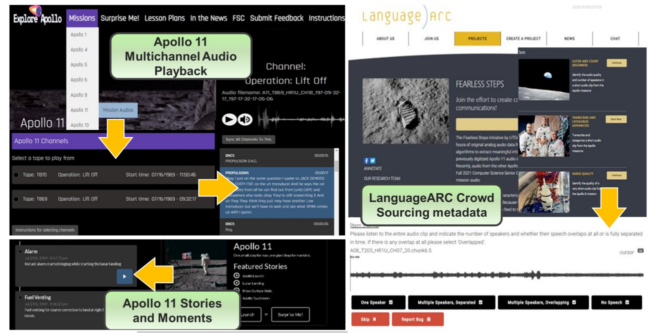

Fig. 5. Overview of key features of ExploreApollo.org *(left)* and LanguageARC *(right)* platforms. ExploreApollo.org offers multi-channel audio playback, including special 'Stories/Moments' from the Apollo-11 mission. LanguageARC facilitates crowd-sourced meta-data collection, allowing users to listen to mission audio and complete specific tasks.

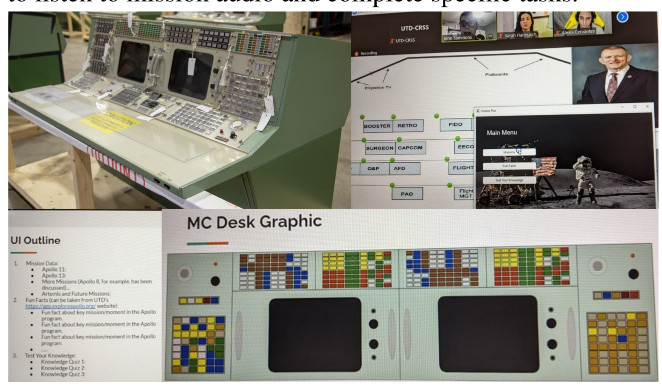

Fig. 6. *Top Left*: The original MCC console from NASA, Houston, Texas *Bottom Left*: Project outline for the Senior Design effort. *Top Right*: The final UofH Project Presentation (team members displayed on top panel). *Bottom Right*: Graphical illustration of the Mission Control Center console. The software demo was built to be displayed through the two screens mounted on the replica.

6. SUMMARY

The FS-APOLLO community resource has successfully established a massive, open-access public repository that serves multiple disciplines. Designed to benefit not only researchers and academics across various fields but also students from diverse age groups. This initiative has led to advancements in speech diarization technologies and offers a venue for studying team-based communications in naturalistic settings. Ongoing efforts will focus on expanding educational outreach, making this resource invaluable for K-12 and STEM education. Additionally, the focus will be on enhancing and further developing advanced semantic and contextual meta-data to augment the value of this ongoing effort. This initiative lays a foundational platform for future research, technology development, and educational programs, thereby making a lasting impact on both current and future generations.

7. REFERENCES

[1] John H.L. Hansen, Abhijeet Sangwan, Aditya Joglekar, et al., "Fearless Steps: Apollo-11 Corpus Advancements for Speech Technologies from Earth to the Moon," in *Proc. Interspeech*, 2018, pp. 2758–2762.

[2] John H.L. Hansen et al., "The 2019 Inaugural Fearless Steps Challenge: A Giant Leap for Naturalistic Audio," in *Proc. Interspeech*, 2019, pp. 1851–1855.

[3] Aditya Joglekar, John H.L. Hansen, Meena MC Shekar, et al., "FEARLESS STEPS Challenge (FS-2): Supervised Learning with Massive Naturalistic Apollo Data," in *Proc. Interspeech*, 2020, pp. 2617–2621.

[4] Aditya Joglekar, SO Sadjadi, Meena MC Shekar, Christopher Cieri, and J.H.L. Hansen, "Fearless Steps Challenge Phase-3 (FSC P3): Advancing SLT for Unseen Channel and Mission Data Across NASA Apollo Audio," in *Proc. Interspeech*, 2021, pp. 986–990.

[5] John HL Hansen et al., "Fearless steps apollo: Advanced naturalistic corpora development," in *Proc. 2nd Workshop NIDCP within LREC*, 2022, pp. 14–19.

