aerospace

MDPI

Article

Toward Effective Aircraft Call Sign Detection Using Fuzzy String-Matching between ASR and ADS-B Data

Mohammed Saïd Kasttet [1,*], Abdelouahid Lyhyaoui [1], Douae Zbakh [1], Adil Aramja [1] and Abderazzek Kachkari [2]

[1] Laboratory of Innovative Technologies (LTI), National Schools of Applied Sciences of Tangier (ENSAT), Tangier 90063, Morocco; lyhyaoui@ensat.ac.ma (A.L.); douae.zbakh@gmail.com (D.Z.); adil.aramja@gmail.com (A.A.)

[2] The Moroccan Airports Authority (ONDA) Tangier-Ibn Battouta Intl. Airport, Tangier 90032, Morocco; kachkari@gmail.com

* Correspondence: mohammedsaid.kasttet@etu.uae.ac.ma

Abstract: Recently, artificial intelligence and data science have witnessed dramatic progress and rapid growth, especially Automatic Speech Recognition (ASR) technology based on Hidden Markov Models (HMMs) and Deep Neural Networks (DNNs). Consequently, new end-to-end Recurrent Neural Network (RNN) toolkits were developed with higher speed and accuracy that can often achieve a Word Error Rate (WER) below 10%. These toolkits can nowadays be deployed, for instance, within aircraft cockpits and Air Traffic Control (ATC) systems in order to identify aircraft and display recognized voice messages related to flight data, especially for airports not equipped with radar. Hence, the performance of air traffic controllers and pilots can ultimately be improved by reducing workload and stress and enforcing safety standards. Our experiment conducted at Tangier's International Airport ATC aimed to build an ASR model that is able to recognize aircraft call signs in a fast and accurate way. The acoustic and linguistic models were trained on the Ibn Battouta Speech Corpus (IBSC), resulting in an unprecedented speech dataset with approved transcription that includes real weather aerodrome observation data and flight information with a call sign captured by an ADS-B receiver. All of these data were synchronized with voice recordings in a structured format. We calculated the WER to evaluate the model's accuracy and compared different methods of dataset training for model building and adaptation. Despite the high interference in the VHF radio communication channel and fast-speaking conditions that increased the WER level to 20%, our standalone and low-cost ASR system with a trained RNN model, supported by the Deep Speech toolkit, was able to achieve call sign detection rate scores up to 96% in air traffic controller messages and 90% in pilot messages while displaying related flight information from ADS-B data using the Fuzzy string-matching algorithm.

Keywords: ATC; ASR; HMM; DNN; RNN; WER; VHF; ADS-B; METAR; GMTT; speech corpus; deep speech; call sign detection; levenshtein distance; fuzzy string matching

Citation: Kasttet, M.S.; Lyhyaoui, A.; Zbakh, D.; Aramja, A.; Kachkari, A. Toward Effective Aircraft Call Sign Detection Using Fuzzy String-Matching between ASR and ADS-B Data. *Aerospace* **2024**, *11*, 32. https://doi.org/10.3390/aerospace11010032

Academic Editors: Hartmut Helmke and Oliver Ohneiser

Received: 29 October 2023
Revised: 13 December 2023
Accepted: 15 December 2023
Published: 29 December 2023

1. Introduction

The purpose of Air Traffic Control (ATC) is to ensure the safe and efficient movement of aircraft within a specific controlled airspace. It helps prevent collisions between different aircraft and between aircraft and the surrounding obstacles, maintaining the order of air traffic and allowing quick support and collaboration in case an aircraft declares an emergency [1].

Air traffic controllers monitor the position of any aircraft assigned to their airspace and ensure aircraft separation and distancing using primary or secondary radars. The communication with pilots is ensured via Very High Frequency (VHF) radio equipment. Any change in the aircraft's heading or assigned flight level is subject to ATC approval, which ensures the appropriate horizontal and vertical separation between aircraft on the ground or in the controlled airspace is thoroughly respected.

L. Rabiner [2] defined ASR as "as a technology that involves the conversion of speech signals into a sequence of words by a computer program". Every ASR system should consider the type of speech recognizer, which can be speaker-dependent or speaker-independent. The first type requires prior training for each user to create voice patterns for hypothesis comparison. This kind of system is more accurate and has better performance. It can be designed for voice command solutions with limited vocabulary pronounced in the flight cockpit. Our application, dedicated to Air Traffic Control Officers (ATCOs), aims to recognize the pilot's spoken message and display flight data captured by the ADS-B receiver. It is a multi-user system or a speaker-independent recognizer, where the implementation is more complex considering the variety of accents and mispronunciations. It thus requires more hardware capabilities, such as memory and processor speed. In the given conditions, such systems could not achieve an accuracy lower than 10% word error rate (WER) [3].

In this paper, we introduce an ASR based on DNN, a new end-to-end RNN, and the Fuzzy string-matching algorithm to enhance ATC efficiency by reducing cognitive workload in dense traffic situations, especially in airports not equipped with radar. We use an Automatic Dependent Surveillance-Broadcast (ADS-B) receiver to provide captured flight data of all surrounding aircraft synchronized with recorded VHF voice communication. After training the ATC datasets and generating both acoustic and language models, the ASR system was able to recognize, with a reasonable WER, the spoken pilot message by matching it with the decoded call sign from ADS-B data. This threshold rate matching enables the call sign detection and display of flight-related information for ATCOs, such as speed, heading, altitude, distance, and bearing to the airport.

2. Related Work

D. Becks [4] briefly reviewed the state of the art of automatic speech recognition systems with types and modes of operation. Additionally, Georgescu [5] provides a comparison study between ASR performance and hardware requirements.

Recently, the FAA's (Federal Aviation Administration) final report on ASR methodologies [6] concluded that transformers have had a significant impact on audio and NLP fields, and their innovative architecture has been successfully integrated into various algorithms [7].

Since 1980, considerable progress has been made in ASR and applied to the ATC domain. A good description of the state-of-the-art ASR systems and their application for ATC was provided by Van Nhan Nguyen [8].

2.1. ASR in ATC

Van Nhan Nguyen [8] described three ASR systems. The Hidden Markov Model (HMM) approach has been the most widely used technique for the last two decades. It is a simple and efficient solution with automatic training, but its main weakness lies in discarding information about time dependencies. A hybrid approach was introduced to overcome this weakness of HMM. This approach combines an Artificial Neural Network and a HMM. A recognition accuracy rate of 94.2% was achieved by Wroniszewska [9] using the K-Nearest Neighbor (KNN) classifier and Genetic Algorithms (GAs). Finally, an interesting approach was proposed by Beritilli [10] using Dynamic Time Warping (DTW) and Vector Quantization-Weighted Hit Rate (VQWHR), which is a robust solution for noisy environments such as ATC.

Although the hybrid approach combines different algorithms and techniques, challenges in ASR systems still exist. To address issues such as poor signal quality from VHF communication, ambiguity in commands and instruction values, or the use of non-standard phraseology and mispronunciation in different accents (native and non-native speakers) [11], a new approach based on utilizing contextual information is introduced to improve the performance and accuracy of ASR in ATC as a post-processing approach based on syntactic, semantic and pragmatic analysis.

2.2. Contextual Knowledge in ATC

Syntactic and semantic analyses [12,13] consist of parsing the result of recognized words from ASR systems and eliminating invalid sentences or words by respecting grammatical rules highly inspired by ICAO standard phraseology. It helps correct misrecognized out-of-vocabulary words with similar ones from valid words of the ATC vocabulary. Semantic analysis is the process of testing the meaning of sentences. It can help resolve ambiguity and recognize words despite background noises [14].

