Advances in Transportation, Vol. 60, 2022, 584–592
https://doi.org/10.54941/ahfe1002493

Early Callsign Highlighting using Automatic Speech Recognition to Reduce Air Traffic Controller Workload

Shruthi Shetty, Hartmut Helmke, Matthias Kleinert, and Oliver Ohneiser

Institute of Flight Guidance, German Aerospace Center (DLR), Lilienthalplatz 7, 38108 Braunschweig, Germany

ABSTRACT

The primary task of an air traffic controller (ATCo) is to issue instructions to pilots. However, the first verbal communication contact is often initiated by the pilot. Hence, the ATCo needs to search for the aircraft radar label that corresponds to the callsign uttered by the pilot. Therefore, it would be useful to have a controller assistance system, which recognizes and highlights the spoken callsign in the ATCo display as early as possible, directly from the speech data. Therefore, we propose to use an automatic speech recognition (ASR) system to first obtain the speech-to-text transcription, followed by extracting the spoken callsign from the transcription. As a high performance in callsign recognition is required, we use surveillance data, which significantly reduces callsign recognition error rates. When using ASR transcriptions for ATCo utterances of Isavia data (HAAWAII project[1]) by SESAR Joint Undertaking (Grant Numbers 874464 resp. 884287)), we initially obtain a callsign recognition error rate of 6.2%, which improves to 2.8% when surveillance data information is used.

Keywords: Callsign highlighting, Automatic speech recognition, Air traffic controller workload

INTRODUCTION

Voice communication over radio is a widely used mode of air traffic control (ATC) communication (Helmke et al. 2021). ATC utterances contain communication between ATCos and aircraft pilots. ATCos are responsible for safe and efficient movement of aircraft in the air and on ground. They issue voice instructions to pilots who are in the area that the ATCos are responsible for. These instructions include altitude and speed changes, headings to certain waypoints and geographical coordinates, etc. Many of these instructions are time critical and need immediate actions to be taken by pilots. The first contact between ATCo and pilot is often initiated by the pilot who reports the current status of the aircraft while entering the area controlled by an ATCo. The ATCo then has to search the radar screen for the callsign corresponding to the aircraft. This task increases ATCo workload, especially during periods

[1] HAWWAII project and PJ.10-96 (Wave-2) are partly funded by SESAR Joint Undertaking (Grant Numbers 874464 resp. 884287).

of high traffic when ATCos communicate with many pilots simultaneously, which may cause reduced situational awareness.

ASR could be used as a viable solution to address this problem. Highlighting relevant callsigns on the radar display could be useful to an ATCo, since it relieves the ATCo from having to search for the aircraft on the radar screen. Recognizing and highlighting callsigns on the radar screen as early as possible helps in reducing ATCo workload and enables ATCos to focus on issuing instructions to aircraft more efficiently. For this, ASR can be used to obtain the speech-to-text transcription of an utterance as a first step. In the second step, automatic language understanding is used to automatically extract relevant callsigns from the transcription. This paper focuses on recognizing and understanding relevant callsigns as early as possible from ATC utterances, even before an utterance is completely spoken. In addition to finding aircraft corresponding to initial pilot utterances on a radar display, early callsign extraction in ATCo utterances also allows ATCos to be completely sure that they are communicating the right instructions to the right aircraft, provided the callsigns are correctly recognized. This requires callsign recognition to have a high recognition rate with a very low error rate.

However, recognizing callsigns with a very low error rate is challenging, because a given callsign can be spoken in different ways. A callsign usually consists of a 3-letter airline designator followed by a sequence of letters and digits. An ATCo or pilot must ideally pronounce a callsign in its full form using the corresponding keyword sequence for the designator and ICAO (International Civil Aviation Organization) phonetic alphabet. However, in real life ATCos and pilots deviate from the standard phraseology and use short forms. For example, "BAW502P", which should be ideally pronounced as "speed bird five zero two papa" could also be pronounced as "speed bird five", "speed bird five zero two" or without the designator as "five zero two papa". These deviations are more often found in pilot utterances as compared to ATCo utterances. Moreover, transcripts obtained from ASR systems are not 100% correct and contain word errors. This leads to misrecognitions in callsign designators, ICAO alphabet and digits in word sequences containing callsigns. For example, "speed bird five zero two papa" could be recognized as "speed bird five zulu papa", making it challenging to correctly recognize the correct callsigns.

