Thirteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)

Characterizing National Airspace System Operations Using Automated Voice Data Processing

A Case Study Exploring Approach Procedure Utilization

Shuo Chen, Hunter Kopald, Rob Tarakan, Gaurish Anand, Karl Meyer
The MITRE Corporation
McLean, Va.
chen@mitre.org

***Abstract*—Air Traffic Control (ATC) radio communications contain a wealth of situational context information. While valuable, this information resource has been difficult and expensive to use for large scale analyses because raw speech audio cannot be directly used in analyses without human or computer interpretation. To help the Federal Aviation Administration (FAA) better understand National Airspace System (NAS) dynamics, The MITRE Corporation (MITRE) has been developing voice data analysis capabilities that can enable information from ATC voice communications to be automatically processed on a large scale and used in post-operational analyses. These capabilities use an array of technologies to segment audio data by speaker role, transcribe the audio to text, and extract semantic entities such as aircraft identifiers and clearances. The data derived by these capabilities can inform large-scale analyses, augmenting existing data sources such as radar tracks and flight plans, and enable studies and the generation of metrics that were previously impractical. This paper describes these voice data processing capabilities and presents one example of the use of voice data: to enable better understanding of Performance-Based Navigation (PBN) procedure utilization in the NAS. This paper describes an initial use of voice data analysis to better understand approach procedure utilization, which opens the door for many new analyses.**

Keywords-automatic speech recognition, natural language processing, neural networks, air traffic control, performance-based navigation

I. INTRODUCTION

Air Traffic Control (ATC) radio voice communications are a critical mechanism for coordinating aircraft movement through the National Airspace System (NAS) and will continue to be even with the expected increase in the use of Data Communications in the future. Controller-pilot communications provide key information on the flight's intent, controller interventions, and the ultimate outcome of each flight operation. In some cases, voice communications offer additional insight into a given event; in other cases, voice communications are the only data source that holds certain information about what happened and why. However, recorded controller-pilot voice communications have been difficult to use for large-scale analyses because raw speech audio cannot be directly used in analyses without human or computer interpretation.

The Federal Aviation Administration (FAA) has been investigating how automatic speech recognition and language processing technologies can be used to extract information from ATC voice communication, specifically to improve the safety of NAS operations in real time and to better assess NAS operations through post-operations analyses. On behalf of the FAA, The MITRE Corporation (MITRE) has been researching and developing ATC-specific applications of these technologies. This research includes improving speech recognition and language processing accuracy and evaluating how particular applications can benefit the FAA enterprise. Some applications require voice data processing in real time—e.g., to detect a controller clearance that may result in a safety risk, such as an instruction for an aircraft to line up and wait where an arrival is on short final approach to the same runway, and immediately alerting the controller to that issue. Other applications involve post-operations analysis of the NAS using automatic processing of large quantities of audio data, often in conjunction with other aviation data sources.

This paper describes automated voice data analysis capabilities that MITRE has developed on behalf of the FAA and how the resulting information can be used to better understand the NAS. The goal of this paper is to expose and promote the value of voice data analysis. To that end, we describe not only the analysis capabilities we have developed, but also specific use cases of the data towards operational analysis of Performance Based Navigation (PBN) and other approach procedures. Large-scale ATC voice data analysis is still relatively new, but we hope this paper provides insight to organizations around the world on ways to integrate voice data into their research, analysis, and decision-making to improve the safety and efficiency of global air traffic management.

Section II provides background on the FAA voice switch and recording infrastructure, the PBN research needs that voice data analysis can help address, previous research on ATC speech recognition, and the challenges unique to large-scale processing of ATC voice data. Section III describes the voice data analysis capabilities that MITRE has developed and is continuing to improve. Section IV describes how the voice data analysis

capability is then leveraged to better understand approach procedure utilization in the NAS. Section V describes next steps.

II. BACKGROUND

A. FAA Voice Recording Infrastructure

DALR (Digital Audio Legal Recorder) is an FAA system for capturing, compressing, encoding, and storing controller-pilot voice communications within a facility [1]. DALR retains the most recent 45 days of recorded audio for legal purposes. In addition, many facilities locally use recorded DALR audio for quality control, training, and other purposes. Users at a facility retrieve audio from the facility's DALR system via a text-based user interface on a proprietary, dedicated DALR computer. The DALR system itself does not provide a mechanism for remote or automated (i.e., computer-to-DALR) access to recorded audio.

Each facility has the discretion to select what audio to record. For example, a facility may record all audio received at the controller position, which can include both air/ground (i.e., speech between controller and pilot) and ground/ground (i.e., speech between different controllers, either within or between facilities) communications, or just audio transmitted and received over the radio, which would include only air/ground communications for specific frequencies. Regardless of the source of the audio, DALR does not record the push-to-talk information that delineates the start and end of each controller radio transmission. When both controller and pilot audio is recorded on the same channel, which is the typical recording configuration, DALR does not retain information about which speech is from the controller and which speech is from pilots. To reduce storage requirements, the DALR system stores continuous stretches of non-silence audio in individual files, retaining the wall-clock time associated with the audio as metadata. These files do not correspond directly to individual transmissions—one DALR file may be eight seconds long and contain one controller transmission and pilot readback, while another DALR file may be two minutes long and contain many transmissions.

The DALR system identifies the stored audio with a unique channel number and allows a facility to configure a channel map that specifies the control position code associated with each channel number. This mapping is valuable when identifying the ATC sector, ATC position, or other FAA ATC identifier associated with an audio recording.

DRAAS (DALR Remote Audio Access System) is a new FAA system designed to overcome some of the access limitations of the DALR system by providing a mechanism for remote and automated access [2]. In addition, through the FAA Comprehensive Electronic Data Analysis and Reporting (CEDAR) system, DRAAS can access the facility channel maps that document the facility sector or position associated with each DALR channel number [3]. Through the DRAAS interface, facility DALR recordings can be retrieved remotely using the facility name, DALR channel number, and a date-time period. Currently, DRAAS provides access to audio from 129 NAS facilities; more than 200,000 hours of silence-reduced audio are recorded each month.

