aerospace

MDPI

Review

Spoken Instruction Understanding in Air Traffic Control: Challenge, Technique, and Application

Yi Lin

College of Computer Science, Sichuan University, Chengdu 610000, China; yilin@scu.edu.cn; Tel.: +86-135-4790-3121

Abstract: In air traffic control (ATC), speech communication with radio transmission is the primary way to exchange information between the controller and aircrew. A wealth of contextual situational dynamics is embedded implicitly; thus, understanding the spoken instruction is particularly significant to the ATC research. In this paper, a comprehensive review related to spoken instruction understanding (SIU) in the ATC domain is provided from the perspective of the challenges, techniques, and applications. Firstly, a full pipeline is represented to achieve the SIU task, including automatic speech recognition, language understanding, and voiceprint recognition. A total of 10 technique challenges are analyzed based on the ATC task specificities. In succession, the common techniques for SIU tasks are categorized from common applications, and extensive works in the ATC domain are also reviewed. Finally, a series of future research topics are also prospected based on the corresponding challenges. The author sincerely hopes that this work is able to provide a clear technical roadmap for the SIU tasks in the ATC domain and further make contributions to the research community.

Keywords: air traffic control; speech communication; automatic speech recognition; spoken instruction understanding; voiceprint recognition

Citation: Lin, Y. Spoken Instruction Understanding in Air Traffic Control: Challenge, Technique, and Application. *Aerospace* **2021**, *8*, 65. https://doi.org/10.3390/aerospace8030065

Academic Editor: Enric Pastor

Received: 25 January 2021
Accepted: 1 March 2021
Published: 5 March 2021

1. Introduction

1.1. Air Traffic Control Safety

As well known, air traffic control (ATC) is a complicated and time-varying system, in which operational safety is always a hot research topic. All achievements of an ATC center can be vetoed without any hesitation if any safety incident occurs. Air traffic safety is affected by various aspects of air traffic operation, from mechanical maintenance, resource management, to air traffic control. The safety of air traffic control is particularly important since the aircraft is already in the air. There is no doubt that any effort deserves to be made to improve ATC safety [1].

Air traffic is an extension of ground transportation, in which the aircraft flies in the three-dimensional (3D) earth space. Since no signs and traffic signals can be designed to provide required guidance for flights in the air, the pilot is almost "blind" once the aircraft has taken off, with few approaches to obtain traffic situation around the aircraft. Considering this issue, a position, called air traffic control, is established to ensure flight safety in a local airspace area. Various infrastructures were developed to collect the global air traffic situation (radar) and then transmit the information between air and ground (communication). Based on the real-time traffic situations and a set of well-designed ATC rules, the air traffic controller (ATCO) is able to direct the flight to their destination in a safe and highly efficient manner.

Although enormous efforts have been made to build a qualified controller pilot data link communication (CPDLC) [2], digital data transmission is still a dilemma of the communication for air traffic control. In the current ATC procedure, speech communication with radio transmission is still the primary way to exchange information between the

ATCO and aircrew. Therefore, the spoken instruction is transmitted in an analog manner and can be easily impacted by environmental factors, such as communication conditions, equipment error, etc. The spoken instruction contains a wealth of contextual situational dynamics that indicates the evolutions of the flight and traffic in the future [3,4], which is highly significant to the air traffic operation.

However, speech communication is also a typical human-in-the-loop (HITL) procedure in the ATC loop, since the current ATC system fails to process the speech signal directly. Any speech error may cause communication misunderstanding between the ATCO and aircrew [5,6]. As a first step of performing an ATC instruction, the communication misunderstanding likely results in incorrect aircraft motion states and further induces a potential conflict (safety risk) during the air traffic operation. Based on the statistics released by EUROCONTROL [7], up to 30% of all incidents related to speech communication errors (rising to 50% in airport environments) and 40% of all runway incursions also involve communication problems. Consequently, understanding the spoken instruction is particularly significant to detect the potential risk and further improve the ATC safety.

1.2. Spoken Instruction Understanding

The main purpose of understanding spoken instruction is to obtain the near-future traffic dynamics in advance and further to detect the communication errors that may cause potential safety risks. It not only enhances the information source of the current ATC system but also is capable of providing reliable warnings before the pilot performs the incorrect instruction (with more prewarning time).

As shown in Figure 1, the upper part presents the typical ATC communication procedure, while the lower part illustrates the required spoken instruction understanding (SIU) task in the ATC domain. In general, the SIU mainly consists of two steps: automatic speech recognition (ASR) and language understanding (LU) [8,9], as described below:

(a) ASR: translates the ATCO's instruction from speech signal into text representation (human- or computer-readable). The ASR technique concerns the acoustic model, language model, or other contextual information.

(b) LU: also known as text instruction understanding, with the goal to extract ATC-related elements from the text instruction since the ATC system cannot process the text directly, i.e., from text to an ATC-related structured data. The ATC elements are further applied to improve the operational safety of air traffic. In general, the LU task can be divided into three parts: role recognition, intent detection, and slot filling (ATC-related element extraction, such as aircraft identity, altitude, etc.).

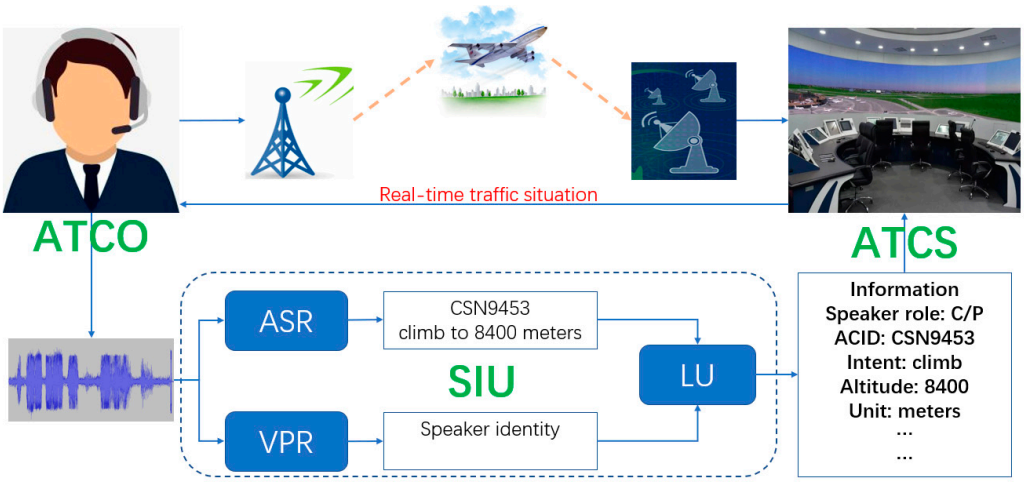

Figure 1. The air traffic control (ATC) procedure and spoken instruction understanding (SIU) roles in ATC.

In addition, since the ATC communication is a multi-speaker and multi-turn conversation system, to support the correlation among different instructions in the same sector, voiceprint recognition (VPR) is also needed to distinguish the identity of different speakers for the LU task. The VPR technique can also be applied for security purposes. For instance, if an ATCO instruction for a certain flight A is incorrectly responded to by the aircrew of flight B (usually the similar aircraft identity), the potential risks may be raised due to the mismatched traffic dynamics. In this way, the VPR technique is expected to be applied to detect this emergency situation from the perspective of the vocal feature of different speakers and further prevent the potential flight conflict (improve operational safety).

All the time, new techniques failed to be applied to the ATC domain promptly due to the various limitations (safety, complex environment, etc.). Although enormous academic studies for speech instruction have been reported in the ATC domain [10–14], currently, there is no valid processing devoted to speech instruction in a real industrial ATC system. The only contribution of speech communication is regarded as the evidence of the post-event analysis, which cannot present its important role in improving air traffic safety. Fortunately, thanks to a large amount of available industrial data storage and widespread applications of information technology, it is possible to obtain extra real-time traffic information from speech communication and further make contributions to the air traffic operation.

In this work, a comprehensive review is made about the spoken instruction understanding in the ATC domain, including the challenge of task specificities, techniques (especially machine learning-based ones), and prospect applications. In addition, future works that deserved to be focused on are also discussed in this work. The ultimate goal of this research is to provide a clear roadmap of concerned techniques for understanding the spoken instruction during the ATC procedure so that other researchers can make continuous contributions to improve air traffic safety.

1.3. Research Design

As illustrated in Figure 1, the SIU task in the ATC domain concerns the ASR, LU, and VPR procedures. Obviously, all the mentioned techniques have been widely studied in common application areas. In the early stage, common techniques have been evaluated to achieve the SIU task [1]; however, the results demonstrated that common techniques failed to complete the SIU task with acceptable accuracy due to task specificities. Thus, the key of the SIU task in the ATC domain is how to apply and improve the common techniques to properly address the task specificities.

In general, a "purpose–problem–solution–application" paradigm is applied to achieve the research design of this work, as shown below:

(1) Purpose: presents the significance of the SIU task to clarify why we study it (Section 1.1).
(2) Problem: presents the SIU system architecture (Section 1.2) and the difficulties we need to address to achieve the ASR task (Section 2).
(3) Solution: indicates the technical roadmap to achieve the SIU task (Section 3) and how can we improve it (Section 5).
(4) Application: introduces the potential application and benefits of applying the SIU task to real industrial systems (Section 4).

To this end, a systematic analysis is highly required to clarify the task specificities of the SIU task for this application, i.e., the ATC domain. In this work, the following ideas are considered to fully analyze the task specificities:

a. Firstly, as mentioned before, thanks to a large amount of available industrial data storage and the development of deep learning techniques, the performance of concerned techniques of the SIU task are greatly improved in recent years. Therefore, this research mainly focuses on improving the deep learning-based approaches to achieve the SIU task.
b. As is well known, the deep learning-based model is a kind of data-driven approach, which achieves the desired tasks (specifically, the pattern recognition tasks) by fitting

the complicated distribution between the input and the output data. That is to say, the training data is essential to the deep learning model, whose performance highly depends on the quality of the training samples.

c. Following the last description, the analysis of the task specificities of the SIU task in the ATC domain will focus on the input and output of the SIU techniques, i.e., ASR, LU, and VPR. In general, the input and output of the SIU model consist of the ATC speech, vocabulary, and ATC-related elements, as can be found in Figure 1.

d. In succession, based on the production mechanism of the ATC speech and ATC rules, a systematic analysis for the task specificities is achieved from various perspectives.