RELATED WORK

In this section, we discuss some of the related work carried out in the area of callsign recognition. Lithuanian tower utterances have been considered for command extraction by DLR (Ohneiser et al. 2021). This automatic extraction focusses on recognizing the semantic meanings of ATCo utterances. A command is composed of various fields such as: callsign, command type, second type, value, unit, etc. Here, callsign recognition is carried out as part of command extraction.

(Nigmatulina et al. 2021) focusses on improving ASR callsign recognitions on word level. This work also uses surveillance information to boost words

Table 1. Examples of callsign annotations and their keyword sequences.

Callsign Annotation	Designator Code	Keyword Sequences
AHO372Q	AHO	air hamburg three seven two quebec, hamburg three seven two quebec, air hamburg seven two quebec, three seven two quebec
BAW515	BAW	speed bird five one five, speed bird one five, five one five, five fifteen
DLH47V	DLH	lufthansa four seven victor, hansa four seven victor, lufthansa seven victor, hansa seven victor, four seven victor
OK1AC	-	oscar kilo one alfa charlie, oscar alfa charlie

belonging to callsigns by using the list of callsigns that are in air at the given time.

Some applications of the HAAWAII project such as readback error detection also have a focus on callsign recognition. Readback error detection is a critical application for which recognition of correct callsigns in the ATCo-pilot communication is of utmost importance as discussed in (Helmke et al. 2021).

Another work (Lasheras et al. 2021) discusses recognition and highlighting of callsigns in ATC communication, where candidate words of the utterance are first classified as callsign, which are then compared with a list of possible callsigns. This work focusses on extracting those callsigns which are spoken using full forms.

(Chen et al. 2021) also works on processing voice utterances in ATC communication by first developing ASR models to get the speech-to-text transcription and using domain-specific parsing algorithms to understand aircraft callsigns, issued controller commands and advisories as well as pilot readbacks.

ASR-BASED AUTOMATIC CALLSIGN RECOGNITION

Automatic callsign recognition from a voice utterance consists of two steps - first obtaining the transcription followed by annotation of the spoken callsign. Transcription refers to the word-by-word representation of the speech data. Annotation refers to the semantic interpretation of the transcription, transforming a sequence of words to a sequence of ATC concepts (Helmke et al. 2021). In this work, we only focus on recognizing and annotating relevant callsigns which are uttered in ATCo-pilot communication.

In a callsign, the airline designator is uttered using a keyword sequence and is annotated using their corresponding code composed of a sequence of three letters as shown in Table 1. The sequence of letters and digits in a callsign is pronounced using the ICAO phonetic alphabet. Some examples of callsigns and their keyword sequences are illustrated in Table 1. Here, the airline designators are shown in black and the digits and letters are shown in blue. The airline designator is optional and small aircraft like OK1AC usually do not have a designator in their callsign (Table 1).

Once the ASR starts outputting transcription of the spoken utterance, our system uses it as input to extract the spoken callsign. This means that we do not wait until the end of utterance to extract callsigns, because our goal is to recognize callsigns in an utterance as early as possible. Surveillance information is also used in callsign extraction as it significantly narrows down the number of possible callsigns. Radar data contains information regarding the list of callsigns corresponding to aircraft currently in air. A callsign which is captured by the radar device is said to be in context.

Callsign extraction is carried out as part of command extraction in four steps. Since we focus on recognizing callsigns in this paper, we shall only discuss about callsign extraction. The four steps are:

i. Recognizing callsigns by exact match by using surveillance data
ii. Using word classification to verify extracted callsign from step 1
iii. Recognizing callsigns by Levenshtein distance (Levenshtein, 1966)
iv. Recognizing callsigns without using surveillance information

In the first step, we generate possible keyword sequences for callsigns which are in context and try to find an exact match in the recognized parts of an utterance. Words which are recognized as part of the callsign are classified as “csgn”.

The callsign extracted in step 1 (if any) is verified by looking into the word classification of the utterance. If a letter or digit, which is classified as “unknown” is found immediately next to a word which is classified as callsign (“csgn”), we discard the callsign extracted in step 1 and continue extraction for the same sequence of words in the next step.

The third step is to extract callsigns by computing Levenshtein distance. Here, we extract callsigns which are in context by comparing their keyword sequences with word sequences from utterances which are candidates for callsign. A Levenshtein distance of two or lesser is permitted under defined conditions. The best matching callsign with the least Levenshtein distance is extracted. For example, if “speed bird two alfa four” is said and there exists a BAW3A4 (speed bird three alfa four) in the surveillance data which is the closest match to “speed bird two alfa four”, then the BAW34A is assumed to be the correct callsign. This step helps in recognizing callsigns which were not previously extracted due to errors made by the ASR or were wrongly uttered in the ATC communication by accident. In addition, this step also helps in recognizing the callsign when “break break” is used in the utterance.