Research sponsored by the Federal Aviation Administration

B. Performance Based Navigation (PBN) Research Need

The FAA has major challenges in meeting future demand for airport and airspace resources, while balancing its need to protect the environment, reduce traffic delays, and improve operational safety. The FAA is addressing these issues through the NextGen Air Transportation System (NextGen), which relies heavily on PBN procedures and optimized airspace. PBN leverages modern navigation technology to facilitate aircraft flying more direct routes and conforming tightly to planned paths. The navigational technology needed for some PBN procedures, the complexity in developing procedures, and the difficulty in integration of PBN operations brings significant costs to both the FAA and NAS users.

The FAA has developed a PBN NAS Navigation Strategy that outlines a roadmap for deployment and maintenance of navigation services, along with goals for increased usage of and conformance on PBN procedures. With the implementation of PBN procedures, there will also be a pressing need to reduce the complexity and cost to the FAA of maintaining legacy Instrument Flight Procedures (IAPs). The FAA is implementing several plans to reduce maintenance costs, including the Very High Frequency (VHF) Omni-directional Range (VOR) Minimum Operational Network (MON) and a process for cancellation of approach procedures as part of the National Procedures Assessment (NPA)

For the reasons outlined above, the FAA and NAS stakeholders have an ongoing need to track the current state of PBN in the NAS and understand PBN procedure conformance and usage in detail. To support that need, MITRE has developed automated capabilities and metrics to characterize PBN operations in the NAS, using a fusion of trajectory-based data, aircraft intent (filed/amended flight plans), and aircraft equipage data.

However, there are significant challenges and limitations associated with tracking procedure conformance and usage through radar track and flight plan data. Approach procedures present a particular challenge because the flight plan does not contain the expected or cleared approach. Relying solely on radar track data to determine approach procedure conformance and usage is not sufficient, because overlays (where approaches have very similar or identical lateral paths) and common waypoints may exist between a conventional and a PBN approach procedure, or vectors to Instrument Landing System (ILS) approaches may be similar to defined PBN paths. Further, aircraft flying straight-in approaches without a downwind leg often follow a very similar trajectory regardless of approach type, making it sometimes impossible to determine usage based solely on track points. As a result, using track data alone, approach procedure usage can only be determined for Required Navigation Performance (RNP) approaches that contain a Radius-to-Fix (RF) leg, because the RF legs provide a procedure geometry (usually not similar to other approach paths) that can be monitored for conformance. RNP approaches with RF legs represent about 1 percent of Instrument Flight Rules (IFR) arrivals in the NAS, thus motivating an exploration of additional information and techniques that can be used to determine approach procedure usage across a wider range of arrivals and approach types.

Controller-pilot voice communications provide an alternative source for determining approach usage. All arrivals flying under IFR where radar services are provided by ATC are issued approach clearances, which include the name (or type) of the approach and arrival runway. To discover the approach clearance for a single flight, manually listening to audio of the controller-pilot interaction is sufficient. However, to track approach procedure utilization across the NAS and over a period of time, and then correlate utilization to other flight metrics (e.g., safety events), manual transcription is not practical. Voice data analysis capabilities that can process ATC communications on a large scale present a solution to this analytical shortfall.

C. Computational Linguistics in ATC

Computational linguistics, of which automatic speech recognition is a subfield, has been present in ATC since the late 1980s [4]. With modern advances in computational linguistics areas such as speaker diarization, automatic speech recognition, semantic parsing, and dialogue modeling [5][6][7][8][9], the research and use of these technologies in the ATC domain has become increasingly prevalent. Although early use cases were focused around training and simulation, where the phraseology could be tightly controlled [10], more recent use cases have diversified into operational scenarios that included controller workload estimation [11], safety monitoring [12][13], and assistant-based efficiency improvements [14], including controller workload reduction by automatically updating clearances in radar data blocks [15].

Automatic speech recognition remains a focal area for research because of how different ATC radio communications are from conversation speech. The speed of speech, the brevity of the structured phraseology, the variations in speaker style by region, and the audio fidelity of ATC radio communications all negatively impact speech recognition accuracy. Furthermore, although some applications can be designed to anticipate and mitigate the impact of errors in speech recognition, many applications in the ATC domain require higher recognition accuracy in order to attain user acceptance. Most recently, Kleinert et al. examined the benefits of performing semi-supervised, site adaptation with context incorporation in their Assistant Based Speech Recognizer and observed notable accuracy improvements [16]. In 2018, Airbus, in collaboration with Institut de Recherche en Informatique de Toulouse (IRIT) and Safety Data Analysis Services hosted an ATC speech recognition challenge focused on audio transcription and callsign extraction of ATC audio at the flight deck [17][18]. The winners of the challenge, a joint team from Vocapia Research and Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (CNRS-LIMSI), achieved very good accuracy with word errors rates between six and eight percent using deep neural network (DNN) models for speech recognition and a combination of regular expressions and consensus network search for callsign extraction [18][19].

D. Challenges Unique to Large-Scale Voice Data Processing

This large-scale processing of audio data has presented some challenges that we had not encountered while researching real-time speech recognition applications. These challenges involve dealing with scalability and audio quality.

One challenge comes from scaling up to handle over 3,000 channels with differing acoustic and language characteristics. It is not practical for us to perform the same degree of manual channel/facility/sector/position-specific customization that we were able to do with single-facility audio. We do perform some site adaptation leveraging digital data sources, but other adaptation techniques would require manual effort beyond our current resources. Another scalability challenge is finding the correct pronunciations for the over 40,000 waypoint and procedure names that could be spoken throughout the NAS.