2.3. Call Sign Detection (CSD)

The ability of an ASR system to detect accurate call signs in ATC communication is measured by the CSD rate. In 2018, in collaboration with IRIT (Institute for Research in Informatics of Toulouse) and Safety Data-CFH, Airbus organized a challenge for 22 teams for automatic speech recognition in ATC and call sign detection [15]. The Airbus dataset consisted of 40 h of manually transcribed voice communication with various accents and a high speech rate over noisy radio channels. The best result achieved was a 7.62% WER and 82.44% CSD rate, scored by the VOCAPIA-LMSI team. In 2020, the ATCO2 project [16] added an NLP module to extract the call sign from a recognized spoken utterance matched with surveillance data (ADS-B and radar) and improved the WER from 33% to 30%. The results showcased in [17,18] reported up to 60.4% relative improvement in call sign recognition by boosting call sign n-grams with the combination of ASR and NLP methods to use surveillance data. Finally, by leveraging surveillance information, Blatt, A et al. [14] significantly improved the accuracy of call-sign recognition in noisy air traffic control environments. The model showed a 20% improvement compared to existing methods. The study by Shetty et al. 2022 [19] focused on command extraction, including the recognition of call signs as part of the semantic meanings of ATCo utterances. Their study emphasized the importance of correctly interpreting various command components, showing that call sign recognition can be achieved within 20 ms after full call sign has been uttered, making it feasible for live data use. The research used gold transcriptions to achieve call sign recognition rates above 95% and error rates below 2.5%. With automatic transcriptions, they obtained recognition rates between 92 and 98% and error rates below 5% for most datasets. Finally, Garcia et al. 2023 [20] focus on how ASR can assist air traffic controllers (ATCos) and flight crews (FCs) in their communication. It describes a project under the SESAR2020 solution for ASR in call sign recognition, which was a collaboration between Enaire, Indra, CRIDA, and EML Speech Technology GmbH. The ASR highlights call signs on the ATCo screen to improve situational awareness and safety. The recognition rates for this system were around 84–87% for controllers and 49–67% for flight crews.

3. Automatic Speech Recognition Pipelines

3.1. Conventional Automatic Speech Recognition

Automatic Speech Recognition (ASR) is the assignment of transducing raw audio signals of spoken language into text transcriptions. It is based on statistical pattern-matching using a combination of acoustic and language models, which depends on the complexity of the application. This discussion covers the history of ASR models, from Gaussian Mixtures (GMMs) and Hidden Markov Models (HMMs) to attention-augmented DNNs. The ASR architecture is represented in Figure 1.

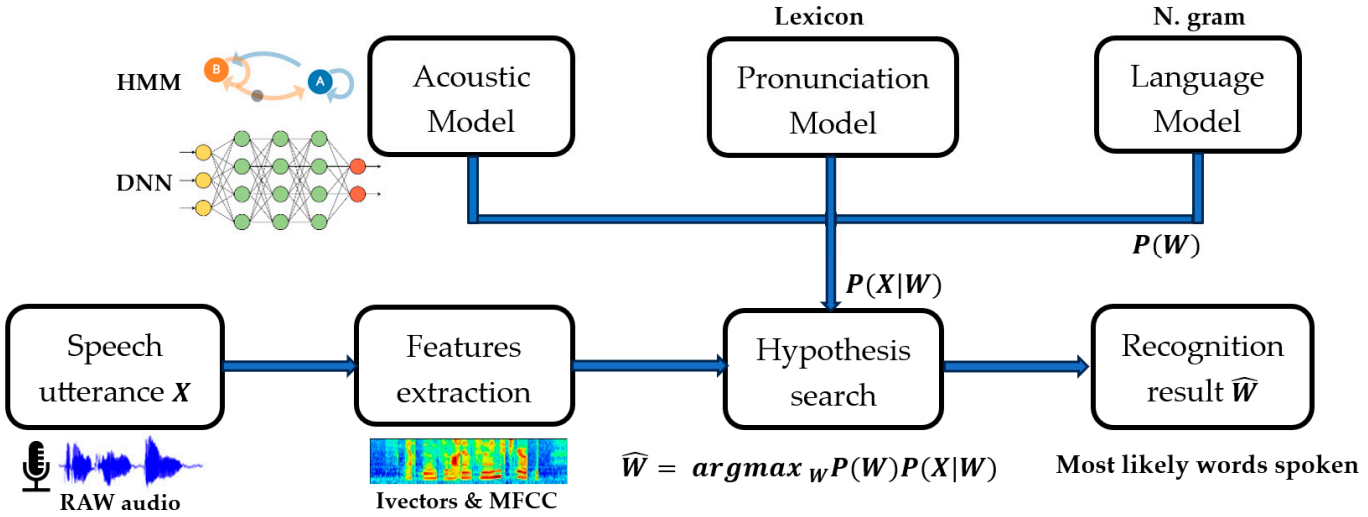

Figure 1. ASR architecture.

3.1.1. Acoustic Model

With reduced vocabulary, the acoustic model converts pronounced words into phonemes as minimal digital units. The speech processor compares the latter with stored word patterns until it matches the spoken utterance. However, in a complex situation, as in connected or continuous speech recognition, the analog voice signal is converted to digital format, typically using a 16 kHz sampling frequency. For feature extraction, the digital signal is transformed into the frequency domain using the Fast Fourier Transform (FFT). Subsequently, standard techniques [21], such as Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficients (MFCCs), are applied. The feature numbers are determined by comparing the resulting frequency graph with stored known sounds, which allows the referencing of each phoneme found.

However, in circumstances involving a speaker with a specific accent and the noisy environment of flight cockpits and radio communications, those feature numbers cannot identify a unique sound to become a particular phoneme. The solution is to use probability techniques such as Hidden Markov Models (HMMs) that represent each phoneme and use feature numbers' probabilities to calculate the transition state's likelihood (high probability).

Recently, many techniques [22] based on neural networks (NNs) have been deployed to replace the GMM and HMM by combining recurrent and convolutional neural networks to predict states efficiently [21].

3.1.2. Language Model

The English language contains 44 phonemes; every word is a sequence of phonemes with a large number of phonetic spelling possibilities. To overcome this problem, we generated a pronunciation dictionary of 907 unique words vocabulary from all ATC datasets, known as a lexicon. All probable words delivered by the acoustic model are compared in a second N-gram model [23] or an NN called a language model [24], which can predict the next word from a set of preceding words by following standard grammatical rules. Finally, a search engine combining all models can decode and continually recognize the most likely word sequence.

The aim of the speech recognizer engine is to find the most probable word $\hat{W}$ given an acoustic signal X as input.

$$\hat{W} = argmax_{\,W} P(W|X) \tag{1}$$

$P(W \mid X)$ is the probability that the word W was uttered, knowing that the evidence X was observed.

Equation (1) can be rewritten using Bayes' law, as shown in Equation (2):

$$P(W|X) = \frac{P(X|W) \cdot P(W)}{P(X)} \tag{2}$$

$P(W)$ is the probability that the word W will be uttered, $P(X \mid W)$ is the probability that the acoustic evidence X will be observed when the speaker speaks the word W, and $P(X)$ is the probability that X will be observed.

So, $P(X)$ can be ignored as $P(X)$ is not dependent on the selected word string. Consequently, Equation (1) can be written as Equation (3):

$$\hat{W} = argmax\ {}_W P(W)P(X|W) \tag{3}$$

where $P(W)$ is determined by the language model, and $P(X \mid W)$ is determined by the acoustic model.

3.2. End-to-End Speech Speech Recognition

For optimization purposes and simplification of the training process of different models, new end-to-end models are deployed for ASR [25]. It typically uses a type of neural network called deep neural network (DNN) or recurrent neural network (RNN) architecture. It is trained on large amounts of audio data with corresponding transcriptions. It has been proven effective at transcription in many cases, especially in a noisy environment, and can potentially simplify the ASR pipeline. The end-to-end model can directly decode a feature-extracted X from spoken utterance to a sequence of words Y+ by integrating the acoustic and the language model in one process, as shown in Figure 2; it is most often used in a reduced and noisy dataset. Moreover, there is a possibility of including an optional language model called a scorer to perform the best results.