As shown in Figure 2, the research design of this work is organized as a top–down architecture to present the aforementioned ideas. In addition, a concept abstract from task specificities to technique targets is also needed to guide the technical improvement, including the data collection, framework, network architecture, etc. An intuitive and efficient way is to apply and improve similar studies to the ATC domain, in which dedicated improvements are also required to enhance the task performance. In this work, existing works for common applications of the SIU tasks are firstly reviewed. Meanwhile, papers for addressing the task issues (related to the task specificities in the ATC domain) are also provided to clarify the SIU research. Finally, the ATC-related applications deserve to be paid more attention since they are the ultimate goal for studying the SIU task, i.e., obtaining real-time information to support the ATC safety improvement.

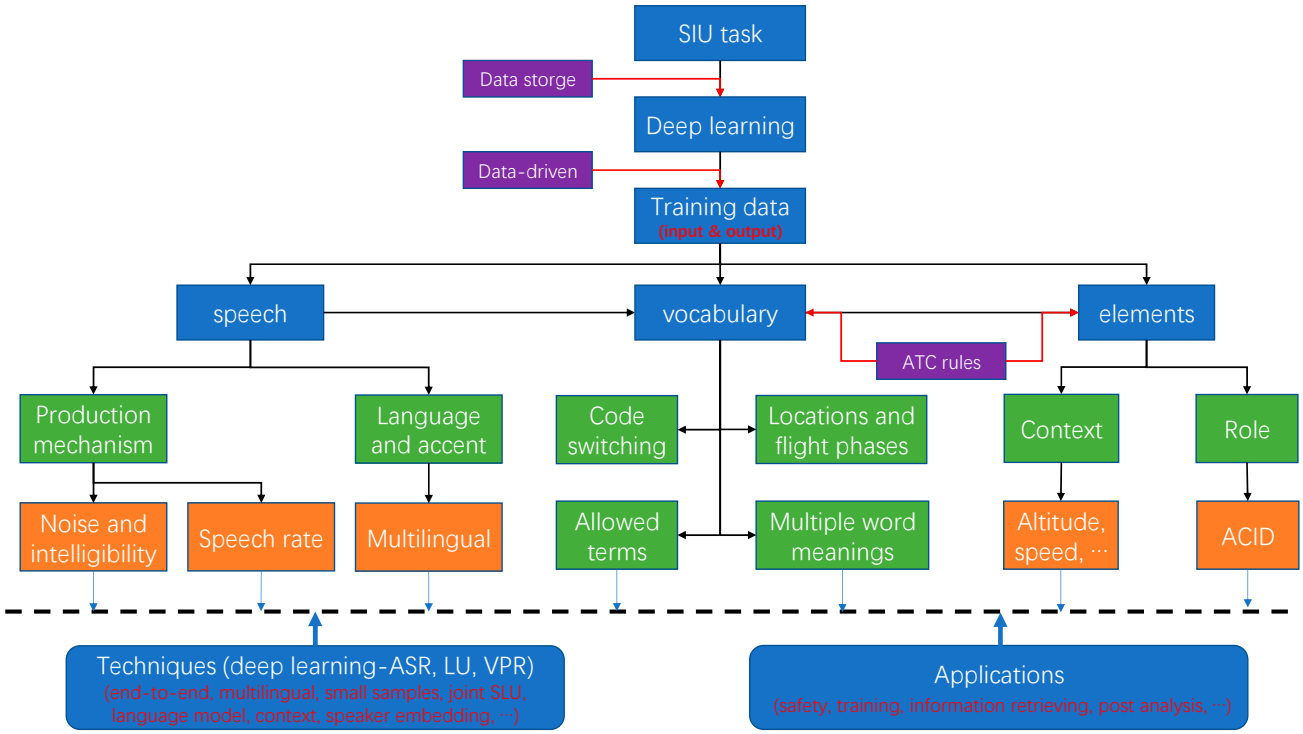

Figure 2. Research design in this work.

1.4. Document Structure

Based on the aforementioned descriptions, the rest of this paper is organized as follows. The technique challenges raised by the ATC task specificities are firstly analyzed and summarized in Section 2. The techniques for common applications are reviewed in Section 3, where the works for the spoken instruction understanding in the ATC domain are also reviewed and discussed. The applications of concerning techniques are prospected in Section 4, including the real-time ATC application and post-analysis. The future works

that have the potential to improve the SIU task performance are provided in Section 5. Finally, this paper is concluded in Section 6.

2. Challenges

In this section, the task specificities of understanding spoken instruction in the ATC domain are firstly summarized, and the technical challenges are also analyzed to infer the required technique improvements in this section. The challenges concern all the techniques in this work in detail in the following sub-sections.

2.1. Data Collection and Annotation

Currently, almost all state-of-the-art ASR/LU/VPR models are constructed by the data-driven mechanism, and the quality of training samples greatly affects the model performance [15]. On the one hand, due to the safety and intellectual property issues in the ATC domain, it is hard to collect sufficient training samples to develop a qualified speech recognition system. Even for the air traffic-related groups, the communication speech cannot be shared with other research institutions or companies. On the other hand, the transcriptions of the spoken instruction are domain-dependent, in which some vocabularies only apply to the ATC domain, such as "squawk". Furthermore, lots of vocabulary words are newly generated based on the ATC rules, such as the waypoint name "PIKAS", "AGULU", etc. That is to say, annotating the ASR training samples in the ATC domain is an expert-dependent work in which staff must learn a lot of required ATC knowledge to be competent for this job.

Thus, collecting and annotating training samples is costly and laborious work. Various speech corpora have been built to study the ASR technique, as shown in Table 1 [16]. From the table, we can see that it is easy to collect a qualified ASR corpus (up to thousands of hours) to develop an ASR system for common applications in different languages. However, due to the domain-specific characteristics, it is extremely difficult to collect and annotate sufficient samples to develop a desired ASR system in the ATC domain, i.e., only tens of hours. Therefore, from the perspective of machine learning techniques, developing ASR models on small annotated speech samples is an evitable research topic in the ATC domain.

Table 1. A summary of the speech corpora [16].

No.	Corpus	Language	Domain	Size (Hour)	Access
1	LibriSpeech [17]	English	novels	960	public
2	TED-LIUM3 [18]	English	TED talks	452	public
3	Switchboard [19]	English	Telephone	260	public
4	THCHS30 [20]	Chinese	newspapers	30	public
5	AISHELL-V1 [21]	Chinese	multidomain	500	public
6	AISHELL-V2 [22]	Chinese	multidomain	1000	application
7	ATCSpeech [23]	Chinese/English	real ATC	59	application
8	ATCOSIM [24]	English	simulated ATC	11	public
9	LDC94S14 [25]	English	airport	70	paid
10	Airbus [26]	English	pilot	40	unavailable

2.2. Volatile Background Noise and Inferior Intelligibility

The volatile background noise and inferior intelligibility are the most prominent specificities of the speech signal for ATC communication, which are analyzed as follows:

a. Due to the resource limitation of the radio transmission, an ATCO usually communicates with several pilots in the same communication frequency. Therefore, the equipment and radio transmission conditions change as the speaker changes [27], which further results in volatile background noise in the same frequency, as shown in Figure 3. It is clear that the feature intensities distribute in different frequency ranges due to the different noise models (communication equipment or conditions).

b. In general, the speech signal of ATC communication is recorded in a very low sample rate (8000 Hz), which degenerates the intelligibility of the speech.
c. Since the spoken instruction is transmitted by radio communication, the robustness of the communication is always a fatal obstacle to receive high-quality speech for both ATCOs and pilots in the ATC domain.
d. In general, the speech rate of ATC speech is higher than that in daily life due to the time constraints of the traffic situation. This fact severely damages the quality and intelligibility of the ATC speech. For example, speaking "two two" in a fast speech rate may probably cause an overlapped speech segment, and the ASR system can only output one "two" (incorrect results).

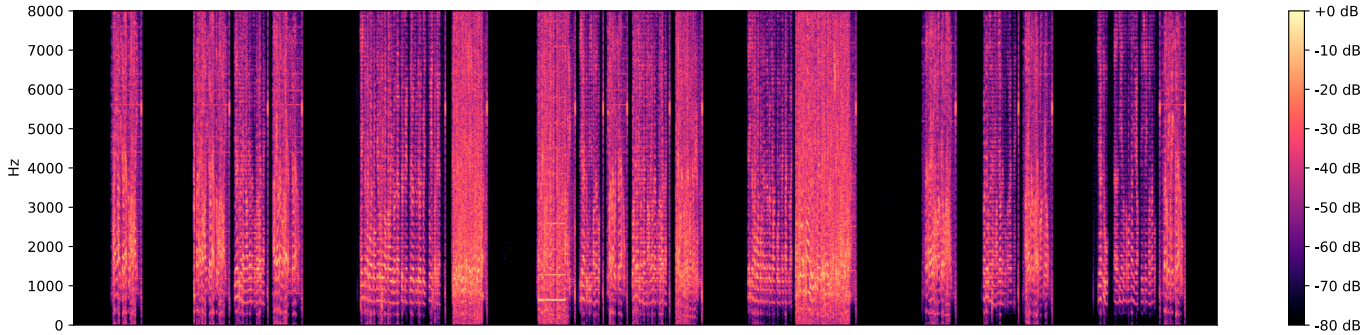

Figure 3. The volatile background noise model in a same frequency.

2.3. Unstable Speech Rate

As illustrated before, the rate of ATC speech is generally higher than that of in daily life. However, the speech rate is also influenced by the following factors:

a. Traffic situation: the ATCO unconsciously speeds up his speech when facing a busy sector or peaking hours.
b. Language: the ATCOs usually speak their native language at a higher speech rate than that of other languages. For example, ATCOs in China speak Chinese at a higher speech rate than English.
c. Emotion: The speech rate is also impacted by the ATCO's emotion and presents an irregular and unstable state.

From the perspective of signal processing, the unstable speech rate shows different temporal resolution, i.e., the speech durations of the same vocabulary are highly varied among speech segments. The unstable speech rate will further increase the difficulties of the feature engineering approach, aiming to extract discriminative features to support the ASR task. From the above descriptions, it can be seen that how to deal with the unstable speech rate is an essential issue for improving the ASR performance in the ATC domain.