Another set of challenges come from the audio path. The systems that record the audio mix audio from multiple sources (i.e., pilot, controller, intercom) and discard the push-to-talk information delimiting individual transmissions. The systems that record and distribute the audio perform multiple lossy encode-decode format conversions, with one step using a particularly high-loss compression algorithm.

One advantage of this large-scale post-processing is that we can bring in non-voice ATC context information and we can use algorithms that look both forward and backward in time to aid with accurate speech recognition. For example, we can know an aircraft's arrival time (and sometimes runway) when we try to recognize its landing clearance. Another advantage is that, while we do need to concern ourselves with overall computer resource usage, we do not need to worry about lag time processing individual transmissions.

III. Voice Data Analysis Capabilities

MITRE's voice data analysis capabilities process controller-pilot voice communications on a large scale to generate text transcripts and extract valuable aviation information that can facilitate post-operations analysis. Using a combination of digital signal processing (DSP) and machine learning (ML) techniques, these capabilities form an audio processing pipeline (Figure 1) that ingests audio files, splits them into audio segments corresponding to individual radio transmissions, and produces text transcriptions, speaker role labels, and semantic tags identifying the presence of key aviation concepts including aircraft identifiers (ACID) and clearances for each audio segment. It also fuses these speech-derived artifacts with surveillance and other data sources to enable enhanced aviation analyses.

The audio processing pipeline is illustrated in Figure 1.

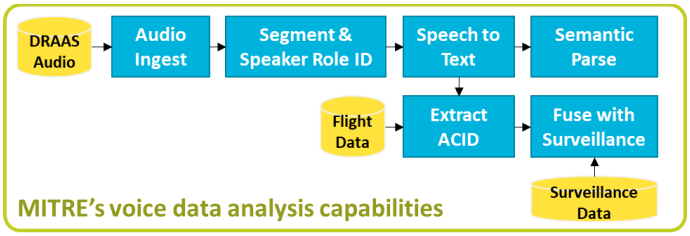

Figure 1. Voice Data Processing Pipeline

To keep up with the rate of audio data generation in the NAS, this audio processing pipeline is hosted on an Apache Hadoop

cluster of computers at MITRE. The DNN models used for speech recognition are generated in MITRE's High-Performance Computing (HPC) facility. The subsections below describe individual components in the audio pipeline, highlighting the difficulties and areas for improvement in specific steps.

A. Audio Acquisition and Ingest

Audio acquisition and ingest is the first component in the voice data analysis capabilities audio pipeline. This component retrieves recorded facility audio files through the DRAAS interface and ingests it into the pipeline [3]. As part of the ingest process, each audio file is packaged with descriptive metadata such as its source facility, radio channel, controller position, and start and end times of the audio files.

Although the audio data undergo decompression at this stage, it is not otherwise changed. As such, each audio file retrieved through the DRAAS interface is similar to the recordings retrieved from the DALR system, i.e. it has variable duration, it may contain speech data from one or more speaker turns (e.g., each controller and pilot transmission is considered a speaker turn), and each speaker turn may be generated by a different speaker. Subsequent components in the pipeline will pinpoint these characteristics and add them to the metadata associated with the audio data.

Currently, the voice data analysis capabilities pipeline ingests audio from 129 FAA ATC facilities, including all Air Route Traffic Control Centers (ARTCCs), most major Air Traffic Control Towers (ATCT), and most major Terminal Radar Approach Control facilities (TRACON).

B. Speaker Role Identification and Segmentation

Individual DRAAS audio files do not directly correspond to individual controller and pilot transmissions. The files are formatted to reduce storage and bandwidth requirements and do not retain any controller or pilot push-to-talk metadata. The speaker role identification component of the pipeline segments DRAAS audio files into individual transmissions and distinguishes the speaker as either a controller or a pilot, as illustrated in Figure 2.

Figure 2. Example Segmentation and Speaker Role Identification

The component employs an acoustics-based approach using speech features to identify when individual transmissions begin and end. It then classifies audio as either controller, pilot, or non-speech, and then post-processes the classifications to determine when transmissions begin and end.

This component is built on the open source software CMU Sphinx-4 from the Carnegie Mellon University [20]. CMU Sphinx is a large-vocabulary, speaker-independent, continuous speech recognition system based on Hidden Markov Models. Custom models trained on a corpus of controller-pilot audio is used with the Sphinx recognition engine to perform speaker role identification. The training corpus consists of audio segments, each manually labeled as a controller transmission, pilot transmission, or a non-speech audio segment.

Diarization Error Rate (DER) is a frame-by-frame measure of this component, where frames are typically 10 milliseconds. However, when human transcribers create the "ground truth" segmentation reference used the measure DER, there is some amount of subjectivity in deciding where each transmission segment starts and ends. Given that many ATC transmissions are just a few seconds long, errors from the subjective placement of the start and end times can significantly inflate the overall error numbers, to the extent that they overwhelm the more relevant errors. As an alternative to basic DER, we have evaluated the component with respect to segmentation (is it producing the correct number of segments for a given DRAAS file) and speaker role identification (is it producing the correct speaker label for a given segment). With both measures, the accuracy is approximately 85-90 percent, with variations between test samples including facility and speech turn patterns (e.g., instances with adjacent pilot transmissions back-to-back).

C. Automatic Speech Recognition

The speech recognition component converts audio to text, producing a text string for each audio segment that can then be parsed for relevant information. This component is built on the Kaldi speech recognition engine [21]. Kaldi is an open-source software package developed and maintained by Johns Hopkins University. It was created for speech recognition research and as such is a versatile software package with domain agnostic algorithms that can be quickly deployed in new domains with different recognition models.

For the ATC domain, MITRE created a Chain Time-Division Neural Network (Chain TDNN) [22][23] acoustic model (AM) and a statistical language model (SLM) trained on a corpus of transcribed controller-pilot audio. The training corpus consists of approximately 500 silence-reduced hours of transcribed controller-pilot audio that was collected from a variety of different audio sources using different types of recording hardware. Most of the audio in the training data corpus was acquired from DRAAS and has acoustic characteristics similar to those in the target audio.