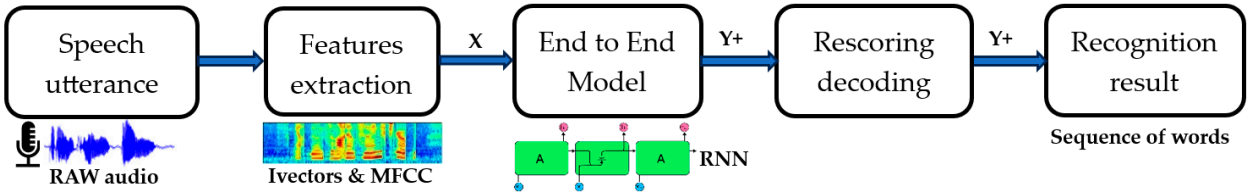

Figure 2. End-to-End ASR.

We notice that Connectionist Temporal Classification (CTC) [26] is the most popular training approach.

4. ATC Speech and Contextual Data Specification

4.1. ATC Communication

The standard communication, known as International Civil Aviation Organization (ICAO) Standard Phraseology, specifies all exchanged messages in Radio Telephony Communication (RTF) between air traffic controllers and pilots in controlled airspace, as well as in face-to-face communication between pilots and aerodrome staff in addition to the communication between pilots in the cockpit [27]. Primarily based on English or the national language, the pronunciation will be distinct between native and non-native English speakers. In some high-traffic situations, ATCOs must speak quickly to provide information and instructions for all aircraft in their allocated airspace [28]. Consequently, any recognizer system will return some broken or missing words due to the high speech rate and noisy radio signals from VHF transmission [29]. However, in some cases, it is possible to compensate for incorrect words by using the standard phraseology, as shown in Table 1, in addition to the structured contextual data such as a Meteorological Airport Report (METAR) and ADS-B flight information.

Table 1. Example of ICAO phraseology.

	Message			
	Tower	**Aircraft (Call Sign)**	**Information**	**Request Instruction**
Pilot	Tangier Approach	ARABIA six four niner	Descending flight level seven zero	Request visual approach runway 10
	Aircraft (call sign)	**Tower**	**Information**	**Instruction**
ATCo	ARABIA six four niner	Tangier Approach	Negative, last wind two seven zero degrees 25 knots	Report established for ILS approach runway 28

Table 1 shows the structured and precise nature of aviation communication exchange between the pilot and ATCO. Initially, the pilot, communicating with the Tangier Approach, identifies their aircraft as ARABIA six four niner and informs the tower that they are descending to flight level seventy FL70. The pilot then requests permission for a visual approach to runway 10. This request is part of standard aviation protocol, where pilots provide their current status and express their intended maneuvers. In response, the ATCO addresses the aircraft with the call sign ARABIA six four niner, indicating that the request is denied, possibly due to wind conditions, which are reported as two seven zero degrees at 25 knots. Instead of the requested visual approach, the ATCO instructs the pilot to prepare for an Instrument Landing System (ILS) approach for runway 28 and to report back once established on this approach. This exchange highlights the dynamic and responsive nature of air traffic communications, where ATCOs provide critical instructions and adjustments based on real-time conditions and operational requirements, ensuring the safety and efficiency of aircraft operations. The dialogue reflects the essential characteristics of air traffic communication: clarity, conciseness, and the conveyance of necessary information for the safe conduct of flights.

4.2. ADS-B Data and Call Sign

Automatic Dependent Surveillance Broadcast (ADS-B) is a technology for monitoring aircraft via satellite information. It improves the efficiency and safety of aircraft on the ground as well as in the air. It contains the flight call sign decoded in 3 letters and numbers for commercial flight, or equal to the aircraft registration number for private and general flights as shown in Table 2, speed, altitude, vertical speed, heading, and GPS latitude and longitude, as shown in Table 3. It is becoming the preferred method of real-time surveillance for ATC. Because of its reduced cost and valuable information on the call sign code, it is well suited for our application concept of ASR systems as the primary key for pilot message identification.

Table 2. Call Sign annotation.

Call Sign Annotation	**Designator**	**Transcription**
RAM982	RAM	royal air maroc niner eight two/air maroc niner height two
MAC146T	MAC	arabia maroc one four six tango/arabia one four six tango
CNTAV		charlie november tango alfa victor/charlie alfa victor

In Table 2, the provided call sign annotation data showcase the intricate and standardized method of communication in air traffic control, particularly in articulating aircraft call signs. For instance, the call sign RAM982 is designated as “RAM”, and its transcription unfolds as “Royal Air Maroc Niner Eight Two”. This transcription method, where numbers are spoken phonetically, is crucial for clarity, particularly in initial radio communication,

where precision is paramount. Later, the call sign transcription can be reduced for more straightforward pronunciation. Similarly, MAC146T, designated "MAC", is transcribed as "Arabia Maroc One Four Six Tango". Each number and the letter 'T' (Tango) are pronounced individually, denoting a specific flight or route. The third example, CNTAV, despite lacking a clear designator, is transcribed using the phonetic alphabet as "Charlie November Tango Alpha Victor". Each letter is articulated using a corresponding word from the phonetic alphabet, ensuring each character is unmistakably understood in potentially noisy or disrupted communication environments. These examples highlight the critical importance of standardized and clear communication in aviation, especially in identifying aircraft, where even minor miscommunications can have significant implications for air traffic safety and efficiency.

Table 3. Example of ADS-B data.

Date	Time	Call Sign	Radar	Alt	Speed	Head	Vertical	Lat	Lon
6 July 2021	08:42:51	RAM982	5320	9000 ft	580 kt	320°	80 ft/min	35.46	−7.48

4.3. METAR

METAR is a weather observation report for an aerodrome and is periodically generated every 30 min. It contains wind direction and speed data, temperature, dew point, cloud cover and heights, visibility, and barometric pressure. Aircraft pilots and controllers primarily use it to determine runway-in-use and flight rules during takeoffs or landing operations.

Table 4: Example of METAR report shows an example of a weather report of Tangier Aerodrome made on 10/06/2021 at 10:30 UTC. The conditions were 15 kt wind from the west with gusts up to 30 kt, temperature of 14 °C, 84% humidity, a pressure of 1012 hPa, visibility of 7000 m, and few clouds at a height of 3000 ft. No significant changes occurred in the next two hours.

Table 4. Example of METAR report.

Aero-Drome	Day/Time	Wind Direction/Speed	Visibility	Clouds	Temp/Dew	Pressure	
GMTT	101,330 Z	27015G30KT	7000	FEW020	14/12	Q1012	NOSIG

5. Methodology and Materials

Our methodology using ASR in the specific domain of ATC involves several steps; the process begins with dataset collection for training and recording new actual speech corpus IBSC for testing; this includes communication between pilots and ATCOs, such as those with ground control and tower control, under different conditions, including varying levels of clarity, background noise, and accents. Once collected, the audio data need to be preprocessed. This stage involves cleaning the audio by reducing noise, normalizing audio levels, and segmenting it into smaller, manageable parts for easier processing. The next step is the accurate transcription of these audio files. This process is crucial and should include not only verbal communication but also annotations for non-verbal elements like flight and metrological information from ADS-B and METAR data, which is especially important in the context of ATC communications. The core of our research will involve training our chosen ASR model using the annotated data with two different toolkits based on DNN and RNN architecture. This process might require substantial computational resources and time. It is crucial to regularly validate and test the model with a separate dataset to ensure its accuracy. Special attention will be paid to how the model performs under various challenging conditions, like heavy accents, rapid speech, and noisy environments. In terms of evaluation and based on the results of our tests and validations, the model may need to be refined; this could involve adapting it with ATC data, tweaking the model parameters,

or experimenting with different sets of features. Finally, the goal in the context of ATC is to achieve the highest possible accuracy and reliability, particularly under challenging conditions, due to the critical nature of ATC communications.