A comparison of the speech rate for common and ATC speech corpora is listed in Table 2. The mean and standard deviation of the speech rate are measured, in which the w/s means words per second. It can be seen that the speech rate of ATC speech is higher than that of the common corpora for both Chinese and English. Specifically, the rate of Chinese speech is higher than that of English speech. In addition, the ATC speech rate is more unstable than that of the common corpora, i.e., higher standard deviation, 1.10 vs. 0.47 for Chinese, 0.75 vs. 0.47 for English.

Table 2. The comparison of the speech rate for different speech corpus [27].

Language	Corpus	Mean (w/s)	Standard Deviation (w/s)
Chinese	ATCSpeech [23]	5.15	1.10
	THCHS-30 [20]	3.48	0.47
English	ATCSpeech [23]	3.28	0.75
	Librispeech [17]	2.73	0.47

2.4. Multilingual and Accented Speech

In general, English is the universal language for ATC communication all over the world. However, some countries, including China, France, Russian, etc., still used to communicate with their domestic flights in local languages, while English is for international flights. Similarly, many greeting words are also in the English ATC speech, such as nihao, xiexie, etc. Most importantly, since the speakers (especially pilots) come from different countries and cultures, and perform flights all over the world, ATC communication is naturally a kind of accented speech, even for the English ATC speech. Therefore, multilingual and accented speech is a prominent characteristic for the SIU research (ASR and LU) in the ATC domain. This fact also causes a situation in which existing approaches and models cannot achieve the SIU task in the ATC domain, and it inspires us to develop dedicated approaches and models for the ATC application. Many technique challenges are indispensably required to be addressed for multilingual and accented ASR tasks, such as the scale of pronunciation and grapheme, vocabulary design, length dilemma of the output text sequence, etc.

2.5. Code Switching

To eliminate the misunderstanding between the ATCO and aircrew, the international civil aviation organization (ICAO) published the standard pronunciation of the vocabulary words, which is called code switching [28]. The pronunciation of the homophone and near-syllable words are switched. Similarly, the ATC department of other concerned countries also published the pronunciation of the ATC terminologies in their local languages. Some examples are shown in Table 3, in which both the English and Chinese are concerned. This specificity burdens the difficulties of sample annotation, and it messes up the correlations with common words for both the ASR and LU tasks. In addition, this real fact in Chinese ATC speech forces us to study/train a special recognition engine, since existing models are never trained by code-switching vocabularies let alone able to predict them correctly.

Table 3. Examples of code-switching words in the ATC domain.

Language	Common	ATC
English	three	tree
	five	fife
	nine	niner
	thousand	tousand
Chinese	ling	dong
	yi	yao
	er	liang
	qi	guai

2.6. Vocabulary Imbalance

To improve communication efficiency and operational safety, ICAO published the standard operation procedure of ATC communication, in which only predefined terminologies are allowed in ATC communication [28]. However, in practice, many out-of-vocabulary (OOV) words are still widespread in ATC speech, such as modal words. As described in Figure 4, the word frequencies for both Chinese and English ATC speech are extremely unbalanced in the ATCSpeech corpus. Up to 40% of words appear less than ten times, while some words appear almost ten thousand times.

The OOV words cause an unbalanced dataset for the machine learning approach, i.e., long-tail problem. From the perspective of model training, this fact may severely degrade the recognition performance, i.e., classification accuracy between speech frames and text labels. Therefore, addressing the class imbalance is also a key to improve the recognition accuracy of the SIU task in the ATC domain.

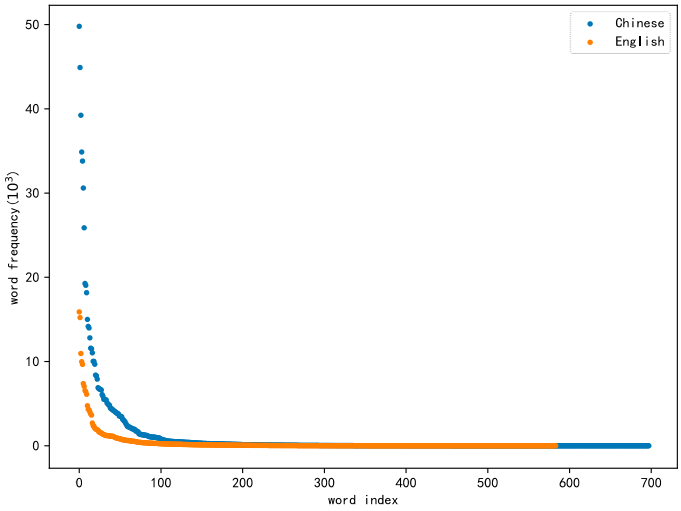

Figure 4. The word frequency of ATC speech for both Chinese and English speech.

2.7. Generalization of Unseen Samples

The generalization is a core evaluation measurement of data-driven models, which is particularly important to the SIU research in the ATC domain. On the one hand, the distributions of speech features are varied based on the device and communication conditions, which highly depend on the model generalization to obtain the desired performance. On the other hand, the vocabularies for different control centers or locations have a distinct and unique set. For instance, the term "line up" is only for the ATCO of an aerodrome tower (flight phase-dependent). Similarly, the waypoint "PIKAS" is only for a specific sector (location-dependent), i.e., the 23rd sector in Chengdu area control center (ACC), China. Therefore, enhancing the generalization of the SIU models among different control centers or locations is a necessary technique to improve the SIU applicability, especially under the limitation of annotating sufficient training samples.

2.8. Ambiguous Word Meaning

In the ATC domain, the digit is the common vocabulary for different goals. For instance, a digit can be used to represent the airline number, flight level, altitude, speed, heading, runway number, etc. The widespread usage of digits may result in the following disadvantages:

(a) Since digits are commonly used in the speech corpus, the distributions or the contextual correlations between digits and other words are extremely similar. This fact reduces the effectiveness of the language model (LM) for text correction to a certain extent for the speech recognition and language understanding task.

(b) For the LU task, it is hard to design a fair and distinct label (slot filling) for digits in the ATC-related corpus. If all the digits are regarded as the same label (i.e., digit), the actual role for different goals (airline number, flight level, altitude, etc.) will be confused. If all the digits are explained as different labels based on their real goal, a large amount of one-to-many relationships will be generated. Both situations have a possibility of degenerating the final performance of the LU model.

In addition, the flight callsign can also be represented in different formats. For instance, both "Lufthansa" and "DLH" denote the Lufthansa airline. In summary, distinguishing the word meaning from ambiguous texts based on the contextual situation is required to achieve a high-accuracy LU task.

2.9. Role Recognition

Based on the ATC procedure, ATC communication can be defined as a task-oriented conversation task, focusing on detecting the potential risks from ATCO speech and repetition errors from pilot speech. In short, role recognition is an indispensable precondition

of any business-related process. Different safety check procedures are applied to the instructions spoken issued by different roles (ATCO or pilot). Typically, resource conflict check is designed for ATCO speech, while repetition check is for pilot's speech. In general, the ICAO requested that the ATCO instruction starts with a valid aircraft identification (ACID) to specify the communication object, while the pilot instruction must end with their ACID during the repetition. However, in practice, some pilots ignore the ATC rules, whose instruction even starts with an ACID. The text-dependent role recognition approach may be confused under this situation, and it further invalidates the subsequent safety check tasks. That is to say, accurate role recognition is an important and fundamental step of the LU task in the ATC domain.

2.10. Contextual Information

All the time, the SIU task is mainly addressed from the perspective of acoustic modeling, in which the LM is applied to correct the results based on semantic meanings. For the SIU research in the ATC domain, the standardized phraseology plays a significant role in improving the LM effectiveness. Most importantly, the contextual situational information from other information sources (such as radar, flight plan, etc.) provides a more accurate and targeted reference for the ASR research. For instance, the ASR result of an ACID is "CSC 7019", while only the flight "CSC 7016" really exists based on radar detection. In this way, a correction from the "6" to "9" based on the contextual information may be a promising way to improve the ASR performance since the pronunciation is easy to be confused (i.e., "j iu" vs. "l iu"). In summary, incorporating the contextual knowledge into the SIU research in a proper manner is a practical and highly efficient way to improve the final performance; i.e., it is more realistic in a certain ATC environment.

3. Technique

In this section, existing works that relate to the SIU task are reviewed, including automatic speech recognition, language understanding, and voiceprint recognition technique. For all the techniques, the research advances in the common application are firstly presented to provide an overall glimpse, and those for the ATC domain are also reviewed here to clarify its current development. In the section, the way that the following techniques can be used to address the aforementioned challenges is also analyzed.

3.1. Automatic Speech Recognition

As the aforementioned illustration, ASR is the first step of the SIU task, which achieves the representation conversion from speech signal to human- or computer-readable texts. As shown in Figure 5, an ASR system consists of the acoustic model (AM) and LM, in which the AM can be hidden Markov model (HMM)-based or end-to-end paradigm, while the LM can be implemented by the n-gram or neural network architecture.

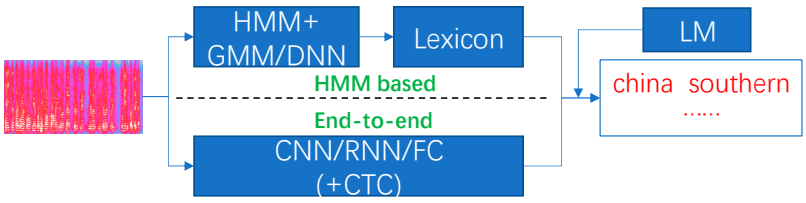

Figure 5. The ASR scheme (hidden Markov model (HMM)-based and end-to-end framework).

The ASR research can be traced back to the 1950s [29], and it has undergone several technical improvements, as described below:

(1) Statistical models: The introduction of statistical models advanced the first technical peak of the ASR research, which achieves the goal of large vocabulary continuous speech recognition (LVCSR). The hidden Markov model (HMM) [29] was proposed to capture the state transitions among continuous phonemes, while the Gaussian

mixture model (GMM) was applied to build the distribution between the state and the vocabulary unit [30]. Currently, the HMM/GMM framework still plays an important role in the ASR research.

(2) Hybrid neural network models: Thanks to the improvement of the deep neural network (DNN), it was also proposed to the ASR research to replace the GMM, which further generates the HMM/DNN framework [31]. As expected, the HMM/DNN showed desired performance improvements over the HMM/GMM framework, which also promotes the ASR research into the deep learning era.