The same AM and SLM are used for speech recognition on audio from all 129 facilities processed by the pipeline. The speech recognition component achieves word accuracy around 85 to 91 percent, with accuracy on controller speech generally better than accuracy on pilot speech. In addition, accuracy varies between facilities. We have tested accuracy against more than 20 ATC facilities and have observed accuracy on controller speech between 83 and 95 percent, and on pilot speech between 75 and 89 percent. The variance in word accuracy results from variations in speakers, pronunciations, phraseology, and acoustics across facilities as well as the amount of facility-specific data in the training corpus. Test set size and composition also influences the variations in accuracy measurement.

D. Aircraft Identifier Extraction

Aircraft identifier (ACID) extraction is a subcategory of semantic content extraction, however, the importance of this

particular semantic concept to analysis, subsequent data fusion, and overall usability of the speech-derived data warrants particular attention.

Because the word accuracy of the DNN-based recognition is fairly high, it is possible to assume that the ACID phrase recognized in a transmission is mostly correct, with only one or two single word errors. Thus, this component approaches ACID extraction post-recognition by attempting to correct errors in the recognized ACID phrase. It maps the ACID phrase to its nearest match in a candidate ACID list that consists of the list of aircraft in the control sector at the time of the radio transmission. Specifically, the algorithm first transforms consecutive sequences of airline callsign words, aircraft type words, and alphanumeric words into parsed ACIDs in symbolic form and compares the parsed symbolic form to a list of candidate ACIDs in their symbolic form as well to find the nearest match, i.e. with the fewest number of omissions, substitutions, or additions. For example, if the phrase "delta twenty oh five" was recognized, but the controller actually said "delta twenty one oh five", the algorithm would first transform the recognized phrase into the symbolic ACID form "DAL2005", then correct it to "DAL2105" after comparison to the candidate ACID list, assuming "DAL2105" is the closest match to the parsed ACID. The candidate list is updated periodically using a combination of track data and transfer of control messages. This approach is also adaptable to accepted truncations of general aviation callsigns. For example, it can correctly map "november three alfa whiskey" to the complete ACID symbolic form "N313AW" even though the spoken general aviation identifier was acceptably truncated to just the last three symbols in the tail number.

When evaluated on a test set of over 1,300 recognized text transmissions from eight tower and TRACON facilities, the ACID extraction component yielded an average accuracy of 85 percent on controller-spoken transmissions and 70 percent on pilot-spoken transmissions. Analysis of the error cases showed that the ACID extraction component is still susceptible to several types of error. First, because this component works on the recognized text transcription (i.e., the output of the speech-to-text component), the recognizer's accuracy will affect the accuracy of the ACID extracted. There is a noticeable drop in word accuracy when there is unusual distortion or noise in the audio; and when the speaker mumbles or trails off at the end of a transmission. Second, because the ACID phrase is parsed independent of other semantic content, transformation of an incorrect alphanumeric sequence, such as an altimeter setting or radio frequency, within the recognized text into an ACID candidate could also result in an incorrect extracted ACID. This error could occur when an ACID is spoken in an unusual position in the transmission, when the ACID was never spoken in the transmissions, or when a segmentation error results in incorrect concatenation of two separate radio transmissions and places ACID phrases in the middle of the recognized text. Finally, speaker error (i.e., saying the wrong words) could lead to a nearest match that selects the wrong ACID.

This component is still being improved in several ways: 1) improving parsing logic to identify and transform the correct word sequence corresponding to the ACID phrase, 2) incorporating dialogue modeling (i.e. the transmissions preceding and following the transmission being processed) to take advantage of conversational context, and 3) prioritizing the candidate ACID list and improving ranking of nearest matches.

E. Semantic Parse

The semantic extraction component is performed in parallel with and independent of the ACID extraction component. This component focuses on the extraction of ATC semantic concepts such as controller clearances, instructions, advisories, and pilot readbacks using a two-tiered, rules-based, domain-specific semantic parser. The first tier of the parser accepts text transcriptions from the speech recognizer and executes a shallow parse, assigning labels to words and phrases that have a semantic role in higher-level ATC concepts. For example, this tier labels individual words like "cleared" with the "Cleared" label. Longer word sequences are favored over shorter ones so that word sequences such as "clear to land" or "cleared to land" are chunked together under the "CTL" label instead of the "Cleared" label. Words with numeric meaning such as "one", "twelve", "thousand" etc. are given generic labels like "Digit", with the expectation that they are likely parts of a higher-level semantic concept but need to be combined with other labeled words and further disambiguated before being tagged as such.

The second tier of the parser operates on labels produced by the first tier and executes a rules-based syntactic parse to combine labeled semantic roles into high-level semantic concepts. The rules configuration of the second tier parser defines what labels can be combined (e.g., it is allowable to combine "Cleared", "ILS", "Runway", "25R", and "Approach"), the order that they can appear in to qualify for combination, the number of labels that can be combined (e.g., how many "Taxiway" labels can be associated with a "Taxi" label) and for certain words such as numeric sequences, the symbol patterns that are acceptable for a specific semantic concept type (e.g., "Climb" and "Flight Level" can only be associated with a sequence of three "Digit" labels that when combined fall within a specified numerical range. This component has been configured with rules to parse over 30 ATC clearance concepts.

The derived data produced by this component allows analysts to focus on their higher-level research goals without having to deal with the task of extracting meaning from text and handling irrelevant variations in controller and pilot phraseology. For example, for an analyst evaluating how many approach clearances were issued at a certain airport on a specific day and what types of approaches were issued, designing a search query for all the possible word combinations denoting an approach clearance would likely be a complex and frustrating task. But with the value-added data from this component, the analyst could query directly for the higher-level approach clearance concept, which would automatically group and retrieve instruction variations such as "expect", "cleared", "join", and "intercept", as well as their associated parameters, such as runway and approach type. Furthermore, to this analyst, an accuracy measure specifying how many approaches were correctly recognized, parsed, and retrieved is much more meaningful than the word accuracy achieved by the recognizer.