The final step is to extract callsigns without using surveillance data. This step is executed only when no callsign was extracted in the previous steps. Here, callsigns are extracted from candidate word sequences of unclassified parts of an utterance, with the assumption that the correct callsign is not in the surveillance data.

When no callsign is extracted in all steps of callsign extraction, we annotate using NO_CALLSIGN, but nothing is highlighted on the radar display.

Table 2. Dataset description[2].

	Area	#Utterances	#Commands	Annotations Gold [h]	WER of A.SR transcriptions	
					ATCo	Pilot
Isavia	Enroute	2930	5803	3.5	4.7%	8.8%
NATS	TMA South & LLAP	4115	7419	4	3.3%	6.3%
Fraport	Ground	3837	9411	3	4.8%	NA
ANS CR Ops	Approach	3040	6121	4.7	10.9%	NA
ANS CR Lab	Approach	4219	6904	4.5	8.2%	NA
ACG Ops	Approach	3092	4912	3.8	10.9%	NA
ACG Lab	Approach	3971	6626	4.5	17.3%	NA
Sol96 Ops	Approach	519	1107	0.5	9.5%	20.8%
Sol96 Lab	Approach	635	1111	0.9	5.7%	NA

EXPERIMENTAL SETUP

The datasets used to evaluate the performance of our callsign recognition were obtained from various Air Navigation Service Providers (ANSP) corresponding to Isavia's enroute airspace, NATS's London Terminal Manoeuvring Area (TMA) South and Heathrow approach sector (HAAWAII project), Fraport's ground traffic (STARFISH project), ANS CR's Prague and ACG's Vienna approach sector (MALORCA project), and Sol96's Ops and Lab room data from ACG for Vienna. Datasets include surveillance data and voice utterances from the above sources. For each test data described in Table 2, manual transcriptions and annotations are available (Helmke et al. 2021).

The quality of callsign extraction is evaluated by comparing the callsigns from automatically extracted commands to that of gold annotations. Gold annotations refer to the annotations which are manually verified and corrected by a human expert. The metrics used to evaluate the quality of the extracted callsigns are: callsign recognition rate (CaRecR), callsign error rate (CaErrR), and callsign rejection rate (CaRejR). Callsign recognition rate is defined as the number of correctly recognized callsigns divided by the total number of callsigns. Callsign error rate is the percentage of wrongly extracted callsigns, which include substitutions and extracting a callsign where no callsign exists (referred to as insertions). A callsign is said to be rejected if NO_CALLSIGN is extracted, but a valid callsign exists in the gold annotation (also referred to as deletions). Callsign rejection rate is the percentage of gold callsigns which are not extracted, i.e., the percentage of NO_CALLSIGN extractions (Kleinert et al. 2021). The metrics are illustrated in Table 3 and the complete table with an example can be found in (Kleinert et al. 2021).

[2] The WERs for all datasets, particularly ANS CR and ACG are much higher as compared to what was reported in (Helmke et al. 2020) because of some updates made to the transcription rules for callsign designators. The gold transcriptions were updated accordingly, but the automatic transcriptions were not modified in order to show the potential of the callsign extraction algorithm also on noise data.

Table 3. Metric definition (Kleinert et al. 2021).

Metric	Calculation
Callsign Recognition Rate (CaRecR)	CaRecR = #matches / #gold
Callsign Recognition Error Rate (CaErrR)	CaErrR = (#substitutions + #insertions) / #gold
Callsign Rejection Rate (CaRejR)	CaRejR = #deletions / #gold

Table 4. Callsign extraction results using different combinations of inputs.

Data	goldtrans+context		autotrans+context		goldtrans		autotrans	
	CaRecR	CaErrR	CaRecR	CaErrR	CaRecR	CaErrR	CaRecR	CaErrR
Isavia ATCo	97.9%	1.5%	96.3%	2.8%	87.2%	4.0%	85.3%	6.2%
Isavia Pilot	96.4%	2.1%	91.9%	4.5%	78.7%	5.0%	75.1%	8.3%
NATSATCo	99.4%	0.5%	98.2%	1.7%	86.0%	5.1%	83.3%	9.4%
NATS Pilot	98.4%	1.2%	94.9%	3.9%	81.7%	6.8%	74.2%	14.2%
Fraport ATCo	98.3%	1.5%	95.7%	3.3%	92.0%	4.8%	82.3%	11.7%
ANS CR Ops ATCo	99.7%	0.2%	98.6%	1.0%	94.7%	2.3%	94.7%	3.0%
ANS CR Lab ATCo	99.6%	0.1%	96.3%	2.1%	94.0%	2.4%	90.6%	3.7%
ACG Ops ATCo	98.2%	1.2%	94.8%	3.7%	81.5%	9.8%	74.2%	17.1%
ACG Lab ATCo	97.2%	1.0%	86.8%	7.7%	80.0%	6.5%	77.2%	10.0%
Sol96 Ops ATCo	93.5%	4.3%	85.3%	8.6%	55.2%	27.0%	60.4%	17.0%
Sol96 Ops Pilot	95.8%	0.7%	84.0%	3.5%	77.9%	6.3%	64.2%	8.4%
Sol96 Lab ATCo	99.8%	0.0%	95.3%	2.4%	82.4%	0.5%	63.0%	19.5%