(3) End-to-end models: Due to the strict requirements of the alignment between speech and vocabulary, Graves et al. proposed a novel loss function called connectionist temporal classification (CTC) [32]. The CTC loss function also formulated a new framework, i.e., it is also known as the end-to-end-based ASR model. The end-to-end ASR model is able to automatically align the speech and text sequence by inserting the blank label, which formulates a more intuitive pipeline [33,34]. The end-to-end framework reduces the requirement of expertise-dependent knowledge and greatly promotes the popularization of the ASR study for common researchers. Many outstanding research outcomes were obtained based on this framework, such as Deep speech 2 (DS2) [35], Jasper [36], CLDNN [37], DeepCNN [38], etc.

(4) Sequence-to-sequence models: Lately, the sequence-to-sequence (S2S) mechanism was also transferred to the ASR research [39,40]. Recently, the attention mechanism [41–44] and Transformer architecture [45–48] were also improved to address the ASR issues and showed desired performance improvement.

As to the multilingual ASR task, a sequence-to-sequence model was proposed to recognize nine Indian dialects [49]. Li et al. proposed a Unicode-based multilingual ASR model, which can also be used for the text-to-speech task [50]. The phoneme was regarded as the modeling unit to achieve the multilingual ASR task [51]. A shared network was designed as the backbone architecture to translate both the Mandarin and English speeches [52]. The code-switching and multi-task mechanism were proposed to improve the performance of the multilingual ASR model [53].

Learning from the approaches and models in common ASR applications, great efforts have been made to achieve the ASR task in the ATC domain. In Table 4, the existing works that relate to the ASR research in the ATC domain are reviewed, in which the framework, the concerned challenges, and technical details are also concerned.

Table 4. The automatic speech recognition (ASR)-related studies in the ATC domain.

Papers	Technique Details	Challenges Concerned
[1]	Independent end-to-end models, DS2, ASR, ATC safety monitoring	Sections 2.4 and 2.5
[27]	An integrated cascaded model, DS2+S2S, multilingual ASR	Sections 2.2–2.7
[16]	Independent end-to-end models, DS2, ASR, pretraining, transfer learning	Sections 2.1 and 2.4–2.6
[8]	Cascaded model, DS2+S2S, SIU, multi-level LMs	Sections 2.2, 2.4, 2.5, 2.8 and 2.9
[54]	Complete end-to-end model, representation learning, multilingual, pretraining	Sections 2.1–2.6
[55]	HMM/DNN, data augmentation, iterative training using unlabeled samples	Sections 2.1 and 2.7
[56]	HMM/DNN, context-aware rescoring, SIU task	Sections 2.5, 2.7, 2.8 and 2.10
[57]	Semi-supervised for transfer learning, DNN based models	Sections 2.1, 2.2, 2.5 and 2.7
[14]	AcListant® based traffic dynamic sensing, Arrival Managers (AMAN)	Sections 2.7 and 2.8
[11]	Cross-task adaption, HMM framework, transfer learning	Sections 2.2, 2.3 and 2.7
[58]	N-best list re-ranking, ATCOSIM, syntactic knowledge	Section 2.10
[59]	traffic dynamic sensing	-
[12]	HMM-based framework, Spanish and English ATC speech, SIU task	Sections 2.4 and 2.5
[60]	French-accented English ATC speech, Time Delay Deep Neural Network (TDNN)	Sections 2.1, 2.2, 2.4 and 2.5
[61]	170 h of ATCO speech, TDNN-based benchmark	Sections 2.2, 2.4 and 2.7
[62]	Improve intelligibility by reducing speech rate	Sections 2.2 and 2.3
[63]	Speech corpus for ASR and text-to-speech task	Sections 2.2, 2.3 and 2.5
[64]	Callsign correlation between ATC speech and surveillance data	Sections 2.1, 2.5 and 2.10

3.2. Language Understanding

For the SIU task in the ATC domain, the language understanding (also known as text instruction understanding) follows the ASR procedure, and it may follow the VPR procedure if a text-independent role recognition is required. The main purpose of the LU task is to extract ATC-related elements from ASR results. The LU consists of the following research topic:

(1) Role recognition: details as illustrated in Section 2.9.
(2) Intent detection: extract the controlling intent (CI) from the text instruction. The CI is a set of predefined ATC-related classes, such as climb, descend, heading, etc.
(3) Slot filling: analyze every word in a text instruction to obtain the contextual types, which are called instruction elements (IE). Similarly, the IE is also a set of predefined ATC-related classes, such as airline, flight number, altitude, speed, etc.

An example of the samples for language understanding is listed in Table 5, in which the prefixes "B-" and "I-" denote the beginning and connection of a semantic element, respectively. AL and CS are the airline company and callsign of the flight, respectively. TL represents "turn left", whose target parameter is 330. Similarly, CL denotes "climb", with the target parameter of 1200 m.

Table 5. An example of language understanding samples [1].

Text	CCA	4012	Turn Left	Heading 330	Climb to	1200 m
Slot filling	B-AL	B-CS	B-TL	I-TL	B-CL	I-CL
Intent	Turn left and climb					

Actually, both the intent detection and slot-filling task can be regarded as a special type of the spoken language understanding (SLU) task. The role recognition is similar to the intent detection task, and it is defined as a text classification task, i.e., classify the speaker role based on the text instruction.

In the early stage, the intent detection and slot-filling task were solved separately. With the development of deep learning techniques [65–67], the two tasks (intent detection and slot filling) were achieved in a joint model. A brief illustration of technical details can be found in Figure 6.

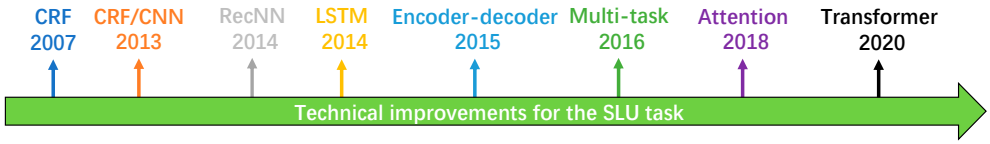

Figure 6. The research improvement of the spoken language understanding (SLU) task.

(1) Intent detection: It is a classification task. Various models were proposed and improved to achieve this task, including the generative machine learning models (such as Bayesian [68], HMM [69], etc.), and discriminative models (such as logistic regression [70], maximum entropy [71], conditional random fields (CRF) [72], support vector machine (SVM) [73], etc.). Deep learning models, including recurrent neural network (RNN) [74] and convolutional neural network (CNN) [75], were also introduced to achieve the intent detection.
(2) Slot filling: A maximum entropy Markov model (MEMM) [76] was proposed to achieve the information extraction and segmentation from texts. The CRF was also improved to achieve the slot filling task in [72]. The RNN block [77] and long short-term memory (LSTM) [78] were also applied to improve the performance by building long-term dependencies among the input text sequence.
(3) Joint model: Liu et al. proposed an LSTM-based model to achieve intent detection and the slot-filling task jointly [79]. A combined model based on the CNN and triangular CRF was also improved to jointly achieve the SLU task [75]. The recursive neural

networks (RecNN) architecture [80] and gated recurrent unit (GRU) [81] were studied to obtain the semantic utterance classification and slot filling jointly. Recently, the attention mechanism [82,83] and transformer architecture [84,85] were also proposed to address existing issues in the SLU research.

Note that role recognition is a special requirement in the ATC domain, and no studies can be found in the literature. It can be defined as a two-class classification problem, i.e., ATCO or pilot, whose technique implementation is similar to that of intent detection. In general, the text instruction is regarded as the input to achieve the role recognition task, where the embedding, RNN, and DNN layers are applied to predict the class label. In addition, to address the issues illustrated in Section 2.9, voiceprint recognition will be introduced in the following Section 3.3, which recognizes the speaker role from the perspective of acoustic features. In summary, a proper combination based on the text instruction and speaker feature is a promising solution for this task.

Language understanding is an essential step to bridge the gap between the ASR and the ATC system; i.e., it converts the text into the predefined data structure. The following works concerning this research topic are summarized in Table 6. It can be seen that the research topic of language understanding focuses on the concept definition, while the extraction approaches still needed to be improved in the future.

Table 6. The language understanding-related studies in the ATC domain.

Papers	Technique Details
[1]	Joint S2S model, 26 controlling intent, safety monitoring
[8]	Joint S2S model, 26 controlling intents, 55 instruction elements
[10]	SESAR 2020 Solution PJ.16-04, extra qualifier, conjunction
[86]	10 ontologies for ATC command extraction
[57]	More label classes (such as QNH), conditional clearances
[87]	Command definition for ASR rescoring

3.3. Voiceprint Recognition

Voiceprint recognition (VPR) is a task to identify the speaker of a given utterance. VPR is one of the desired alternatives for text-dependent role recognition in the SIU system, which directly takes the speech utterance as input to predict the speaker identification. Thus, it is capable of reducing the cascaded errors raised by voice activity detection (VAD) and the ASR model. The VRP technique is also the essential component of task-orient conversation management in the ATC domain. As illustrated in Figure 7, the VPR is generally divided into two pipelines: training and application. The training pipeline is to extract a discriminative feature representation from an input speech waveform, while the application pipeline makes a decision (accept/reject) based on the similarity evaluation between the template feature vector and real-time feature vector to be recognized.

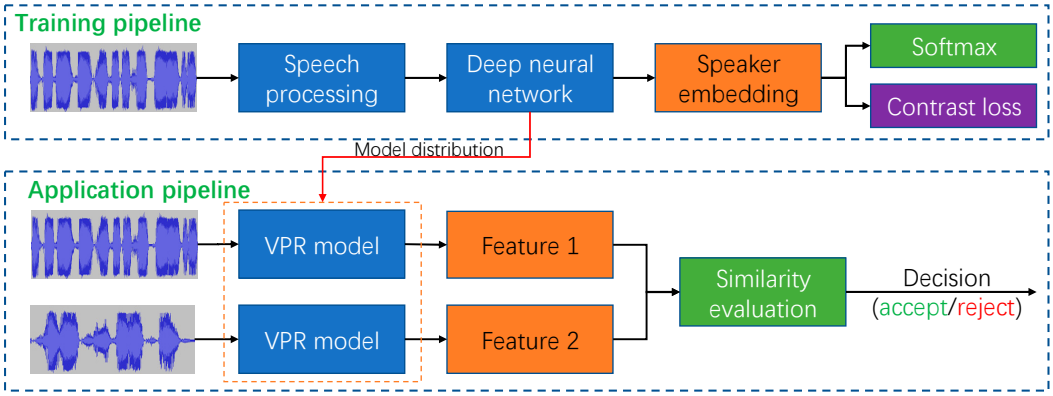

Figure 7. The voiceprint recognition (VPR) pipelines.