As a rules-based parser, this component is susceptible to error when faced with previously unseen vocabulary or concept combinations outside the defined configurations. Thus, this component is still evolving in several ways. First, MITRE engineers are adapting the parser to support multi-hypothesis parsing for when labels are in contention between different higher-level concepts. For example, the transcript, "jetblue forty seven twenty seven cleared to land" could be parsed as either a complete ACID, "JBU4727", and a clearance to land without a runway; or a shorter ACID, "JBU4720", and a clearance to land on runway seven. Second, engineers are researching the use of deep learning based semantic interpretation to improve adaptability and anticipate unseen data.

F. Track Fusion

Fusion with trajectory data is the final component in the audio processing pipeline. It fuses each voice transmission and its derived metadata with a corresponding flight in Threaded Track Flight Story (TTFS) using the ACID extracted from the voice transmission and the start and end times on the transmission. TTFS is an existing analytic suite that fuses a number of surveillance data sources, including Airport Surface Detection Equipment – Model X (ASDE-X), National Offload Program (NOP), En Route Automation Modernization (ERAM) / Traffic Flow Management System (TFMS), and most recently Automatic Dependent Surveillance-Broadcast (ADS-B), to provide complex flight trajectory data and other performance metrics for high-fidelity analysis [24]. This final component adds voice transmissions and derived artifacts as a new analytic layer within TTFS's extensive data store, enabling expanded analysis of ATC and pilot intent, including approach procedure clearances, the case study described in the subsequent section.

IV. Case Study: Approach Procedure Utilization

This section describes how information extracted from voice data processing can be used to determine the approach clearance issued to arrival flights. A basic assumption of this technique for determining approach procedure usage is that the approach *issued* by the controller via voice communications is the approach *flown* by the aircraft.

A. Input Data

As described in Section III, the voice data processing produces a speaker role, detected ACID, and semantic concepts spoken in the transmission. One of the semantic concepts that is parsed is the approach clearance, including the name/type of approach (e.g., visual, ILS, RNP) and the runway (e.g., 26L, spoken as "two six left"). Depending on the approach type, the controller may also provide an approach suffix, such as "yankee" or "zulu".

The approach clearance semantic tag does not differentiate between an approach clearance issued by a controller and the readback of the approach clearance spoken by the pilot. Because many (but not all) readbacks contain the same information as the clearance, the semantic parsing component may identify approach clearances for the same ACID in consecutive transmissions. There are several techniques for assigning a single approach clearance to a given flight; we have chosen to select approach clearances by applying two criteria to a transcription record: (1) it must be labeled as controller speech, and (2) it must contain an approach clearance semantic tag.

However, ATC generally does not mention the arrival airport in the transmission and the DALR channel information is not sufficient for mapping a control position to a specific airport without facility-specific knowledge and a process for handling combined positions. The fusion of the transcription record with a flight track in TTFS enables the arrival airport to be determined automatically. As described in Section III, the detected ACID and the transmission times are used to find a unique flight track, so that the transcription record can be associated with the track and all other information that the track record contains, such as the arrival airport. This fusion of voice and track data enables analysis by airport.

The analysis presented in the following sections is based on processing of calendar year 2017 voice data for seven TRACONs and the fusion of those voice data analysis results to track information. The seven TRACONs are:

- Atlanta TRACON (A80)
- Boston TRACON (A90)
- Chicago TRACON (C90)
- Minneapolis TRACON (M98)
- New York TRACON (N90)
- Northern California TRACON (NCT)
- Potomac Consolidated TRACON (PCT)
- Southern California TRACON (SCT)

Some of the analysis presented is based on a subset of the data processed.

B. Approach Clearance Detection Accuracy

Each step in the voice data processing pipeline can introduce errors that may result in a failure to recognize an approach clearance, an incorrect approach clearance, or a failure to identify the flight within clearance.

- Audio ingest – if audio is missing from the archive, then the approach clearances contained within that audio will not be detected.
- Segmentation and speaker role identification – if an approach clearance spoken by the controller is incorrectly classified as pilot speech, then that approach clearance will be missed, per the approach clearance criteria described in the previous subsection. This part of the process is roughly 85-90 percent accurate, depending on the specific classification being measured.
- Speech-to-text – if the speech recognition system produces a text transcription with substitution errors (where the spoken word is substituted for a word not spoken) or deletion errors (where a spoken word is not transcribed at all), and those errors are critical to the approach clearance, then the approach clearance will either be missed or will be incorrect.

- Semantic parse – if approach clearance phraseology significantly varies from standard, or if the parser is not appropriately configured, then the approach clearance will likely be missed.
- ACID detection – if the ACID detected is incorrect, then the fusion of voice data and track data will be incorrect, and the flight-specific approach clearance will be incorrect. This may cause the approach clearance for a flight to be missed. In general, ACID detection can range from 70 to 95 percent accurate, depending on the facility, type of call signs, and speaker role (i.e., the current system is more accurate on controller speech and on commercial aviation call signs).
- Track fusion – if the flight track isn't accurate enough to assign the true arrival airport/runway, a correctly-detected approach clearance may not be fused to the appropriate flight.

Several filters were used to eliminate approach clearance records that are likely incorrect. First, we ensured that the identified ACID for a record corresponds to a flight that arrived at an airport within the ATC facility. For this analysis, we also excluded a small portion of records (less than 1.5 percent) where the detected approach clearance is more than 30 minutes from the flight's landing time.

By comparing the number of known IFR arrivals at a given site to the number of approach clearances detected, we can estimate an upper bound of the number of flights with no associated approach clearance. Using voice data processed from 2017 from seven TRACONs, we found that upper bound to be around 20 percent for most TRACONs. We looked for trends within the missing flights and found that general aviation flights made up a bigger share (23 percent) of flights without an associated approach clearance than would be expected from their proportion of the total flights (13 percent).