[6] Per Fallgren, Zofia Malisz, and Jens Edlund, "How to annotate 100 hours in 45 minutes," *Proc. Interspeech 2019*, pp. 341–345, 2019.

[7] Anastasios Vafeiadis, *Machine Learning for Human Activity Detection in Smart Homes*, Ph.D. thesis, De Montfort University, 2020.

[8] Jens Heitkaemper, Joerg Schmalenstroeer, Joerg Ullmann, Valentin Ion, and Reinhold Haeb-Umbach, "A Database for Research on Detection and Enhancement of Speech Transmitted over HF links," in *Speech Communication; 14th ITG Conference*, 2021, pp. 1–5.

[9] Per Fallgren and Jens Edlund, "Edyson: rapid human-in-the-loop browsing, exploration and annotation of large speech and audio data," 2022.

[10] Matthew Lehet et al., "Circumspection in using automated measures: Talker gender and addressee affect error rates for adult speech detection in the Language ENvironment Analysis (LENA) system," *Behavior research methods*, vol. 53, no. 1, pp. 113–138, 2021.

[11] Lakshmish Kaushik et al., "Multi-Channel Apollo Mission Speech Transcripts Calibration," in *Proc. Interspeech*, 2017, pp. 2799–2803.

[12] Douglas W Oard, John HL Hansen, et al., "Toward access to multi-perspective archival spoken word content," in *ICADL, Proc. 18*, 2016, pp. 77–82.

[13] Chelzy Belitz and J.H.L. Hansen, "Challenges in Metadata Creation for Massive Naturalistic Team-Based Audio Data," in *Proc. Interspeech*, 2022, pp. 5210–5214.

[14] Aditya Joglekar and John HL Hansen, "DeepComboSAD: Spectro-Temporal Correlation based Speech Activity Detection for Naturalistic Audio Streams," *IEEE Signal Processing Letters*, 2023.

[15] Craig S Greenberg, Douglas A Reynolds, et al., "The NIST 2014 speaker recognition i-vector machine learning challenge," in *ISCA Odyssey*, 2014, pp. 224–230.

[16] Meena M.C. Shekar and J.H.L. Hansen, "Historical Audio Search and Preservation: Finding Waldo Within the Fearless Steps Apollo 11 Naturalistic Audio Corpus ," *IEEE Sig. Proc. Mag.*, vol. 40, no. 3, pp. 30–38, 2023.

[17] Meena Chandra Shekar, "Knowledge Based Speaker Analysis Using a Massive Naturalistic Corpus: Fearless Steps Apollo-11," *Master's Thesis, The University of Texas at Dallas*, 2020.

[18] Meena M.C. Shekar and John H.L. Hansen, "Speaker Tracking using Graph Attention Networks with Varying Duration Utterances across Multi-Channel Naturalistic Data: Fearless Steps Apollo-11 Audio Corpus," *Proc. Interspeech*, pp. 1459–1463, 2023.

[19] Szu-Jui Chen, Wei Xia, and John HL Hansen, "Scenario Aware Speech Recognition: Advancements for Apollo Fearless Steps & CHiME-4 Corpora," in *IEEE ASRU*. IEEE, 2021, pp. 289–295.

[20] Szu-Jui Chen, Jiamin Xie, and John HL Hansen, "FeaRLESS: Feature Refinement Loss for Ensembling Self-Supervised Learning Features in Robust End-to-end Speech Recognition," *Proc. Interspeech*, 2022.

[21] James Fiumara, Christopher Cieri, Jonathan Wright, and Mark Liberman, "LanguageARC: Developing Language Resources through Citizen Linguistics," in *Proceedings of the LREC Workshop Citizen Linguistics in Language Resource Development*, 2020.

[22] Christopher Cieri and James Fiumara, "LanguageARC - a Tutorial," in *Proceedings of the LREC 2020 Workshop on" Citizen Linguistics in Language Resource Development"*, 2020, pp. 58–64.

[23] James Fiumara, Christopher Cieri, Mark Liberman, Chris Callison-Burch, Jonathan Wright, and Robert Parker, "The NIEUW Project: Developing Language Resources Through Novel Incentives," in *Proc. 2nd Workshop NIDCP within LREC*, 2022, pp. 1–7.

[24] "NIST Fearless Steps Challenge Phase-03 2021 Webplatform," `https://sat.nist.gov/fsc3`, Accessed: 2022-03-01.