For call sign detection, we will implement fuzzy string matching, which is particularly important in fields like automatic speech recognition using the Levenshtein algorithm. This method centers on calculating the number of edits-insertions, deletions, or substitutions needed to transform one string into another. This algorithm is readily available in many programming languages; Python offers libraries like Fuzzy Wuzzy for this purpose [30]. We set a matching threshold based on our accuracy needs—a lower threshold means more lenient matching, while a higher one requires a closer match. We applied the algorithm to our dataset, compared each string to our target string ASR hypothesis, and calculated the similarity score. The results were evaluated and adjusted to finetune the threshold parameter of the algorithm as necessary.

5.1. Data Collection: The Ibn Battouta Speech Corpus

The Ibn Battouta Speech Corpus is a synchronized dataset of voice communication between pilots and ATCOs with weather observation data originated from Tangier's airport and current activated aircraft flight information [31], which has a very rich pronunciation accents of native and nonnative speakers thanks to its vital geographic position linking different airspaces from Morocco (GMMM, GMTT), Spain (LEZL), and Gibraltar (LXGB). The purpose is to detect and record audio speech with various accents and related captured ADS-B data plus METAR report provided by the NWS Server [32], as described in Figure 3.

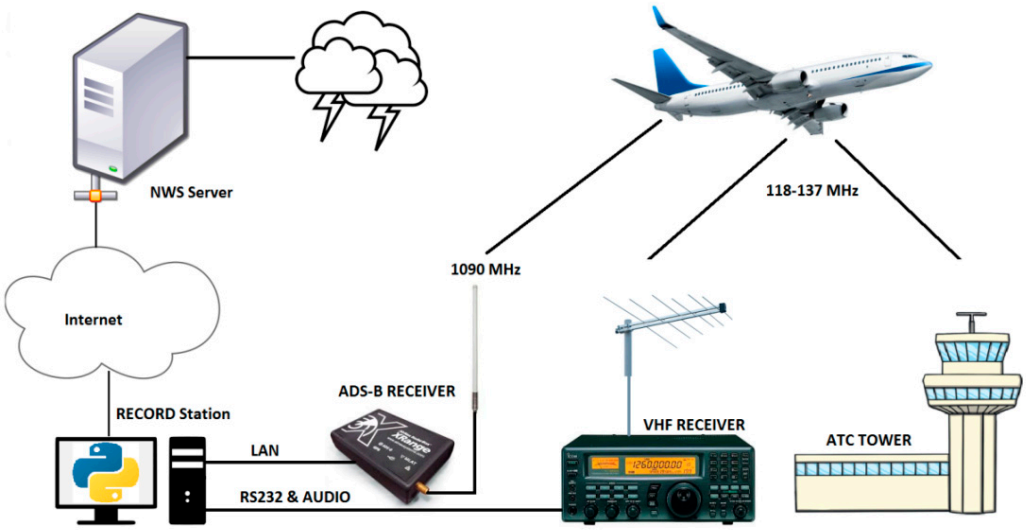

Figure 3. Ibn Battouta dataset architecture.

The voice recording was obtained with Voice Activity Detection (VAD) [33] at a rate of 16 kHz from the VHF receiver [34] tuned to the airport frequency connected to the workstation's audio input. At the same time, the ADS-B receiver provided by AirNav System [35] logs flight data, as shown in Figure 4, including the call sign code of activated aircraft with approximately 200 Nm circumference. In addition, a weather report is saved separately after downloading updated data from the US National Weather Service (NWS) Server.

Tracked	Status	Mode S	Flight ID	Registration	Aircraft	Airline	Route	Altitude	GS	Hdg	VRate	Squawk	Company
09:28:12	Cruise	3004C2	NOS1446	I-NEOZ	B738	neos	GCFV-LIPE	35 000	450	240	0	4057	Neos
09:28:12	Cruise	4CA806	RYR1JX	EI-EKH	B738	RYANAIR	GMMX-LEGE	38 000	450	060	0	6461	Ryanair
09:28:12	Departure	020118	MAC377Z	CN-NMI	A320	airarabia.com		16 275	360	040	1720	6475	Air Arabia Maroc
10:20:27	Leveled	4CADF7		EI-IHM	B38M			12 600			0		Ryanair
10:23:09	Cruise	502D5E	BTI6WU	YL-ABM	BCS3	airBaltic	EYVI-GCTS	37 000	450	230	0	1174	Air Baltic
10:19:39	Cruise	346689	IBB18XQ	EC-NPU	E295	BinterCanarias		36 000	430	040	0	6212	Binter Canarias
	Landing	347302	BCS932	EC-NXU	B738	DHL	LEMD-GMTT				0	6774	Swiftair
10:20:25	Cruise	4B027D	EZS73RW	HB-AYN	A20N	easyJet	GMMX-LFSB	38 025	430	360	0	6463	easyJet Switzerland
	Timeout	39D311	TVF68DK	F-HUYR	B738	transavia.com		37 000	360	040	0	6452	Transavia France

Figure 4. ADS−B flight data.

5.2. Transcription and Logging

All utterances were transcribed and manually annotated by real pilots and ATCOs who authored this paper [36]. It is a time-intensive task that requires ten man-hours to transcribe one hour of speech. In order to estimate the distance and radial information, which are frequently requested by ATCOs from pilots, we integrated a Python 3 code-based program into the flight data from the ADS-B receiver using the Haversine formula [37] to calculate and save the distance and bearing between the aircraft GPS position and D-VOR installed on the Tangier airport runway. The call sign, date, and time are tagged in transcription files to enhance context-free grammar. Table 5 summarizes the dataset characteristics.

Table 5. The Ibn Battouta dataset characteristics.

	Speaker		Gender		Total
	Pilot	**ATCO**	**Female**	**Male**	
Number of utterances	992	1500	544	1948	2492
Duration (sec)	5416	10,040	3720	11,736	15,456
Number of words	12,936	22,224	8084	27,076	35,160
Signal Average (dB)	106	95	90	102	101
Aircraft Call Sign	832	1180	440	1572	2012

The Ibn Battouta dataset is a rich and complex collection of communications in the air traffic control context, encapsulating a wide array of spoken interactions between pilots and air traffic control officers (ATCOs). It comprises a total of 2492 utterances, divided between 992 from pilots and 1500 from ATCOs, indicating the more extensive communicative role of ATCOs in managing airspace. Notably, the dataset reveals a gender imbalance in communication, with female speakers contributing 544 utterances against 1948 from male speakers, highlighting the male predominance in this sector. The total duration of these communications is 15,456 s (4 h 20 min), with pilots accounting for 5416 s and ATCOs for a larger share of 10,040 s, reflecting the extensive and detailed nature of ATCO communications. Regarding word usage, the dataset records 35,160 words, with a significant portion (22,224 words) used by ATCOs, further emphasizing the complexity of their verbal exchanges. The signal strength, measured in decibels, averages 101 dB across the dataset, with a higher average for pilots (106 dB) compared to ATCOs (95 dB), possibly due to different communication environments or equipment. The dataset also includes a diverse array of 2012 aircraft call signs, with pilots using 832 and ATCOs 1180, adding to the complexity of speech recognition challenges in this domain. Overall, the Ibn Battouta dataset offers invaluable insights into linguistic characteristics and communicative dynamics in air traffic control.

5.3. Datasets of Training and Adapting Models

We collected the following available ATC datasets with different accents and environments (real operational and laboratory simulation) for model training phases, as summarized in Table 6.

Table 6. Dataset splitting for model training.