C. Approach Utilization Analysis

This section presents several examples of approach procedure utilization analysis made possible by fusing voice data to track data and other aviation data sources.

1) Approach Type Analysis Trends

Using voice data analysis to determine the approach clearance issued for specific flights, we can break down the number of approaches flown into each airport by the type of approach. See Figure 3, which presents approach types flown at each airport in a calendar month of 2017.

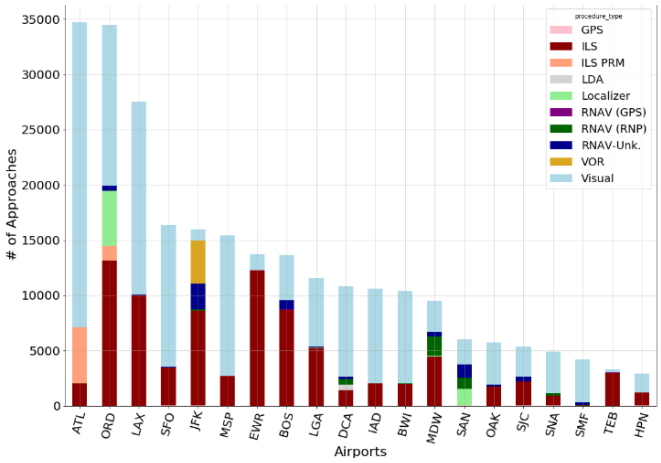

Figure 3. Approach Clearance Breakdown by Airport

This type of breakdown is not possible using only track data analysis. Voice communications are a NAS-wide data source that can be used to understand which procedures are used when. The same information may be collected using Flight Management System (FMS) data, but that would need to be collected from each airline individually.

The fused voice and track data can also be analyzed to understand trends in where flights are when they receive their approach clearances, to compare approach procedure clearance trends for differing aircraft equipage, and to understand how weather conditions affect approach procedure utilization at different facilities.

2) Comparison with RNP AR Conformance

As described in Section II.B, while there are some RNP AR procedures for which conformance can be accurately detected by conformance to the RF leg, there are some RNP procedures that may closely resemble visual approaches to the same runway; DCA Runway 19 and MDW Runway 22L are two such examples. Due to the similar lateral path between the visual and RNP AR approaches to these runways, RNP AR procedure usage counts based on track data conformance have always been caveated appropriately. We can use voice data to better understand the limitations of trajectory-based algorithms in distinguishing the visual and RNP AR approaches to the same runway in these challenging cases.

Figure 4 illustrates the similarities in lateral paths between the visual and RNP AR approaches to DCA Runway 19. The tracks are color-coded based on the approach clearance issued to each flight. For flights cleared for the LDA approach, the tracks are shown in grey.

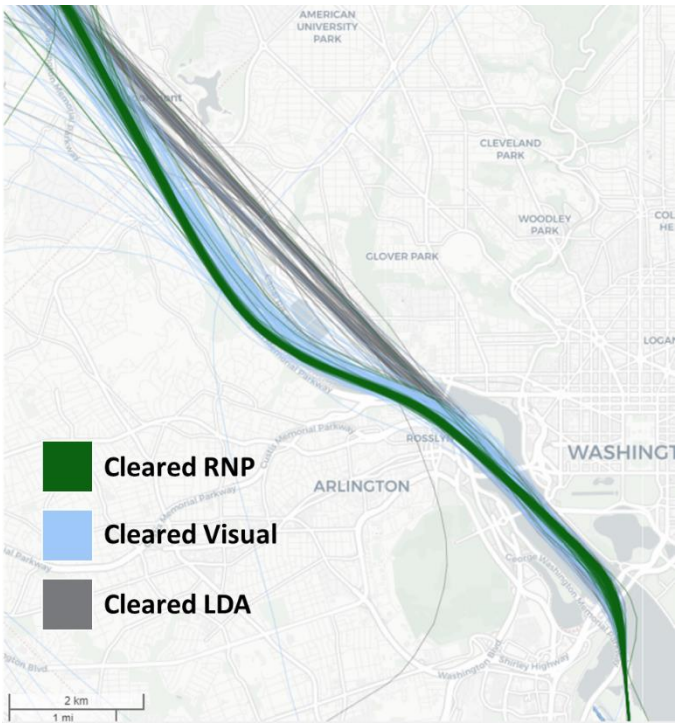

Figure 4. Track Similarity Between RNP and River Visual Approaches to DCA Runway 19

Figure 5 presents detected approach clearances for DCA Runway 19 during a calendar month in 2017, broken down by whether the flight conformed to the RNP within 0.12 Nautical Miles (NM) or not—shown in blue and red, respectively.

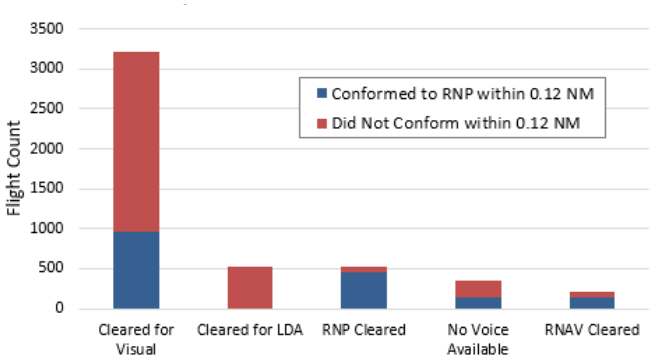

Figure 5. Approach Clearance and RNP Conformance Comparison for DCA Runway 19

Of more than 3,000 flights cleared for the visual approach, thirty percent of them (just under 1,000) conformed to the RNP. Only about 500 flights were cleared for the RNP. Thus, this analysis indicates that a conformance-based estimate of RNP utilization for DCA Runway 19 was roughly triple the number of RNP AR approaches actually flown.