The VPR research can be traced back to the 1960s [88], and it has undergone several technical improvements, as described below:

(1) Template matching: In the early stage, the VPR approaches directly calculated the similarity between the time-frequency spectrum to determine whether two utterances come from the same speaker [89]. Then, this type of approach was improved to consider the speech diversity in the temporal dimension, which further generated other approaches, such as the dynamic time warping (DTW) [90] and vector quantization (VQ) [91], etc.
(2) Statistical models: As the GMM model has made great progress in the ASR research, it was also explored to build a robust text-independent VPR system [92]. Moreover, other models were further introduced to improve the performance and robustness, such as the universal background model [93] and support vector machine [94].
(3) Factor analysis models: To compensate for the channel mismatching and the independence problems of Gaussian components, the joint factor analysis (JFA) [95] and i-vector [96] approaches were studied, which are widely popular in industrial applications, making the VPR technique into a new stage.
(4) Deep neural network models: With the d-vector [97] proposed in 2014, DNN-based models showed the ability to directly optimize the discriminations among different speakers. Subsequently, both metric learning and representation learning were also widely used in the VPR research. In this pipeline, the DNN architecture is used to extract high-level abstract embeddings as voiceprint representation features, while metric learning is applied to optimize the networks. Enormous research outcomes were generated based on this core idea, such as Deep Speaker [98], X-vector [99,100], j-vector [101,102], SincNet [103], etc.

To the best of the authors' knowledge, there are no studies that aim to recognize the speaker role (ATCO or pilot) in the ATC domain, which is further applied to detect communication errors.

4. Applications

As is well known, the spoken instruction understanding task is to extract ATC-related concepts and elements (i.e., traffic dynamics) from the ATC communication speech, which serves as an extra information source of the current ATC system. The obtained information by the SIU task can be performed in both real-time and post-analysis applications. In this section, the possible subsequent ATC applications are reviewed and prospected based on the spoken instruction understanding techniques in this work.

4.1. Information Enhancement

After obtaining the real-time traffic dynamics from the ATC speech, a natural and intuitive idea is to feed the obtained information into the current ATC system, which takes the ATC speech into the information processing loop in an automatic manner. The ATC system providers, including the INDRA (Madrid, Spain) and Thales (La Défense, France) cooperation, have studied the way to feed the real-time information to the ATC system based on open-source ASR tools [10], such as the electronic strip system. This approach not only enhances the source of real-time information but also improves the timeliness of information sensing (extraction in advance).

4.2. Communication Error Detection (CED)

Once the information of ATC communication speech is obtained, various applications can be applied to detect the potential communication risks, as below:

(1) The instruction completeness: Confirm whether essential elements are embedded in the ATCO's instruction based on the ATC rules, such as target altitude for climb instruction. The purpose of this application is to encourage the ATCO to issue standard instructions to eliminate misunderstandings between ATCO and aircrew during the ATC communication.

(2) Resource incursion: Check whether the concerned resources of the ATCO instruction are valid or have conflicted with other operators from temporal and spatial dimensions, such as the closed runway detection, ground obstacle, etc. [104]. In this way, the potential risks can be detected in the stage of instruction issue and greatly improve the operational safety in advance.
(3) Repetition check: Check whether the pilot receives the ATCO's instruction in a correct and prompt manner. The repetition check error includes no response from aircrew, repetition error (intent or elements), etc. [5]. This application is able to reduce the risks raised by the incorrect transmission and understanding of the pilot instruction, which can eliminate the potential safety risk during the issue of instruction (before changing the aircraft motion states).

4.3. Conflict Detection Considering Intent

Currently, the ATC system failed to process ATC speech directly, let alone to understand it and further be applied to improve flight safety. Existing works only relied on the current aircraft state (radar) or flight plan to predict the flight trajectory [105,106] and conflict detection [107]. If the SIU task can be achieved with considerable high confidence, the flight trajectory can be predicted based on the consideration of controlling intent and its ATC-related elements [108]. In this way, the accuracy of the conflict detection approach can also be improved by a more reliable predicted trajectory. Most importantly, conflict detection can be performed at the issue time of the instruction (before the aircraft changes its motion state) and allows the ATCO to cope with an emergency with more pre-warning time.

4.4. Post Analysis and Processing

Currently, the ATC speech is completely recorded to serve as evidence of post-event analysis. Taking the recorded ATC speech, a large number of post-processing applications can be achieved, including (but not limited to) the following:

(1) Workload measurement: Evaluate the workload of an ATCO from the time and sector dimensions, such as flight peak hours or busy sectors [13,109]. Based on the evaluation results, more efficient and effective designs for the airspace sector are expected to be achieved to balance the ATCO workload, which is also helpful to improve the operational safety of air traffic. For instance, a frequent "correct" instruction may indicate that an ATCO is in a fatigued state, so that too many incorrect instructions appeared in the ATC speech.
(2) Performance evaluation: The ATC speech is a side view of real-time air traffic operation, in which the ATCO performance is enclosed as the conversation speech. Thus, the ATCO performance also deserves to be considered to detect the improper ATC actions and further improve ATCO's skills. For example, excessive extra instructions for changing aircraft motion state may indicate that the sector always faces potential risk so that the ATCO has to adjust the aircraft motion to resolve the potential conflict. Facing this situation, it is necessary to improve operational safety by enhancing the ATCO skills or designing a more proper standard operating procedure (SOP) during the ATC communication.
(3) Information retrieving: Currently, human hearing is the only way to search the ATC speech for a certain goal. Intuitively, based on the SIU technique, it is easy to search the target information (speech) from a long-duration continuous record speech, such as a certain flight number or a certain ATCO. This is strong support to the post-incident analysis in an automatic manner, since it is laborious and costly work undertaken by human staff.
(4) Event detection: Detect anomaly speech to support other analyses in the ATC domain. For instance, the "confirm" instruction is issued by many speakers in a certain sector or time period and may indicate that the communication condition between ATCO

and aircrew in the sector or time period is needed to be improved, such as the infrastructure malfunction or signal interference.

4.5. ATCO Training

In the ATC domain, ATCO training is given a particularly high priority, since ATCO is the core of the air traffic operation. Only a licensed ATCO is allowed to compete for a position in a real ATC environment. A series of knowledge and operation training requirements were published and requested by ICAO [110] to bring up a qualified ATCO. Currently, due to the technique limitation, a dedicated person is also needed to act as the pilot to assist the ATCO training, which requires the extra cost of ATCO training, i.e., human resource and training device (position). By combining the SIU approach and other advanced techniques (i.e., instruction generation and text-to-speech), an autonomous pilot agent is expected to be developed to serve as a human–machine interface and further replace the human-acted pilot during the air traffic controller training. It is clear that the training agent is able to greatly save the training and maintenance cost and improve the utilization of the training devices. This will speed up the progress of skill upgrade for ATCO facing a new system or SOP and further benefit to improve the air traffic safety. Most importantly, facing the limitation of physical attendance (such as COVID-19), the autonomous training agent is capable of achieving a virtual training system through online, which solves the current dilemma of ATCO training.

5. Future Research

Based on the aforementioned technique challenges and exiting works, the possible research topics related to the SIU task in the future are prospected, from the perspective of automatic speech recognition, language understanding, and voiceprint recognition, as summarized below:

5.1. Speech Quality

(1) Speech enhancement: Facing the inferior speech quality in the ATC domain, an intuitive way is to achieve the speech enhancement to further improve the ASR and VPR performance. With this technique, a high-quality ATC speech is expected to be obtained to support the SIU task and further benefit to achieve the high-performance subsequent ATC applications.

(2) Representation learning: Facing the diverse distribution of speech features raised by different communication conditions, devices, multilingual, unstable speech rate, etc., there are reasons to believe that the handcrafted feature engineering algorithms (such as MFCC) may fail to support the ASR and VPR research to obtain the desired performance. The representation learning, i.e., extracting speech features by a well-optimized neural network, may be a promising way to improve the final SIU performance.

5.2. Sample Scarcity

(1) Transfer learning: Although a set of standardized phraseology has been designed for the ATC procedure, the rules and vocabulary still depend on the flight phases, locations, and control centers. It is urgent to study the transfer learning technique among different flight phases, locations, and control centers to save the sample requirement and formulate a unified global technical roadmap.

(2) Semi-supervised and self-supervised research: Since the data collection and annotation is always an obstacle of applying advanced technology to the ATC domain, the semi-supervised and self-supervised strategies are expected to be a promising way to overcome this dilemma, in which the unlabeled data samples can also be applied to contribute the model optimization based on their intrinsic characteristics, such as that in the common application area.

(3) Sample generation: Similar to the last research topic, sample generation is another way to enhance the sample size and diversity and further improve the task performance, such as text instruction generation.

5.3. Contextual Information

(1) Contextual situational incorporation: As illustrated before, contextual situational information is a powerful way to improve SIU performance. Due to the heterogeneous characteristics of the ATC information, existing works failed to take full advantage of this type of information. Learning from the state-of-the-art studies, the deep neural network may be a feasible tool to fuse the multi-modal input by encoding them as a high-level abstract representation using the learning mechanism and further make contributions to improve the SIU performance.
(2) Multi-turn dialog management: Obviously, the ATC communication in the same frequency is a multi-turn and multi-speaker dialog with a task-oriented goal (ATC safety). During the dialog, the historical information is able to provide significant guidance to current instruction based on the air traffic evolution. Thus, it is important to consider the multi-turn history information to enhance the SIU task of current dialog, similar to what is required in the field of natural language processing.

5.4. Other Research Topics

(1) Joint SIU model: Currently, the ASR and LU tasks are achieved separately, i.e., a cascaded pipeline, which also leads to cascaded errors (reduces the overall confidence). In the future, a joint SIU model for automatic speech recognition understanding (ASRU) deserves to be studied to capture the task compatibility to promote the final performance, similar to that of the joint SLU model. In this way, the SIU task can be achieved in a more intuitive and clear processing paradigm.
(2) On-board SIU system: Currently, all the SIU studies are developed based on the requirements of the ground systems. The computational resource is heavily required due to the applications of the deep learning model. For future development, it is also attractive to achieve the SIU task for the on-board purpose (i.e., cockpit) and further construct a safety monitoring framework for the aircrew. In this way, a bi-directional safety-enhancing system is constructed for both the ATCO and aircrew, which is expected to ensure flight safety in a reinforced manner. To this end, the model transfer from the X86 platform to the embed system (such Jetson, NVIDIA, CA, USA) is the primary research to save the computational resource requirements, such as model compression, power reduction, etc.