RESULTS

This section presents the results of callsign extraction. With respect to run time, we are able to recognize the callsign within 20ms after a callsign is uttered, thereby making it feasible to be used with live data. The results presented do not include callsign rejection rates (CaRejR), which can be computed using (1 – CaRecR – CaErrR). Table 4 illustrates the results of callsign extraction carried out using different combinations of inputs for the above-mentioned datasets. The meaning of the columns is explained below:

- goldtrans+context: using gold transcriptions and context information
- autotrans+context: using ASR transcriptions and context information
- goldtrans: using gold transcriptions without using context information
- autotrans: using ASR transcriptions without context information

From Table 4, we see that we obtain the best extraction rates (CaRecR and CaErrR) when using gold transcriptions in the presence of context information. Using ASR or automatic transcriptions decreases CaRecR and increases CaErrR. Not using context information deteriorates callsign extraction rates for both gold and automatic transcriptions. This difference is found to be

Table 5. Callsign extraction results on applying various filters.

Data	Disable- LDAndNoCtxExtr		Enable-LD-DisableNoCtxExtr		Disable-LD-Enable-NoCtxExtr		Disable-ImprvFromFrstWords	
	CaRecR	CaErrR	CaRecR	CaErrR	CaRecR	CaErrR	CaRecR	CaErrR
Isavia ATCo	93.7%	1.3%	95.2%	1.8%	95.1%	3.4%	96.3%	2.8%
Isavia Pilot	87.3%	1.5%	90.3%	2.8%	88.8%	6.1%	91.9%	4.5%
NATS ATCo	93.5%	1.1%	97.2%	1.3%	94.4%	5.2%	98.1%	1.8%
NATS Pilot	87.4%	0.9%	93.9%	2.5%	88.7%	8.7%	94.7%	4.0%
Fraport ATCo	86.8%	0.6%	92.3%	2.2%	90.2%	8.4%	95.7%	3.4%
ANS CR Ops ATCo	96.6%	0.1%	98.2%	1.6%	97.0%	2.2%	98.2%	1.6%
ANS CR Lab ATCo	94.6%	1.8%	96.2%	1.9%	95.8%	2.1%	96.3%	2.1%
ACG Ops ATCo	89.2%	1.0%	93.9%	5.3%	89.6%	8.2%	93.8%	5.6%
ACG Lab ATCo	82.7%	4.7%	84.7%	6.5%	85.9%	7.3%	86.9%	7.8%

more significant for Isavia, NATS and ACG data sets as compared to others. Additionally, for datasets such as Isavia and NATS, where both ATCo and pilot utterances are available, this difference is more significant for pilot utterances. This could imply the presence of large phraseology deviations and short forms for callsigns in these datasets and particularly in pilot utterances. Furthermore, from Table 2 we see that the WER of pilot utterances is higher than that of ATCo utterances. Correspondingly, we observe that the callsign extraction rates of ATCo utterances is better than that of pilot utterances for most datasets. However, as opposed to other datasets, for Sol96 Ops ATCo data when context information is not used, the callsign extraction rates are better when automatic transcription is used as compared to gold transcription. This could suggest that sometimes the ASR is able to recognize certain words in the callsign that a human transcriber is not.