	Data Set	Accent	Environment	Utterance	Duration	Call Sign Annotation
Training (103 h) 46,732 utterances	LDC94S14A [38]	USA	Operational	25,120	60 h	No
	ZCU_CZ [39]	Czech	Operational	6435	15 h	No
	ATCOSIM [40]	FR/DE/CH	Simulation	8078	8 h	No
	HIWARE [41]	FR/GK/ES/IT	Simulation	7099	25 h	No
Validation (21 h) 10,024 utterance	Mixed/Unseen	Mixed	Mixed	10,024	21 h	No
Test (11 h) 5382 utterances	ATCO2 [42]	CZ/DE/CH/AU	Operational	2890	6 h	2817
	IBSC	MAR/ES/FR/EN	Operational	2492	5 h	2012

This dataset is specifically tailored for research in automatic speech recognition within the air traffic control sector, comprising a diverse range of accents, environments, and operational scenarios. It is segmented into three primary sections: training, validation, and testing, cumulatively spanning 135 h. The training set, with a substantial 103 h of audio, incorporates a wide array of utterances from the USA, Czech Republic, and a mix of countries like France, Germany, Switzerland, Greece, Spain, and Italy, covering both operational and simulation environments. The validation set offers a 21 h mixed compilation from unseen sources in varied environments. Lastly, the testing segment, totaling 11 h, includes specific datasets like ATCO2 and IBSC, representing a range of Morocco and Spain airspace in operational settings like En route and approach flight situations. This section is unique as it includes call sign annotation.

5.4. Vocabulary and Accuracy

A limited or medium ATC vocabulary size, estimated at around 500 words, and the standard phraseology of ATC grammar with its substantial semantic restrictions both allow better accuracy by increasing the probabilities of valid words and their sequences despite the noisy environment and high speech rate.

The word error rate (WER) is the standard metric for measuring the accuracy of any ASR system [43,44]. It is calculated using the formula given in Equation (4):

$$WER = \frac{I + D + S}{N} \tag{4}$$

where I is the number of insertions, D is the number of deletions, S is the number of substitutions, and N is the number of words in the sentence.

The Real-Time Factor (RTF) is included to measure the speed of ASR. It can be computed using the ratio expressed in Equation (5):

$$RTF = {}^{P}/_{I} \tag{5}$$

where P is the necessary time to process an input of duration I.

5.5. Fuzzy String-Matching

For call sign detection, we applied a string-matching algorithm called the Fuzzy, which determines the closeness of two strings. It is a technique used to identify two elements of text strings that match partially but not precisely. This algorithm is based on the

Levenshtein distance [45], a metric that evaluate the dissimilarity between two sequences of words. This measure calculates the least number of modifications required to transform one sequence of words into another.

Mathematically, the Levenshtein distance between two strings a, b is given by $lev_{a,b}(|a|, |b|)$ in Equation (6), where:

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) \; if \; \min(i,j) = 0, \\ min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} otherwise \end{cases} \quad (6)$$

where $1_{(a_i \neq b_j)}$ is the indicator function equal to 0 when $a_i = b_j$, and equal to 1 otherwise.

A threshold value will be determined during the experimentation to assess the similarity ratio of the string matching between ADS-B Call Signs and Hypothesis transcription.

6. Experimentation

6.1. Overview

In our experiment employing the IBSC, we searched for the call sign in recognized messages in all ADS-B line data stored and synchronized with voice utterances. The call sign code in ADS-B data was parsed using an airline call sign designator database [46] from the International Air Transport Association (IATA) and phonetic transcription. The highest score of string matching allows the appropriate call sign to be identified.

We used the Pocketsphinx toolkit [47] with HMM-DNN topologies and the Deep Speech recognition toolkit [48] based on a Recurrent Neural Network (RNN) for training, adapting, and testing the IBSC dataset on an HP Z4 workstation equipped with an Nvidia GeForce RTX 2070 GPU for training acceleration, and Ubuntu 18.04 as the OS.

Precisely, we implemented the Mozilla Deep speech version 0.9.3; the RNN is fully connected and has bidirectional layers with 512 hidden units per layer. Initially, it contains three layers with clipped rectified-linear (ReLU) activation, a Long Short-Term Memory (LSTM) layer, followed by another layer with ReLU activation. Lastly, it is capped by a softmax classifier to predict the most likely alphabet letter at each point in an audio utterance.

6.2. Experimental Setup

First, we trained [49,50] two new acoustic models with five hidden layers using the Pocketsphinx toolkit and the Deep speech toolkit with TensorFlow. We then adapted [51,52] each toolkit's default English model with all ATC datasets, as indicated in Table 6. For audio data representation, we computed spectrograms of 80 linearly spaced log filter banks and an energy term. The filter banks were computed over 20 ms windows with strides of 10 ms each. The language model was a 3 g model with a 907 unique words vocabulary from all ATC datasets, which contained 55,338 utterances with a total duration of 128 h, and from the AirNav Systems database history, from which the call sign was extracted and decoded. To enhance and train this language model using the SRLIM toolkit [53], we added all air waypoints from Morocco, Spain, and Gibraltar's nearby airspaces and decoded meteorological reports, in addition to all existing commercial and private company call sign designator [54]. Finally, it took about 32 h to train each new acoustic model for 200 epochs.

For call sign detection, we used a pragmatic analysis based on ADS-B data, including flight information, to detect the call sign in a recognized pilot message that represents essential information for ATCOs to identify the aircraft; we implemented the fuzzy Wuzzy Python function using the token_set_ratio () method [55]. It returned the highest similarity ratio score for fuzzy string matching in all ADS-B data lines stored in the dataset for each recognized utterance.

6.3. Results and Discussion

After training, adapting, and following the assumptions given above, our ASR model was tested on a different corpus from Morocco, Spain, and Gibraltar airspaces to cover different accents, airspace information, and role speakers, as shown in Figure 5.

In an analysis of two prominent ASR models, PocketSphinx and Deep Speech, applied to a dataset of air traffic control communications, distinct trends emerge in their performance across various metrics. Both models were evaluated under three distinct training conditions: pretrained English, trained on ATC only, and adapted pretrained English + ATC on three key metrics:

- word error rate (WER) as defined in (4).
- fuzzy string-matching ratio given by the percentage of similarity between two strings.
- and call sign detection rate which gives the percentage of correct call signs detected among total transcribed utterances.

The baseline results of model training and adapting ATC data sets are shown in Table 7.

Table 7. Model training and adaptation.

Model		Word Error Rate			Fuzzy String-Matching Ratio			Call Sign Detection Rate		
		ATCO	PILOT	Both	ATCO	PILOT	Both	ATCO	PILOT	Both
Pocket sphinx HMM-DNN	Pretrained English	83%	91%	87%	17%	09%	13%	0%	0%	0%
	Trained on ATC only	14%	20%	17%	68%	60%	64%	94%	84%	89%
	Adapted Pretrained English + ATC	11%	13%	12%	78%	68%	73%	95%	87%	91%
Deep Speech RNN	Pretrained English	81%	89%	85%	27%	19%	23%	0%	0%	0%
	Trained on ATC only	10%	12%	11%	80%	66%	73%	93%	89%	91%
	Adapted Pretrained English + ATC	08%	10%	09%	85%	77%	81%	96%	90%	93%

For PocketSphinx, the pretrained English model trained on approximately 6500 h of data not related to the ATC condition showed in Table 7 high WERs (83% for ATCO, 91% for the pilot, and 87% overall) and low fuzzy string-matching ratios (17% for ATCO, 9% for the pilot, and 13% overall), along with a 0% call sign detection rate across all categories. However, when trained exclusively on ATC data, there was a substantial improvement in all metrics, with the call sign detection rate reaching as high as 94% for ATCO, 84% for the pilot, and 89% overall. The adaptation of pretrained English with ATC data further enhanced performance, reducing WERs to 11–13% and increasing the fuzzy string-matching ratio to around 70–78%.