6. Conclusions

In this work, a comprehensive review is made for the spoken instruction understanding in the ATC domain. The whole paper is categorized into three parts: challenge, technique, and application. The concerning techniques for the SIU task are firstly specified, in which a total of 10 challenges are summarized based on the ATC specificities. Lately, extensive works of concerned techniques are reviewed for both the common and ATC applications. A brief summary of this work can be found in Table 7, in which both the findings and conclusions are presented to provide the development of this issue. It can be seen that although great efforts have been made in this field, some key issues are still needed to be addressed properly. Finally, a series of future research topics are sketched in this work. The author sincerely hopes that this work can contribute to the research community of the spoken instruction understanding in the ATC domain.

Table 7. A summary of this work.

Section	Item	Findings	Conclusions or Future Research Topics
Challenges	Data collection and annotation	English corpus [24,26] Chinese/English corpus [23]	More corpora are required to build large-scale SIU systems in the ATC domain.
	Volatile background noise and inferior intelligibility	Multi-scale CNN [27]	Representation learning may be a promising way to overcome the mentioned issue.
	Unstable speech rate	Multi-scale CNN [27]	
	Multilingual and accented speech	Cascaded pipeline [8,27] Independent system [1,23]	The end-to-end multilingual framework.
	Code switching	Language model [27]	The author believes that the most efficient way is to build sufficient training samples.
	Vocabulary imbalance	Phoneme-based vocabulary [8,27] Data augmentation [1,26]	Sub-word-based vocabulary is a better tradeoff between the vocabulary size and sequence length.
	Generalization of unseen samples	Transfer learning [16]	Transfer learning from other domains is a feasible way to address this issue.
	Ambiguous word meaning	Currently, no literature is for this issue.	An intuitive way is to build a dictionary for synonyms pairs.
	Role recognition	Text-dependent SLU model [1,8]	VPR is a powerful text-independent way to achieve this task.
	Contextual information	Enumeration of possible information [56,87].	Deep information fusion using neural network is expected to improve the performance of this issue.
Techniques	Automatic speech recognition	Monolingual: HMM-based [26], deep learning based [1,16,23,60]. Multilingual: deep learning based [8,27,54].	Great efforts deserve to be made to promote the ASR task into an industrial level, including speech quality, contextual information, etc.
	Language understanding	Concept extraction [10], deep learning based SLU model [1,8].	More concept classes are required to cover the ATC-related elements, especially for the rarely used terms.
	Voiceprint recognition	Currently, there is no literature for this issue.	Building a corpus for the ATC environment is the key to train a qualified VPR system.
Applications	Information enhancement	Electronic strip system [10].	More applications are expected to be achieved based on the SIU task.
	Communication error detection	Studies based on ASR tools [5,12,14,60,62].	A way to improve the air traffic safety.
	Conflict detection considering intent	Flight trajectory considering intent [107].	Conflict detection considering intent should be studied to provide more warning time for ATCO.
	Post analysis and processing	(1) Workload measurement and performance evaluation [13,109]. (2) Currently, there is no literature on the information retrieving and event detection.	More applications are required to be explored to take full advantage of the SIU research outcomes.
	ATCO training	There is no literature for this issue.	It is very important to emphasize the SIU task in the ATC domain.

Funding: This research was funded by National Natural Science Foundation of China, grant number 62001315.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lin, Y.; Deng, L.; Chen, Z.; Wu, X.; Zhang, J.; Yang, B. A Real-Time ATC Safety Monitoring Framework Using a Deep Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 4572–4581. [CrossRef]
2. Rossi, M.A.; Lollini, P.; Bondavalli, A.; Romani de Oliveira, I.; Rady de Almeida, J. A Safety Assessment on the Use of CPDLC in UAS Communication System. In Proceedings of the 2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC), Colorado Springs, CO, USA, 5–9 October 2014; pp. 6B1-1–6B1-11.
3. Kopald, H.D.; Chanen, A.; Chen, S.; Smith, E.C.; Tarakan, R.M. Applying Automatic Speech Recognition Technology to Air Traffic Management. In Proceedings of the 2013 IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC), East Syracuse, NY, USA, 5–10 October 2013; pp. 6C3-1–6C3-15.
4. Nguyen, V.N.; Holone, H. Possibilities, Challenges and the State of the Art of Automatic Speech Recognition in Air Traffic Control. *Int. J. Comput. Electr. Autom. Control. Inf. Eng.* **2015**, *9*, 1916–1925.
5. Geacăr, C.M. Reducing Pilot/Atc Communication Errors Using Voice Recognition. In Proceedings of the 27th International Congress of the Aeronautical Sciences, Nice, France, 19–24 September 2010; pp. 1–7.
6. Glaser-Opitz, H.; Glaser-Opitz, L. Evaluation of CPDLC and Voice Communication during Approach Phase. In Proceedings of the 2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC), Prague, Czech, 13–18 September 2015; pp. 2B3-1–2B3-10.
7. Isaac, A. Effective Communication in the Aviation Environment: Work in Progress. *Hindsight* **2007**, *5*, 31–34.
8. Lin, Y.; Tan, X.; Yang, B.; Yang, K.; Zhang, J.; Yu, J. Real-Time Controlling Dynamics Sensing in Air Traffic System. *Sensors* **2019**, *19*, 679. [CrossRef]
9. Serdyuk, D.; Wang, Y.; Fuegen, C.; Kumar, A.; Liu, B.; Bengio, Y. Towards End-to-End Spoken Language Understanding. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 5754–5758.
10. 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.
11. De Cordoba, R.; Ferreiros, J.; San-Segundo, R.; Macias-Guarasa, J.; Montero, J.M.; Fernandez, F.; D'Haro, L.F.; Pardo, J.M. Air Traffic Control Speech Recognition System Cross-Task and Speaker Adaptation. *IEEE Aerosp. Electron. Syst. Mag.* **2006**, *21*, 12–17. [CrossRef]
12. Ferreiros, J.; Pardo, J.M.; de Córdoba, R.; Macias-Guarasa, J.; Montero, J.M.; Fernández, F.; Sama, V.; D'Haro, L.F.; González, G. A Speech Interface for Air Traffic Control Terminals. *Aerosp. Sci. Technol.* **2012**, *21*, 7–15. [CrossRef]
13. Helmke, H.; Ohneiser, O.; Muhlhausen, T.; Wies, M. Reducing Controller Workload with Automatic Speech Recognition. In Proceedings of the 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25–29 September 2016; pp. 1–10.
14. Gurluk, H.; Helmke, H.; Wies, M.; Ehr, H.; Kleinert, M.; Muhlhausen, T.; Muth, K.; Ohneiser, O. Assistant Based Speech Recognition—Another Pair of Eyes for the Arrival Manager. In Proceedings of the 2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC), Prague, Czech Republic, 13–18 September 2015; pp. 3B6-1–3B6-14.
15. Lin, Y.; Li, L.; Jing, H.; Ran, B.; Sun, D. Automated Traffic Incident Detection with a Smaller Dataset Based on Generative Adversarial Networks. *Accid. Anal. Prev.* **2020**, *144*, 105628. [CrossRef]
16. Lin, Y.; Li, Q.; Yang, B.; Yan, Z.; Tan, H.; Chen, Z. Improving Speech Recognition Models with Small Samples for Air Traffic Control Systems. *Neurocomputing* **2021**, *1*, 1–21.
17. Panayotov, V.; Chen, G.; Povey, D.; Khudanpur, S. Librispeech: An ASR Corpus Based on Public Domain Audio Books. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 19–24 April 2015; pp. 5206–5210.
18. Hernandez, F.; Nguyen, V.; Ghannay, S.; Tomashenko, N.; Estève, Y. TED-LIUM 3: Twice as Much Data and Corpus Repartition for Experiments on Speaker Adaptation. In Proceedings of the 20th International Conference on Speech and Computer, Leipzig, Germany, 18–22 September 2018; pp. 198–208.
19. Godfrey, J.; Holliman, E. Switchboard-1 Release 2. Available online: https://Catalog.Ldc.Upenn.Edu/LDC97S62 (accessed on 25 January 2021).
20. Wang, D.; Zhang, X. THCHS-30: A Free Chinese Speech Corpus. *arXiv* **2015**, arXiv:1512.01882.