The same analysis for MDW Runway 22L indicated a much smaller proportion of "false positives" in conformance-based RNP AR counts. Out of approximately 350 flights cleared for the visual approach to Runway 22L, approximately 150 conformed to the RNP, and the total number of flights that conformed to the RNP was approximately 1,300. Thus, the utilization estimate for RNP AR approaches to MDW Runway 22L appears to be accurate within 10-15 percent.

3) Identification of Circling Approaches

Circling approaches sometimes enable lower minima than straight-in approaches to the same runway for some terrain- or obstacle-constrained paths, enabling operations in poor weather conditions. They are also useful when instrument approaches are not available to the desired arrival runway. However, they are infrequently used across the NAS and add to the FAA maintenance footprint of the IAP inventory. The FAA thus would benefit from better insight into where and how often circling approaches are used to inform whether some can be eliminated.

Although the semantic parsing component was not configured to extract "circle" commands (e.g., "cleared ILS 31C circle 22L", which is a clearance for the arrival to land on Runway 22L), we were able to detect some circling approaches by examining clearances for approach types that were not published for the arrival runway. In the case of MDW, there is no ILS approach published for Runway 22L, but the semantic parser was returning approach clearances of type ILS and Runway 22L. On four particular days in a calendar month of the analysis, 30 percent of the 22L arrivals were given ILS clearances according to the semantic parser. Further investigation of the transcriptions revealed that these clearances were actually circling instructions: cleared ILS 31C circle 22L.

Figure 6 shows the track path of flights issued the circling approach (red) compared to those issued RNP (green) and RNAV (blue) approaches to 22L.

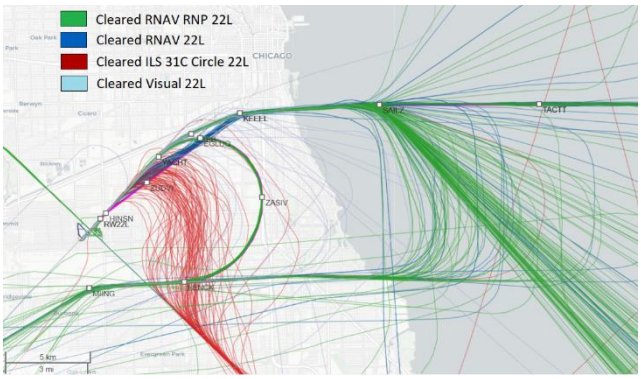

Figure 6. Circling (red), RNP (green), and RNAV (blue) Clearances to MDW Runway 22L

We have since added the detection of circling approaches to the semantic parser; expanding on the above analysis will enable an automated method for understanding how often circling approaches are used in the NAS.

D. Lessons Learned

Analysis of the approach clearance data revealed circling and special procedure clearances at some sites, and some phraseology variation in PBN approach clearances. Fusion to trajectory-based metrics within TTFS further allowed for comparison of results to approach procedure conformance algorithms, comparison of missed approach rates by approach type.

While speech recognition showed promise in our analysis of approach clearance detection, improvements to the capabilities will further enable meaningful analysis, especially when identifying outlier events. Detecting such off-nominal events revealed limitations of semantic parsing such as corrected clearances, where more than one approach type is issued in the same sentence by the controller. Such potential improvements identified from the approach analysis work will be incorporated as we mature our voice data analysis capabilities.

V. NEXT STEPS

The previous two sections have described large-scale voice data analysis capabilities and provided specific use cases that highlight how voice data can be leveraged for analyses that are not feasible with other data sources. While we are still improving the voice data analysis capabilities, these early analyses exemplify how voice data can provide new and valuable information to improve our understanding of air traffic operations.

A. Voice Data Analysis Capability Enhancements

While some ongoing enhancements to the voice data analysis capabilities are incremental improvements to existing features (e.g., modifying the semantic parser so that it can detect circling approach clearances), other enhancements are more transformative.

One significant planned enhancement is processing transmissions as a dialogue. The current set of capabilities treat each transmission in isolation, but in many cases, information from preceding or succeeding transmissions is needed to disambiguate or otherwise make sense of a given transmission. One example is ACID detection in cases when the ACID is not spoken. When a controller and pilot exchange multiple transmissions in a row about the same topic, they may drop the ACID from the intermediate transmissions. If that transmission is processed in isolation from all others, the automatic processing has no ability to determine which ACID should be associated with the transmission. By processing transmissions as a series, automatic capabilities can be set up to take advantage of this dialogue context. For example, a pilot readback of "two two zero two four zero" is ambiguous on its own, but it can be understood if the preceding controller instruction is known to be "united two four zero reduce speed to two two zero knots".

B. Other Voice Data Use Cases

Within the realm of PBN-related analyses, voice data can also be used to better understand arrival and departure procedure usage. Voice data can help assess procedure conformance—including why a flight may not conform to an arrival or departure procedure—and provide insights into the operational integration of capabilities supporting Trajectory-Based Operations (TBO).

Voice data also contain valuable information pertinent to safety analyses. In some cases, voice data can provide supplemental information about an event—e.g., whether the controller or pilot initiated a missed approach/go around. In other cases, voice communications may be the only data source that can provide context to an event—e.g., whether pilot-applied visual separation has been established.

Finally, voice data continues to be a valuable information source for reviewing individual events for a complete understanding of special cases. Even if speech recognition and understanding are not perfect, automatically-processed voice data make it easier to find the audio of interest among the thousands of channels recorded in the NAS throughout the day.

As voice data processing improves in accuracy, our ability to understand air traffic operations will correspondingly improve. Research questions previously seen as being unanswerable will become possible and allow for new, more complex questions to be posed and studied as the NAS transitions to TBO.