Deep speech mirrored these trends but with consistently better outcomes. Under the pretrained English condition, it had slightly lower WERs and higher string-matching ratios than PocketSphinx but still had no call sign detection. Training on ATC data alone introduced significant enhancements, especially in call sign detection, reaching up to 93%. The adaptation of pretrained models with ATC data yielded the best results, with WERs dropping to as low as 8–10%, fuzzy string-matching ratios climbing to 81–85%, and call sign detection rates peaking at 93–96%.

Overall, these results clearly demonstrate that both PocketSphinx and Deep Speech significantly improve accuracy and reliability when trained on ATC-specific data, with Deep Speech showing slightly superior performance in all tested scenarios.

The results show that, in general, for low-resource data, adapting the pretrained default English model offers better performance [56] than training a new model, and using the RNN Deep Speech toolkit achieved better results in noisy environments, especially in-flight cockpit pilot transmission. Because the ATC dataset has a short duration, the

model does not need to increase the depth parameter, as it may lead to overfitting. For call sign detection, the similarity ratio of the fuzzy string matching was improved when the WER was low. This means the better the message recognition accuracy, the better the fuzzy string-matching score between the decoding call sign in ADS-B data and the recognized message.

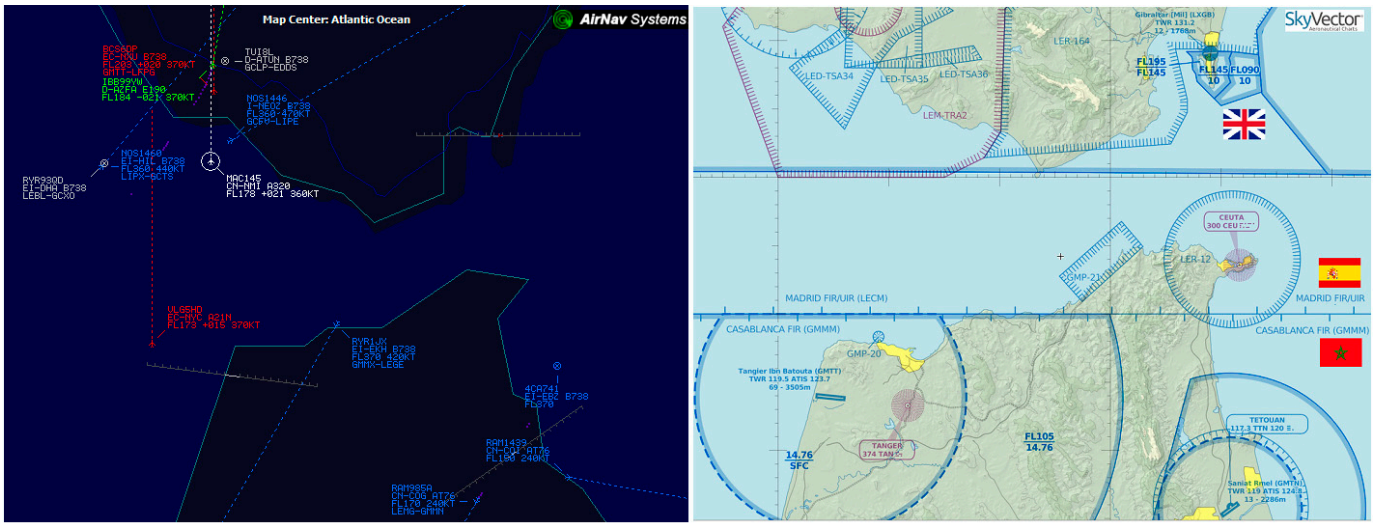

Figure 5. Test flight information region.

Table 8 details an example of a voice message by a Royal Air Maroc company pilot during the approach. The ASR hypothesis was confirmed with a WER of 12.5% by deleting the unknown word "um".

Table 8. ASR hypothesis example.

File Name	12120_20200319_170603_170606.trs
Tracking Date Time	19 March 2020 17:06:03
Real transcription	Tangier AIR MAROC zero seven four roger um continue approach
ASR hypothesis	Tangier AIR MAROC zero seven four roger continue approach
Word Error Rate WER	12.5%

By employing the data captured by the ADS-B receiver, we can assess the similarity between each detected call sign shown in Table 9; compared to the result of the ASR hypothesis from Table 8, the fuzzy string-matching score was calculated for each candidate. The model returned the call sign leading to the highest score, i.e., 89%. It corresponds to the RAM074 flight phonetic transcription (Tangier AIR MAROC zero seven four roger continue approach). A threshold of 80% is fixed to avoid the situation when all call signs have the same designator or do not concern the ASR hypothesis.

Table 9. Call sign candidates from ADS-B data.

Call Sign	Phonetic Transcription	Score	Type	Altitude	Speed	DME	Radial
BEL271	Beeline two seven one	45%	A333	35,000	470	506	225°
RAM075	Royal Air Maroc zero seven five	79%	B738	41,000	430	780	310°
BAW669	Speed bird six six niner	34%	A21N	36,000	460	380	198°
RAM074	Royal Air Maroc zero seven four	89%	B738	3325	210	20	98°
RYR8073	Ryanair eight zero seven tree	51%	B738	30,375	390	240	254°

6.4. Limitations

For private and general flights, the aircraft registration number is used as call sign instead of flight number in ADS-B data. We can apply the same Fuzzy algorithm based to search and match the phonetic transcription of aircraft registration number with the ASR hypothesis, since it is mandatory to exist in every ADS-B information and pronounced in every standard communication.

For light aircraft not equipped with ADS-B transmitters e.g., in VFR flight, we can only rely on ASR performance to detect the registration number based on phonetic transcription.

7. Conclusions and Further Work

Although the application of ASR in ATC is more challenging due to the high-security level required for air traffic management in the aviation domain, it remains possible to benefit from standard communication, a small vocabulary, and contextual information to implement simple and low-cost ASR solutions using ADS-B data to minimize the workload of ATCOs in high-traffic situations located in an airport not equipped with radar.

In our experiment, after training and adapting the ATC dataset using the Deep Speech toolkit and building the acoustic and language models based on a vocabulary dataset, we were able to demonstrate the successful detection of multiple aircraft call signs in recognized voice messages at a string-matching similarity rate starting from 60%. For safety obligations, we recommend a threshold of 80% for the fuzzy string-matching rate.

Further work in the ATC domain can present us a chance to try Whisper, the new advanced ASR system developed by OpenAI trained on 680,000 h of multilingual and multitask supervised data collected from the web, known for its high accuracy in transcribing speech, even in challenging conditions such as noisy environments or with speakers having different accents. It supports multiple languages, making it versatile for global applications. Whisper is designed to understand the conversation's context, which helps provide more accurate transcriptions.

An NLP module investigates the string position between the call sign and airport entity name; in addition, grammar rules such as gerund and key verbs like "request" and "report" in a recognized transmission would allow the detection of the speaker's role during standard ATCO and pilot communication.

Airport meteorological information and the runway are usually delivered to the pilot before takeoff and landing. Decoding the METAR report, extracting the wind direction, and calculating the runway in use will help confirm the acknowledgment between the pilot and ATCO communication.

Author Contributions: Conceptualization; M.S.K.; Methodology; A.L.; Software; A.A.; Validation; D.Z.; Formal Analysis; A.L.; Investigation; A.L.; Resources; M.S.K.; Data Curation; A.K.; Writing—Original Draft Preparation; M.S.K. and A.L.; Writing—Review & Editing: A.L., A.A., A.K. and D.Z.; Visualization; Supervision; Project Administration; Funding Acquisition: M.S.K. and A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by ULTRA CONTROL Technical Center and DETROIT TELECOM Company under agreement number DTUC-1923-ACRT.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from the European Language Resources Association (ELRA) and Aero Club Royal of Tangier; data are available under request.