21. Bu, H.; Du, J.; Na, X.; Wu, B.; Zheng, H. AISHELL-1: An Open-Source Mandarin Speech Corpus and a Speech Recognition Baseline. In Proceedings of the 2017 20th Conference of the Oriental Chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment (O-COCOSDA), Seoul, Korea, 13 November 2017; pp. 1–5.
22. Du, J.; Na, X.; Liu, X.; Bu, H. AISHELL-2: Transforming Mandarin ASR Research into Industrial Scale. *arXiv* **2018**, arXiv:1808.10583.
23. Yang, B.; Tan, X.; Chen, Z.; Wang, B.; Ruan, M.; Li, D.; Yang, Z.; Wu, X.; Lin, Y. ATCSpeech: A Multilingual Pilot-Controller Speech Corpus from Real Air Traffic Control Environment. In Proceedings of the Interspeech 2020, ISCA, Shanghai, China, 25–29 October 2020; pp. 399–403.
24. Hofbauer, K.; Petrik, S.; Hering, H. The ATCOSIM Corpus of Non-Prompted Clean Air Traffic Control Speech. In Proceedings of the 6th International Conference on Language Resources and Evaluation—LREC, Marrakech, Morocco, 26 May–1 June 2008.
25. Godfrey, J. Air Traffic Control Complete. Available online: https://catalog.ldc.upenn.edu/ldc94s14a (accessed on 25 January 2021).
26. 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. In Proceedings of the Interspeech 2019, Graz, Austria, 15–19 September 2019; pp. 2993–2997.
27. Lin, Y.; Guo, D.; Zhang, J.; Chen, Z.; Yang, B. A Unified Framework for Multilingual Speech Recognition in Air Traffic Control Systems. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, 1–13. [CrossRef]
28. ICAO. *Manual on the Implementation of ICAO Language Proficiency Requirements*, 2nd ed.; International Civil Aviation Organization: Montréal, QC, Canada, 2010; ISBN 9789292315498.
29. Benzeghiba, M.; De Mori, R.; Deroo, O.; Dupont, S.; Erbes, T.; Jouvet, D.; Fissore, L.; Laface, P.; Mertins, A.; Ris, C.; et al. Automatic Speech Recognition and Speech Variability: A Review. *Speech Commun.* **2007**, *49*, 763–786. [CrossRef]
30. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal. Process. Mag.* **2012**. [CrossRef]
31. Abe, A.; Yamamoto, K.; Nakagawa, S. Robust Speech Recognition Using DNN-HMM Acoustic Model Combining Noise-Aware Training with Spectral Subtraction. In Proceedings of the Annual Conference of the International Speech Communication Association—INTERSPEECH, Dresden, Germany, 6–10 September 2015; pp. 2849–2853.
32. Graves, A.; Fernández, S.; Gomez, F.; Schmidhuber, J. Connectionist Temporal Classification. In Proceedings of the 23rd International Conference on Machine Learning—ICML '06, Pittsburgh, PA, USA, 25–29 June 2006; ACM Press: New York, NY, USA, 2006; Volume 32, pp. 369–376.
33. Graves, A.; Mohamed, A.; Hinton, G. Speech Recognition with Deep Recurrent Neural Networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.
34. Graves, A.; Jaitly, N. Towards End-to-End Speech Recognition with Recurrent Neural Networks. In Proceedings of the 31st International Conference on Machine Learning—ICML, Beijing, China, 21–26 June 2014.
35. Amodei, D.; Anubhai, R.; Battenberg, E.; Case, C.; Casper, J.; Catanzaro, B.; Chen, J.; Chrzanowski, M.; Coates, A.; Diamos, G.; et al. Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. In Proceedings of the 33rd International Conference on Machine Learning—ICML 2016, New York, NY, USA, 19–24 June 2016.
36. Li, J.; Lavrukhin, V.; Ginsburg, B.; Leary, R.; Kuchaiev, O.; Cohen, J.M.; Nguyen, H.; Gadde, R.T. Jasper: An End-to-End Convolutional Neural Acoustic Model. In Proceedings of the Interspeech 2019—ISCA, Graz, Austria, 15–19 September 2019; pp. 71–75.
37. Sainath, T.N.; Vinyals, O.; Senior, A.; Sak, H. Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 19–24 April 2015; pp. 4580–4584.
38. Zeghidour, N.; Xu, Q.; Liptchinsky, V.; Usunier, N.; Synnaeve, G.; Collobert, R. Fully Convolutional Speech Recognition. *arXiv* **2018**, arXiv:1812.06864.
39. Chan, W.; Jaitly, N.; Le, Q.; Vinyals, O. Listen, Attend and Spell: A Neural Network for Large Vocabulary Conversational Speech Recognition. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 4960–4964.
40. Chiu, C.-C.; Sainath, T.N.; Wu, Y.; Prabhavalkar, R.; Nguyen, P.; Chen, Z.; Kannan, A.; Weiss, R.J.; Rao, K.; Gonina, E.; et al. State-of-the-Art Speech Recognition with Sequence-to-Sequence Models. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4774–4778.
41. Watanabe, S.; Hori, T.; Kim, S.; Hershey, J.R.; Hayashi, T. Hybrid CTC/Attention Architecture for End-to-End Speech Recognition. *IEEE J. Sel. Top. Signal. Process.* **2017**, *11*, 1240–1253. [CrossRef]
42. Bahdanau, D.; Chorowski, J.; Serdyuk, D.; Brakel, P.; Bengio, Y. End-to-End Attention-Based Large Vocabulary Speech Recognition. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 4945–4949.
43. Chorowski, J.; Bahdanau, D.; Serdyuk, D.; Cho, K.; Bengio, Y. Attention-Based Models for Speech Recognition. *arXiv* **2015**, arXiv:1506.07503.
44. Pham, N.-Q.; Nguyen, T.-S.; Niehues, J.; Müller, M.; Waibel, A. Very Deep Self-Attention Networks for End-to-End Speech Recognition. In Proceedings of the Interspeech 2019—ISCA, Graz, Austria, 15–19 September 2019; pp. 66–70.
45. Lian, Z.; Li, Y.; Tao, J.; Huang, J. Improving Speech Emotion Recognition via Transformer-Based Predictive Coding through Transfer Learning. *arXiv* **2018**, arXiv:1811.07691.

46. Karita, S.; Soplin, N.E.Y.; Watanabe, S.; Delcroix, M.; Ogawa, A.; Nakatani, T. Improving Transformer-Based End-to-End Speech Recognition with Connectionist Temporal Classification and Language Model Integration. In Proceedings of the Annual Conference of the International Speech Communication Association—INTERSPEECH, Graz, Austria, 15–19 September 2019.
47. Zhou, S.; Dong, L.; Xu, S.; Xu, B. Syllable-Based Sequence-to-Sequence Speech Recognition with the Transformer in Mandarin Chinese. In Proceedings of the Annual Conference of the International Speech Communication Association—INTERSPEECH, Hyderabad, India, 2–6 September 2018.
48. Jiang, D.; Lei, X.; Li, W.; Luo, N.; Hu, Y.; Zou, W.; Li, X. Improving Transformer-Based Speech Recognition Using Unsupervised Pre-Training. *arXiv* **2019**, arXiv:1910.09932.
49. Toshniwal, S.; Sainath, T.N.; Weiss, R.J.; Li, B.; Moreno, P.; Weinstein, E.; Rao, K. Multilingual Speech Recognition with a Single End-to-End Model. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 4904–4908.
50. Li, B.; Zhang, Y.; Sainath, T.; Wu, Y.; Chan, W. Bytes Are All You Need: End-to-End Multilingual Speech Recognition and Synthesis with Bytes. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 5621–5625.
51. Hu, K.; Bruguier, A.; Sainath, T.N.; Prabhavalkar, R.; Pundak, G. Phoneme-Based Contextualization for Cross-Lingual Speech Recognition in End-to-End Models. In Proceedings of the Interspeech 2019—ISCA, Graz, Austria, 15–19 September 2019; pp. 2155–2159.
52. Zhang, S.; Liu, Y.; Lei, M.; Ma, B.; Xie, L. Towards Language-Universal Mandarin-English Speech Recognition. In Proceedings of the Interspeech 2019—ISCA, Graz, Austria, 15–19 September 2019; pp. 2170–2174.
53. Zeng, Z.; Khassanov, Y.; Pham, V.T.; Xu, H.; Chng, E.S.; Li, H. On the End-to-End Solution to Mandarin-English Code-Switching Speech Recognition. In Proceedings of the Interspeech 2019—ISCA, Graz, Austria, 15–19 September 2019; pp. 2165–2169.
54. Lin, Y.; Yang, B.; Li, L.; Guo, D.; Zhang, J.; Chen, H.; Zhang, Y. ATCSpeechNet: A Multilingual End-to-End Speech Recognition Framework for Air Traffic Control Systems. *arXiv* **2021**, arXiv:2102.08535.
55. Srinivasamurthy, A.; Motlicek, P.; Singh, M.; Oualil, Y.; Kleinert, M.; Ehr, H.; Helmke, H. Iterative Learning of Speech Recognition Models for Air Traffic Control. In Proceedings of the Annual Conference of the International Speech Communication Association—INTERSPEECH, Hyderabad, India, 2–6 September 2018.
56. Oualil, Y.; Klakow, D.; Szaszak, G.; Srinivasamurthy, A.; Helmke, H.; Motlicek, P. A Context-Aware Speech Recognition and Understanding System for Air Traffic Control Domain. In Proceedings of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)—IEEE, Sentosa, Singapore, 14–18 December 2017; pp. 404–408.
57. Srinivasamurthy, A.; Motlicek, P.; Himawan, I.; Szaszák, G.; Oualil, Y.; Helmke, H. Semi-Supervised Learning with Semantic Knowledge Extraction for Improved Speech Recognition in Air Traffic Control. In Proceedings of the Interspeech 2017—ISCA, Stockholm, Sweden, 20–24 August 2017; pp. 2406–2410.
58. 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)—IEEE, Jeju, Korea, 13–16 October 2016; pp. 1309–1314.
59. Cordero, J.M.; Dorado, M.; de Pablo, J.M. Automated Speech Recognition in ATC Environment. In Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems, London, UK, 29–31 May 2012; pp. 46–53.
60. Delpech, E.; Laignelet, M.; Pimm, C.; Raynal, C.; Trzos, M.; Arnold, A.; Pronto, D. A Real-Life, French-Accented Corpus of Air Traffic Control Communications. In Proceedings of the LREC 2018—11th International Conference on Language Resources and Evaluation, Miyazaki, Japan, 7–12 May 2018.
61. Zuluaga-Gomez, J.; Motlicek, P.; Zhan, Q.; Veselý, K.; Braun, R. Automatic Speech Recognition Benchmark for Air-Traffic Communications. In Proceedings of the Interspeech 2020—ISCA, Shanghai, China, 25–29 October 2020; pp. 2297–2301.
62. Hou, N.; Tian, X.; Chng, E.S.; Ma, B.; Li, H. Improving Air Traffic Control Speech Intelligibility by Reducing Speaking Rate Effectively. In Proceedings of the 2017 International Conference on Asian Language Processing (IALP)—IEEE, Singapore, 5–7 December 2017; pp. 197–200.
63. Šmídl, L.; Švec, J.; Tihelka, D.; Matoušek, J.; Romportl, J.; Ircing, P. Air Traffic Control Communication (ATCC) Speech Corpora and Their Use for ASR and TTS Development. *Lang. Resour. Eval.* **2019**, *53*, 449–464. [CrossRef]
64. 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. *Proceedings* **2020**, *59*, 14. [CrossRef]
65. Lin, Y.; Zhang, J.; Liu, H. Deep Learning Based Short-Term Air Traffic Flow Prediction Considering Temporal–Spatial Correlation. *Aerosp. Sci. Technol.* **2019**, *93*, 105113. [CrossRef]
66. Liu, H.; Lin, Y.; Chen, Z.; Guo, D.; Zhang, J.; Jing, H. Research on the Air Traffic Flow Prediction Using a Deep Learning Approach. *IEEE Access* **2019**, *7*, 148019–148030. [CrossRef]
67. Li, L.; Lin, Y.; Du, B.; Yang, F.; Ran, B. Real-Time Traffic Incident Detection Based on a Hybrid Deep Learning Model. *Transp. A Transp. Sci.* **2020**, 1–21. [CrossRef]