Table 5 illustrates the callsign extraction rates when one or more steps of the callsign extraction process (as described in Section ASR-BASED AUTOMATIC CALLSIGN RECOGNITION) are disabled. Here, we try to look at the effect of each step of callsign extraction. The results in Table 5 are all obtained from automatic transcriptions in the presence of context information. The meaning of the columns in the table are explained below:

- Disable-LDAndNoCtxExtr: Disable both callsign extraction by calculating Levenshtein distance and extracting callsign without context information
- Enable-LD-DisableNoCtxExtr: Enable Levenshtein distance calculation, but disabling callsign extraction without using context information
- Disable-LD-Enable-NoCtxExtr: Disable Levenshtein distance callsign extraction, but enabling extraction without context information
- Disable-ImprvFromFrstWords: Disable improving callsign extraction by looking into word classification

From Table 5 we see that each of the three steps improves both callsign recognition and error rates for all datasets and the best extraction rates are obtained when all three steps are enabled (Table 4). Out of the above four filters, the worst recognition rate is obtained in 'Disable-LDAndNoCtxExtr' when steps 3 and 4 of callsign extraction (see Section ASR-BASED AUTOMATIC CALLSIGN RECOGNITION) are disabled. The recognition rate improves in 'Disable-LD-Enable-NoCtxExtr' when callsign extraction without using context information is enabled i.e., when step 3 is disabled and step 4 enabled. However, the callsign error rate worsens in this case. Instead, if step 3 is enabled and step 4 disabled ('Enable-LD-DisableNoCtxExtr'), both callsign recognition and error rates further improve for all datasets. This shows that callsign extraction using Levenshtein distance contributes significantly in the extraction of correct callsigns.

CONCLUSION AND FUTURE WORK

Early callsign highlighting enables ATCos to spot callsigns easily on the radar screens as soon as they are said. But this requires callsign extraction to have very low error rates, in order to avoid adverse effects from highlighting a wrong callsign. In our work, we obtain callsign recognition rates above 95% and error rates below 2.5% for almost all datasets when using gold transcriptions. However, the real challenge is to recognize correct callsigns when using automatic transcriptions. Considering that the used automatic transcriptions have WERs between 2 to 17%, we obtain good recognition rates between 92 to 98% and error rates below 5% for all datasets except Vienna lab and Sol96 ops room data, where relatively high WERs are observed. This shows that we are able to recognize most callsigns. However, our goal for future work is to make our callsign extraction more robust and further reduce callsign error rates when working with automatic transcriptions. Applications such as readback error detection require callsigns to be extracted with a very high level of accuracy (Helmke et al. 2021). The data also suggests that it is worth to invest more time and effort in good recognition performance on word level. Callsign extraction performance highly correlates with error rates on word level.

REFERENCES

Chen, S., Kopald, H., Ma, W., Tarakan, R., Wie, Y.J. (2021) "Air Traffic Control Speech Recognition", Interspeech Satellite Workshop, Brno, Czech Republic.

Helmke, H., Kleinert, M., Ohneiser, O., Ehr, H., Shetty, S. (2020) "Machine Learning of Air Traffic Controller Command Extraction Models for Speech Recognition Applications", proceedings of the IEEE/AIAA 39th Digital Avionics Systems Conference (DASC), virtual event.

Helmke, H., Kleinert, M., Shetty, S., Ohneiser, O., Ehr, H., Arilíusson, H., Simiganoschi, T.S., Prasad, A., Motlicek, P., Veselý, K., Ondřej, K., Smrz, P., Harfmann, J., Windisch, C. (2021) "Readback Error Detection by Automatic Speech Recognition to Increase ATM Safety", proceedings of the Fourteenth USA-/Europe Air Traffic Management Research and Development Seminar ATM2021, virtual event.

Kleinert, M., Helmke, H., Shetty, S., Ohneiser, O., Ehr, H., Prasad, A., Motlicek, P., Harfmann, J. (2021) "Automated Interpretation of Air Traffic Control Communication: The Journey from Spoken Words to a Deeper Understanding of the Meaning", proceedings of the IEEE/AIAA 40th Digital Avionics Systems Conference (DASC), San Antonio, Texas, USA.

Lasheras, R.G., Fabio, A., Celorrio, F., Albarrán, J., Ceñal, N., Oliveira, C.P., Martín, C.B., Chaves, J., Fillal, M. (2021) "Flight call sign identification on a Controller Working Position", Interspeech Satellite Workshop, Brno, Czech Republic.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In: Soviet Physics – Doklady 10.8.

Nigmatulina, I., Braun, R., Zuluaga, J.P., Motlicek, P. (2021) "Improving callsign recognition with air-surveillance data in air-traffic communication", Interspeech Satellite Workshop, Brno, Czech Republic.

Ohneiser, O., Sarfjoo, S.S., Helmke, H., Shetty, S., Motlicek, P., Kleinert, M., Ehr, H., Murauskas, S. (2021) "Robust Command Recognition for Lithuanian Air Traffic Control Tower Utterances", Interspeech, Brno, Czech Republic.