Acknowledgments: The authors are grateful for the cooperation of pilots flying in Aero Club Royal of Tangier and air traffic controllers working at Tangier's Airport for their valuable help and guidance. The authors thank Khalid Choukri for providing the Datasets resources and Captain Nourddine Mountassir for his expertise and assistance in flight procedures. We would also like to thank the research team members in the Laboratory of Innovative Technologies (LTI) based at the National School of Applied Sciences of Tangier.

Conflicts of Interest: The authors declare that this study received funding from ULTRA CONTROL Technical Center and DETROIT TELECOM Company under agreement number DTUC-1923-ACRT. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Emergency Response Guidance for Aircraft Incidents Involving Dangerous Goods 2023–2024 (Doc 9481). Available online: https://store.icao.int/en/emergency-response-guidance-for-aircraft-incidents-involving-dangerous-goods-doc-9481 (accessed on 29 November 2023).
2. Rabiner, L.R.; Juang, B.H. *Fundamentals of Speech Recognition*; PTR Prentice Hall: Upper Saddle River, NJ, USA, 1993; ISBN 978-0-13-015157-5.
3. 3Play Media Study Finds Artificial Intelligence Innovation Has Led to Significant Improvements in Automatic Speech Recognition (ASR). Available online: https://www.businesswire.com/news/home/20230503005160/en/3Play-Media-Study-Finds-Artificial-Intelligence-Innovation-Has-Led-to-Significant-Improvements-in-Automatic-Speech-Recognition-ASR (accessed on 30 November 2023).
4. Beeks, D.W. Speech Recognition and Synthesis. In *Digital Avionics Handbook*; CRC Press: Boca Raton, FL, USA, 2015; ISBN 978-1-315-21698-0.
5. Georgescu, A.-L.; Pappalardo, A.; Cucu, H.; Blott, M. Performance vs. Hardware Requirements in State-of-the-Art Automatic Speech Recognition. *EURASIP J. Audio Speech Music Process.* **2021**, *2021*, 28. [CrossRef]
6. Achour, G.; Salunke, O.; Payan, A.P.; Harrison, E.; Sahbani, C.; Carannante, G.; Ditzler, G.; Bouaynaya, N.; Georgia Institute of Technology; Aerospace Systems Design Laboratory; et al. *Review of Automatic Speech Recognition Methodologies*; Federal Aviation Administration; William J. Hughes Technical Center: Egg Harbor Township, NJ, USA, 2023.
7. Wang, Y.; Mohamed, A.; Le, D.; Liu, C.; Xiao, A.; Mahadeokar, J.; Huang, H.; Tjandra, A.; Zhang, X.; Zhang, F.; et al. Transformer-Based Acoustic Modeling for Hybrid Speech Recognition. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; IEEE: Barcelona, Spain, 2020; pp. 6874–6878.
8. Nguyen, V.N.; Holone, H. Possibilities, Challenges and the State of the Art of Automatic Speech Recognition in Air Traffic Control. *World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control. Inf. Eng.* **2015**, *9*, 10.
9. Wroniszewska, M.; Dziedzic, J. Voice command recognition using hybrid genetic algorithm. *TASK Q.* **2010**, *14*, 377–396.
10. Beritelli, F.; Serrano, S. A Robust Low-Complexity Algorithm for Voice Command Recognition in Adverse Acoustic Environments. In Proceedings of the 2006 8th International Conference on Signal Processing, Guilin, China, 16–20 November 2006; IEEE: Guilin, China, 2006; p. 4129154.
11. Jahchan, N.; Barbier, F.; Gita, A.D.; Khelif, K.; Delpech, E. Towards an Accent-Robust Approach for ATC Communications Transcription. In Proceedings of the Interspeech 2021, ISCA, Brno, Czech Republic, 30 August–3 September 2021; pp. 3281–3285.
12. Joakim, K. *The Integration of Automatic Speech Recognition into the Air Traffic Control System*; Flight Transportation Laboratory, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology: Cambridge, MA, USA, 1990.
13. Schmidt, A.; Oualil, Y.; Ohneiser, O.; Kleinert, M.; Schulder, M.; Khan, A.; Helmke, H.; Klakow, D. Context-Based Recognition Network Adaptation for Improving on-Line ASR in Air Traffic Control. In Proceedings of the 2014 IEEE Spoken Language Technology Workshop (SLT), South Lake Tahoe, NV, USA, 7–10 December 2014; pp. 13–18.
14. Blatt, A.; Kocour, M.; Veselý, K.; Szöke, I.; Klakow, D. Call-Sign Recognition and Understanding for Noisy Air-Traffic Transcripts Using Surveillance Information. *arXiv* **2022**, arXiv:2204.06309.
15. Pellegrini, T.; Farinas, J.; Delpech, E.; Lancelot, F. The airbus air traffic control speech recognition 2018 challenge: Towards atc automatic transcription and call sign detection. *arXiv* **2018**, arXiv:1810.12614.
16. Zuluaga-Gomez, J.; Veselý, K.; Blatt, A.; Motlicek, P.; Klakow, D.; Tart, A.; Szöke, I.; Prasad, A.; Sarfjoo, S.; Kolčárek, P.; et al. Automatic Call Sign Detection: Matching Air Surveillance Data with Air Traffic Spoken Communications. In Proceedings of the 8th OpenSky Symposium 2020, Online, 3 December 2020; p. 14.
17. Nigmatulina, I.; Braun, R.; Zuluaga-Gomez, J.; Motlicek, P. Improving Call sign Recognition with Air-Surveillance Data in Air-Traffic Communication. *arXiv* **2021**, arXiv:2108.12156.
18. Nigmatulina, I.; Zuluaga-Gomez, J.; Prasad, A.; Sarfjoo, S.S.; Motlicek, P. A Two-Step Approach to Leverage Contextual Data: Speech Recognition in Air-Traffic Communications. *arXiv* **2022**, arXiv:2202.03725.
19. Shetty, S.; Helmke, H.; Kleinert, M.; Ohneiser, O. Early Call sign Highlighting Using Automatic Speech Recognition to Reduce Air Traffic Controller Workload. In Proceedings of the 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022), New York, NY, USA, 24–28 July 2022.
20. García, R.; Albarrán, J.; Fabio, A.; Celorrio, F.; Pinto De Oliveira, C.; Bárcena, C. Automatic Flight Callsign Identification on a Controller Working Position: Real-Time Simulation and Analysis of Operational Recordings. *Aerospace* **2023**, *10*, 433. [CrossRef]
21. Hasan, M.R.; Hasan, M.M.; Hossain, M.Z. How Many Mel-frequency Cepstral Coefficients to Be Utilized in Speech Recognition? A Study with the Bengali Language. *J. Eng.* **2021**, *2021*, 817–827. [CrossRef]
22. Deshmukh, A.M. Comparison of Hidden Markov Model and Recurrent Neural Network in Automatic Speech Recognition. *Eur. J. Eng. Res. Sci.* **2020**, *5*, 958–965. [CrossRef]