ACKNOWLEDGMENTS

The authors thank Marshall Koch and Katie Shepley for their leadership in defining the work and enabling a reusable capability to be developed. We also thank other critical members of the technical team for developing the core speech processing capabilities—Dr. Weiye Ma and Dr. Yuan-Jun Wei—and the infrastructure for large-scale data processing—Vick Fisher, Kevin Ray, and Tao Yu. Finally, we thank Donna Creasap and Mark Steinbicker at the FAA for sponsoring this work.

REFERENCES

[1] FAA. "Digital Audio Legal Recorder Operations and Maintenance Manual," FAA, Washington, DC, 6670-16-tib, October 21, 2008.

[2] FAA. "Digital Audio Legal Recorder (DALR) Remote Audio Access System (DRAAS) Concept of Operations," FAA, Washington, DC, Version 8, August 1, 2014.

[3] ASIAS. "Incorporation of voice data into ASIAS." The MITRE Corporation, Mclean, VA, June 2017

[4] Hamel, Cheryl J., David Kotick and Mark Layton. "Microcomputer system integration for air control training." Special Report SR89-01, Naval Training Systems Center, Orlando, FL, January 1989.

[5] Garcia-Romero, Daniel, David Snyder, Gregory Sell, Daniel Povey, and Alan McCree. "Speaker diarization using deep neural network embeddings." Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, March 2017.

[6] Amodei, Dario et al. "Deep speech 2: end-to-end speech recognition in english and mandarin." Proceedings of the 33rd International Conference on Machine Learning. New York, NY, June 2016.

[7] Yu, Dong and Li Deng. "Automatic speech recognition: a deep learning approach." Springer-Verlag London, 2015.

[8] Goldberg, Yoav. "A primer on neural network models for natural language processing." Journal of Artificial Intelligence Research, Vol. 57, Issue 1, pp 345-420. AI Access Foundation, September 2016.

[9] Chen, Yun-Nung, Asli Celikyilmaz, and Dilek Hakkani-Tur. "Deep learning for dialogue systems." Proceedings of Association for Computational Linguistics (ACL), Vancouver, Canada, July 2017

[10] Kopald, Hunter D., Ari Chanen, Shuo Chen, Elida C. Smith, and Robert M. Tarakan. "Applying automatic speech recognition technology to air traffic management." 32nd AIAA/IEEE Digital Avionics Systems Conference (DASC), East Syracuse, New York, October 2013.

[11] Cordero, Jose Manuel, Natalia R. Uncles, Jose M. de Pablo, and Manuel Dorado. "Automatic speech recognition in controller communications applied to workload measurement." 3rd SESAR Innovation Days, Stockholm, Sweden, November 2013.

[12] Chen, Shuo, Hunter D. Kopald, Adel Elessawy, Zach Levonian, Robert M. Tarakan. "Speech inputs to surface safety logic systems." 34th AIAA/IEEE Digital Avionics Systems Conference (DASC), Prague, Czech Republic, September 2015.

[13] Chen, Shuo, Hunter D. Kopald, Ron Chong, Yuan-Jun Wei, and Zach Levonian. "Read back error detection using automatic speech

recognition." 12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017), Seattle, Washington, June 2017

[14] Helmke, Hartmut and Oliver Ohneiser. "Increasing ATM efficiency with assistant based speech recognition." 12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017), Seattle, Washington, June 2017.

[15] Helmke, Hartmut, Petr Motlicek, Dietrich Klakow, Christian Kern, and Petr Hlousek. "Cost Reductions Enabled by Machine Learning in ATM." 13th USA/Europe Air Traffic Management Research and Development Seminar (ATM2019), Vienna, Austria, June 2019.

[16] Kleinert, Matthias et al. "Semi-supervised adaptation of assistant based speech recognition models for different approach areas." 37th AIAA/IEEE Digital Avionics Systems Conference (DASC). London, United Kingdom, September 2018.

[17] AIRBUS. "Airbus Artificial Intelligence Challenges." AIRBUS, [Online]. Available: http://AiGym.Airbus.com. [accessed 11 February 2019].

[18] Pellegrini, Thomas, Jerome Farinas, Estelle Delpech, and Francois Lancelot. "The airbus air traffic control speech recognition 2018 challenge: towards ATC automatic transcription and call sign detection." https://arxiv.org/abs/1810.12614, October 2018.

[19] Despres, Julien, Jodie Gauvain, Viet-Bac Le, Igor Swiecicki, Lori Lamel, and Jean-Luc Gauvain. "Airbus air traffic control speech recognition challenge." IRIT/SAMOVA, [Online]. Available: https://www.irit.fr/recherches/SAMOVA/assets/files/Seminaires/AIRBUS_ATC_Challenge_2018/Presentations/1-LIMSI-VOCAPIA.pdf. [Accessed 11 February 2019].

[20] Carnegie Mellon University. "CMUSphinx," Carnegie Mellon University, [Online]. Available: https://cmusphinx.github.io. [Accessed 8 January 2019].

[21] Povey, Dan, "Kaldi," Johns Hopkins University, [Online]. Available: http://www.kaldi-asr.org. [Accessed 31 July 2018].

[22] Povey, Dan, "Kaldi," Johns Hopkins University, [Online]. Available: http://kaldi-asr.org/doc/chain.html. [Accessed 8 January 2019].

[23] Peddinti, Vijayaditya, Daniel Povey and Sanjeev Khudanpur. "A time delay neural network architecture for efficient modeling of long temporal contexts." INTERSPEECH (2015).

[24] Eckstein, Adric C., Chris Kurcz, and Marcio O. Silva. "Threaded Track: geospatial data fusion for aircraft flight trajectories." MTR120423. The MITRE Corporation, McLean, Virginia, August 2012.

NOTICE

This work was produced for the U.S. Government under Contract DTFAWA-10-C-00080 and is subject to Federal Aviation Administration Acquisition Management System Clause 3.5-13, Rights In Data-General, Alt. III and Alt. IV (Oct. 1996).