68. Dumais, S.; Platt, J.; Heckerman, D.; Sahami, M. Inductive Learning Algorithms and Representations for Text Categorization. In Proceedings of the Seventh International Conference on Information and Knowledge Management—CIKM '98, Bethesda, MD, USA, 3–7 November 1998; ACM Press: New York, NY, USA, 1998; pp. 148–155.
69. Collins, M. Discriminative Training Methods for Hidden Markov Models. In Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing—EMNLP '02, Philadelphia, PA, USA, 6–7 July 2002; Association for Computational Linguistics: Morristown, NJ, USA, 2002; Volume 10, pp. 1–8.
70. Yu, H.-F.; Huang, F.-L.; Lin, C.-J. Dual Coordinate Descent Methods for Logistic Regression and Maximum Entropy Models. *Mach. Learn.* **2011**, *85*, 41–75. [CrossRef]
71. Malouf, R. A Comparison of Algorithms for Maximum Entropy Parameter Estimation. In Proceedings of the 6th Conference on Natural Language Learning—COLING-02, Taipei, Taiwan, 26–30 August 2002; Association for Computational Linguistics: Morristown, NJ, USA, 2002; Volume 20, pp. 1–7.
72. Raymond, C.; Riccardi, G. Generative and Discriminative Algorithms for Spoken Language Understanding. In Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, Antwerp, Belgium, 27–31 August 2007; Volume 1, pp. 413–416.
73. Haffner, P.; Tur, G.; Wright, J.H. Optimizing SVMs for Complex Call Classification. In Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing—ICASSP '03, Hong Kong, 6–10 April 2003; Volume 1, pp. I-632–I-635.
74. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.
75. Xu, P.; Sarikaya, R. Convolutional Neural Network Based Triangular CRF for Joint Intent Detection and Slot Filling. In Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding—IEEE, Olomouc, Czech Republic, 8–12 December 2013; pp. 78–83.
76. McCallum, A.; Freitag, D.; Pereira, F. Maximum Entropy Markov Models for Information Extraction and Segmentation. In Proceedings of the Seventeenth International Conference on Machine Learning, Stanford, CA, USA, 29 June–2 July 2000; pp. 591–598.
77. Mesnil, G.; Dauphin, Y.; Yao, K.; Bengio, Y.; Deng, L.; Hakkani-Tur, D.; He, X.; Heck, L.; Tur, G.; Yu, D.; et al. Using Recurrent Neural Networks for Slot Filling in Spoken Language Understanding. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 530–539. [CrossRef]
78. Yao, K.; Peng, B.; Zhang, Y.; Yu, D.; Zweig, G.; Shi, Y. Spoken Language Understanding Using Long Short-Term Memory Neural Networks. In Proceedings of the 2014 IEEE Spoken Language Technology Workshop (SLT), South Lake Tahoe, NV, USA, 7–10 December 2014; pp. 189–194.
79. Liu, B.; Lane, I. Joint Online Spoken Language Understanding and Language Modeling with Recurrent Neural Networks. In Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Los Angeles, CA, USA, 13–15 September 2016; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 22–30.
80. Guo, D.Z.; Tur, G.; Yih, W.T.; Zweig, G. Joint Semantic Utterance Classification and Slot Filling with Recursive Neural Networks. In Proceedings of the 2014 IEEE Workshop on Spoken Language Technology, SLT, South Lake Tahoe, NV, USA, 7–10 December 2014; pp. 554–559.
81. Xiaodong, Z.; Houfeng, W. A Joint Model of Intent Determination and Slot Filling for Spoken Language Understanding. In Proceedings of the IJCAI International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; Brewka, G., Ed.; AAAI Press: New York, NY, USA, 2016; pp. 2993–2999.
82. Li, C.; Li, L.; Qi, J. A Self-Attentive Model with Gate Mechanism for Spoken Language Understanding. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing—EMNLP 2018, Brussels, Belgium, 31 October–4 November 2018.
83. Liu, B.; Lane, I. Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling. In Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, San Francisco, CA, USA, 8–12 September 2016; pp. 685–689.
84. Radfar, M.; Mouchtaris, A.; Kunzmann, S. End-to-End Neural Transformer Based Spoken Language Understanding. In Proceedings of the Interspeech 2020—ISCA, Shanghai, China, 25–29 October 2020; pp. 866–870.
85. Huang, C.-W.; Chen, Y.-N. Adapting Pretrained Transformer to Lattices for Spoken Language Understanding. In Proceedings of the 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Sentosa, Singapore, 14–18 December 2019; pp. 845–852.
86. Nguyen, V.N.; Holone, H. N-Best List Re-Ranking Using Semantic Relatedness and Syntactic Score: An Approach for Improving Speech Recognition Accuracy in Air Traffic Control. In Proceedings of the 2016 16th International Conference on Control, Automation and Systems (ICCAS)—IEEE, Gyeongju, Korea, 16–19 October 2016; pp. 1315–1319.
87. Oualil, Y.; Schulder, M.; Helmke, H.; Schmidt, A.; Klakow, D. Real-Time Integration of Dynamic Context Information for Improving Automatic Speech Recognition. In Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, Dresden, Germany, 6–10 September 2015; ISCA: Dresden, Germany, 2015; pp. 2107–2111.
88. Kersta, L.G. Voiceprint Identification. *Nature* **1962**, *196*, 1253–1257. [CrossRef]
89. Pruzansky, S. Pattern-Matching Procedure for Automatic Talker Recognition. *J. Acoust. Soc. Am.* **1963**, *35*, 354–358. [CrossRef]
90. Furui, S. Cepstral Analysis Technique for Automatic Speaker Verification. *IEEE Trans. Acoust.* **1981**, *29*, 254–272. [CrossRef]

91. Soong, F.; Rosenberg, A.; Rabiner, L.; Juang, B. A Vector Quantization Approach to Speaker Recognition. In Proceedings of the ICASSP '85, IEEE International Conference on Acoustics, Speech, and Signal Processing, Tampa, FL, USA, 26–29 March 1985; Volume 10, pp. 387–390.
92. Reynolds, D.A.; Rose, R.C. Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. *IEEE Trans. Speech Audio Process.* **1995**, *3*, 72–83. [CrossRef]
93. Reynolds, D.A.; Quatieri, T.F.; Dunn, R.B. Speaker Verification Using Adapted Gaussian Mixture Models. *Digit. Signal. Process.* **2000**, *10*, 19–41. [CrossRef]
94. Campbell, W.M.; Sturim, D.; Reynolds, D.A. Support Vector Machines Using GMM Supervectors for Speaker Verification. *IEEE Signal. Process. Lett.* **2006**, *13*, 308–311. [CrossRef]
95. Kenny, P.; Ouellet, P.; Dehak, N.; Gupta, V.; Dumouchel, P. A Study of Interspeaker Variability in Speaker Verification. *IEEE Trans. Audio. Speech. Lang. Process.* **2008**, *16*, 980–988. [CrossRef]
96. Dehak, N.; Kenny, P.J.; Dehak, R.; Dumouchel, P.; Ouellet, P. Front-End Factor Analysis for Speaker Verification. *IEEE Trans. Audio, Speech Lang. Process.* **2011**. [CrossRef]
97. Variani, E.; Lei, X.; McDermott, E.; Moreno, I.L.; Gonzalez-Dominguez, J. Deep Neural Networks for Small Footprint Text-Dependent Speaker Verification. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 4–9 May 2014.
98. Li, C.; Ma, X.; Jiang, B.; Li, X.; Zhang, X.; Liu, X.; Cao, Y.; Kannan, A.; Zhu, Z. Deep Speaker: An End-to-End Neural Speaker Embedding System. *arXiv* **2017**, arXiv:1705.02304.
99. Snyder, D.; Garcia-Romero, D.; Povey, D.; Khudanpur, S. Deep Neural Network Embeddings for Text-Independent Speaker Verification. In Proceedings of the Interspeech 2017—ISCA, Stockholm, Sweden, 20–24 August 2017; pp. 999–1003.
100. Snyder, D.; Garcia-Romero, D.; Sell, G.; Povey, D.; Khudanpur, S. X-Vectors: Robust DNN Embeddings for Speaker Recognition. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 5329–5333.
101. Liu, Y.; Qian, Y.; Chen, N.; Fu, T.; Zhang, Y.; Yu, K. Deep Feature for Text-Dependent Speaker Verification. *Speech Commun.* **2015**. [CrossRef]
102. Chen, N.; Qian, Y.; Yu, K. Multi-Task Learning for Text-Dependent Speaker Verification. In Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, Dresden, Germany, 6–10 September 2015; pp. 185–189.
103. Ravanelli, M.; Bengio, Y. Speaker Recognition from Raw Waveform with SincNet. In Proceedings of the 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, 18–21 December 2018; pp. 1021–1028.
104. Kopald, H.; Chen, S. Design and Evaluation of the Closed Runway Operation Prevention Device. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2014**, *58*, 82–86. [CrossRef]
105. Lin, Y.; Zhang, J.; Liu, H. An Algorithm for Trajectory Prediction of Flight Plan Based on Relative Motion between Positions. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 905–916. [CrossRef]
106. Lin, Y.; Yang, B.; Zhang, J.; Liu, H. Approach for 4-D Trajectory Management Based on HMM and Trajectory Similarity. *J. Mar. Sci. Technol.* **2019**, *27*, 246–256. [CrossRef]
107. Chen, Z.; Guo, D.; Lin, Y. A Deep Gaussian Process-Based Flight Trajectory Prediction Approach and Its Application on Conflict Detection. *Algorithms* **2020**, *13*, 293. [CrossRef]
108. Yepes, J.L.; Hwang, I.; Rotea, M. New Algorithms for Aircraft Intent Inference and Trajectory Prediction. *J. Guid. Control. Dyn.* **2007**, *30*, 370–382. [CrossRef]
109. Cordero, J.M.; Rodríguez, N.; Miguel, J.; Pablo, D.; Dorado, M. Automated Speech Recognition in Controller Communications Applied to Workload Measurement. In Proceedings of the Third SESAR Innovation Days, Stockholm, Sweden, 26–28 November 2013; pp. 1–8.
110. ICAO. *Manual on Air Traffic Controller Competency-Based Training and Assessment*, 1st ed.; International Civil Aviation Organization: Montréal, QC, Canada, 2016.