23. Pauls, A.; Klein, D. Faster and Smaller N-Gram Language Models. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; Lin, D., Matsumoto, Y., Mihalcea, R., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2011; pp. 258–267.
24. Song, Y.; Jiang, D.; Zhao, W.; Xu, Q.; Wong, R.C.-W.; Yang, Q. Chameleon: A Language Model Adaptation Toolkit for Automatic Speech Recognition of Conversational Speech. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations, Hong Kong, China, 3–7 November 2019; Association for Computational Linguistics: Hong Kong, China, 2019; pp. 37–42.
25. Xue, B.; Hu, S.; Xu, J.; Geng, M.; Liu, X.; Meng, H. Bayesian Neural Network Language Modeling for Speech Recognition. *arXiv* **2022**, arXiv:2208.13259. [CrossRef]
26. Graves, A.; Jaitly, N. Towards End-to-End Speech Recognition with Recurrent Neural Networks. *arXiv* **2017**, arXiv:1701.02720.
27. Kovtun, O.; Khaidari, N.; Harmash, T.; Melnyk, N.; Gnatyuk, S. Communication in Civil Aviation: Linguistic Analysis for Educational Purposes. 2020. Available online: https://www.researchgate.net/publication/344876536_Communication_in_Civil_Aviation_Linguistic_Analysis_for_Educational_Purposes (accessed on 28 October 2023).
28. Ohneiser, O.; Helmke, H.; Kleinert, M.; Ehr, H.; Balogh, G.; Tønnesen, A.; Rinaldi, W.; Mansi, S.; Piazzolla, G.; Murauskas, Š.; et al. Understanding Tower Controller Communication for Support in Air Traffic Control Displays. In Proceedings of the 12th SESAR Innovation Days, SESASR Innovation Days 2022, Budapest, Hungary, 5–8 December 2022.
29. Bollmann, S.; Fullgraf, J.; Roxlau, C.; Feuerle, T.; Hecker, P.; Krishnan, A.; Ostermann, S.; Klakow, D.; Nicolas, G.; Stefan, M.-D. Automatic Speech Recognition in Noise Polluted Cockpit Environments for Monitoring the Approach Briefing in Commercial Aviation. *Proc. Int. Workshop ATM/CNS* **2022**, *1*, 170–175. [CrossRef]
30. Fuzzy String Matching. Available online: https://pypi.org/project/fuzzywuzzy/ (accessed on 20 November 2023).
31. Saïd, K.M.; Abdelouahid, L. The IBN BATTOUTA Air Traffic Control Corpus with Real Life ADS-B and METAR Data. In *Artificial Intelligence and Industrial Applications*; Masrour, T., Cherrafi, A., El Hassani, I., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2021; Volume 1193, pp. 371–384. ISBN 978-3-030-51185-2.
32. USA NWS Server. Available online: https://www.aviationweather.gov/metar (accessed on 30 April 2022).
33. Burileanu, D.; Pascalin, L.; Burileanu, C.; Puchiu, M. An Adaptive and Fast Speech Detection Algorithm. In *Text, Speech and Dialogue*; Sojka, P., Kopeček, I., Pala, K., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2000; Volume 1902, pp. 177–182. ISBN 978-3-540-41042-3.
34. ICOM ICOM 8500. Available online: https://www.icomeurope.com/files/IC-R8500_E_20100108.pdf (accessed on 10 February 2021).
35. AirNav RadarBox. Available online: https://www.radarbox.com/presenting-the-radarbox-xrange-receiver (accessed on 18 February 2021).
36. Helmke, H.; Slotty, M.; Poiger, M.; Herrer, D.F.; Ohneiser, O.; Vink, N.; Cerna, A.; Hartikainen, P.; Josefsson, B.; Langr, D.; et al. Ontology for Transcription of ATC Speech Commands of SESAR 2020 Solution PJ.16-04. In Proceedings of the 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, UK, 23–27 September 2018; pp. 1–10.
37. Diyasa, I.G.S.M.; Prasetya, D.A.; Idhom, M.; Sari, A.P.; Kassim, A.M. Implementation of Haversine Algorithm and Geolocation for Travel Recommendations on Smart Applications for Backpackers in Bali. In Proceedings of the 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 16–17 November 2022; pp. 504–508.
38. Godfrey, J.J. *Air Traffic Control Complete 1994, 4170704 KB*; Linguistic Data Consortium: Philadelphia, PA, USA, 1994.
39. Šmídl, L. *Air Traffic Control Communication*; University of West Bohemia: Plzen, Czech Republic, 2011.
40. Hofbauer, K.; Petrik, S.; Hering, H. The ATCOSIM Corpus of Non-Prompted Clean Air Traffic Control Speech. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*; Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Tapias, D., Eds.; European Language Resources Association (ELRA): Marrakech, Morocco, 2008.
41. Segura, J.C.; Ehrette, T.; Potamianos, A.; Fohr, D.; Illina, I.; Breton, P.-A.; Clot, V.; Gemello, R.; Matassoni, M.; Maragos, P. The HIWIRE Database, a Noisy and Non-Native English Speech Corpus for Cockpit Communication. 2007. Available online: https://www.academia.edu/24112615/The_HIWIRE_database_a_noisy_and_non_native_english_speech_corpus_for_cockpit_communication (accessed on 28 October 2023).
42. Zuluaga-Gomez, J.; Veselý, K.; Szöke, I.; Blatt, A.; Motlicek, P.; Kocour, M.; Rigault, M.; Choukri, K.; Prasad, A.; Sarfjoo, S.S.; et al. ATCO2 Corpus: A Large-Scale Dataset for Research on Automatic Speech Recognition and Natural Language Understanding of Air Traffic Control Communications. *arXiv* **2023**, arXiv:2211.04054.
43. How to Evaluate Speech Recognition Models. Available online: https://www.assemblyai.com/blog/how-to-evaluate-speech-recognition-models/ (accessed on 28 October 2023).
44. Chowdhury, S.A.; Ali, A. Multilingual Word Error Rate Estimation: E-WER3. *arXiv* **2023**, arXiv:2304.00649.
45. Yujian, L.; Bo, L. A Normalized Levenshtein Distance Metric. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 1091–1095. [CrossRef] [PubMed]
46. IATA Airline and Airport Code. Available online: https://www.iata.org/en/publications/directories/code-search/ (accessed on 7 March 2022).
47. Nguyen, V.N.; Holone, H. N-Best List Re-Ranking Using Syntactic Score: A Solution for Improving Speech Recognition Accuracy in Air Traffic Control. In Proceedings of the 2016 16th International Conference on Control, Automation and Systems (ICCAS), Gyeongju, Republic of Korea, 16–19 October 2016; pp. 1309–1314.

48. Hannun, A.; Case, C.; Casper, J.; Catanzaro, B.; Diamos, G.; Elsen, E.; Prenger, R.; Satheesh, S.; Sengupta, S.; Coates, A.; et al. Deep Speech: Scaling up End-to-End Speech Recognition. *arXiv* **2014**, arXiv:14125567.
49. CMUSphinx Training an Acoustic Model. Available online: https://cmusphinx.github.io/wiki/tutorialam/ (accessed on 15 January 2023).
50. Mozilla Training Your Own Model. Available online: https://deepspeech.readthedocs.io/en/r0.9/TRAINING.html# (accessed on 14 June 2022).
51. CMUSphinx Adapting the Default Acoustic Model. Available online: https://cmusphinx.github.io/wiki/tutorialadapt/ (accessed on 15 June 2022).
52. Mozilla Deep Speech Fine-Tuning. Available online: https://deepspeech.readthedocs.io/en/r0.9/TRAINING.html#fine-tuning-same-alphabet (accessed on 19 July 2022).
53. Stolcke, A. SRILM - an Extensible Language Modeling Toolkit. In Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002), Denver, CO, USA, 16 September 2002; pp. 901–904.
54. Aircraft Company/Telephony/Three−Letter Designator Encode. Available online: https://www.faa.gov/air_traffic/publications/atpubs/cnt_html/chap3_section_1.html (accessed on 22 September 2023).
55. Wong, J. String Matching with FuzzyWuzzy. Available online: https://towardsdatascience.com/string-matching-with-fuzzywuzzy-e982c61f8a84 (accessed on 2 December 2023).
56. Kleinert, M.; Venkatarathinam, N.; Helmke, H.; Ohneiser, O.; Strake, M.; Fingscheidt, T. Easy Adaptation of Speech Recognition to Different Air Traffic Control Environments Using the DeepSpeech Engine; Virtual. 2021. Available online: https://elib.dlr.de/145397/ (accessed on 28 October 2023).