Pilot-Controller Communication Problems and an Initial Exploration of Language-Engineering Technologies as a Potential Solution

Bettina Bajaj and Arnab Majumdar

Centre for Transport Studies
Imperial College London
London, SW7 2AZ, UK

ABSTRACT

This paper examines pilot-controller miscommunications and a potential solution to minimise these. Since there is no up-to-date literature on the current level of such problems, a small-scale study was conducted. This involved analysing 531 British accident/incident reports, 531 CHIRP (Confidential Human Factors Incident Reporting Programme) reports, and 21 four-week communication flight logs recorded by eight British pilots. The results indicate that miscommunications remain numerous and directly affect pilots and controllers in terms of workload and situation awareness. The results also show that the number of communication problems identified in the accident/incident and CHIRP reports differ to those reported by the pilots, who recorded higher numbers. This seems to indicate a more problematic communication situation than the analysis of communication problems in the CHIRP and accident/incidents reports suggests. We hence argue that since voice communications remain beset with problems and the planned expanded controller-pilot data link communications (CPDLC) will only resolve some of the problems, it is vital to reduce these issues by developing an additional communication system using language-engineering technologies in automatic speech recognition, machine translation, and terminology extraction. The paper outlines how such a system might work and ameliorate pilot and controller workload and situation awareness.

Keywords: Automatic Speech Recognition, Communication Awareness, Machine Translation, Mental Workload, Pilot-Controller Communications, Situation Awareness, Standard Phraseology, Terminology Extraction

1. INTRODUCTION

Despite considerable research into communication problems between pilots and controllers, conducted from a variety of angles, such problems continue to exist. While Barshi and Farris have recently stated that "misunderstandings occur with an alarming frequency" (2011: 15), the data on which this quote is based stem from research published almost 35 years ago. With no current studies available that analyse the current levels of communication issues, this paper therefore presents a small-scale study consisting of three independent data analyses with the aim of determining the extent to which such issues still present a problem.

Following the review of related work, the subsequent three chapters will present and discuss this study. Chapter 6 then explores how an additional communication system based on language-engineering technologies may work and reduce communication issues while having a positive effect on workload and situational awareness.

2. RELATED WORK

Since the early 1980s pilot-controller communication problems have been the subject of considerable research.[1] One of the earliest collection of research studies was compiled by Billings and Cheaney in 1981. In one of these studies, Billings and Reynard (1981) analysed a total of 12,373 Aviation Safety Reporting System (ASRS) reports, using a typological approach, as part of which they classified information transfer issues into several groups. Billings and Reynard's research is of particular relevance to the present study as they quantified the number of miscommunications between pilots and controllers and their results are still quoted in current research. According to them, over 73.3% of the reports have revealed information transfer problems (1981: 11). They argue that because so many voice communications are impeded, verbal communication is not the best medium for effective information transfer (1981: 12). It is interesting to note that as far back as 1981, Billings and Reynard proposed that one way to improve such issues would be to use data link methods, although they surmised that these may cause different issues, e.g. misread numbers on displays (1981: 12-13). With regard to the current state of pilot-controller communication, the remarks with which Billings and Cheaney conclude their overall report are worth highlighting since they are still valid, as will be shown later in this chapter.

> These and previous studies lead us to conclude that there is a real and present need for better information transfer […]. […]. We conclude that there is insufficient awareness of the pervasive nature of the information transfer problem in its various manifestations, and that this lack of awareness may be in part responsible for nonstandard and inadequate communications practices on the part of both controllers and pilots. (1981: 92-93)

Subsequently, numerous studies have applied similar typological approaches (e.g. Monan 1983, Morrison and Wright 1989), although the objects of these investigations differed. For instance, ATC audio and/or video tapes were also analysed (e.g. Cardosi and Boole 1991) and human-in-the-loop simulations conducted (e.g. Kanki and Foushee 1989). A systematic review of these is provided in Prinzo and Britton's (1993) comprehensive overview of pilot-controller communication literature, which also lists studies based on other approaches, such as acoustical and cognitive-psycholinguistic approaches. By citing the final report of the *Work Group of Human Factors Relating to Controller and Pilot Errors* (1992), Prinzo and Britton (1993: 4) highlight that all these studies are largely descriptive without providing systematic research into the causes of communication problems. The literature does, however, record various studies which have not only done this, but have also looked at the causes from various perspectives. For instance, Goguen, Linde and Murphy (1985) have taken a communicative approach, whereas Cushing (1994) sees generic communication problems as the cause for pilot-controller miscommunications. More recently, a cognitive-linguistic approach was applied by Barshi and Farris (2013). A number of linguistic studies have led to vital recommendations for improvements, e.g. suggestions have been made regarding the length and complexity of ATC messages and phraseology wordings. Other studies have put forward new methods and tools for testing and training in ATC communications (e.g. Elliot 1997, Alderson 2011). Non-linguistic studies have addressed such issues as frequency congestion, and noise reduction, and, as suggested by Billings and Cheaney (1981), have examined alternative methods of conveying ATC messages using Controller Pilot Data Link Communications (CPDLC), e.g. Schneider, Healy, Barshi and Kole (2011).

Since Billings and Reynard's research, there appear to have been no studies that examine whether communication problems still amount to 73.3%. Barshi and Farris (2013), for instance, quote this figure without questioning its validity, and in other recent literature we can only get a vague idea. For example, in a study by EUROCONTROL[2] (2006), 535 communication occurrences (reported between Oct. 2004 and Mar. 2005) and 344 pilot and controller surveys were analysed. While the aim of this study was not to determine an overall percentage of communication issues, the fact that 535 occurrences were reported within six months nevertheless appears to be indicative of a problematic situation. A more recent study by the International Air Transport Association, the International Federation of Air Line Pilots' Associations and the International Federation of Air Traffic Controllers' Associations (IATA, IFALPA and IFATCA 2011) has examined the non-use of standard phraseology and related language issues using questionnaires. The results show, for example, that 44% of the 2070 participating pilots encounter non-standard phrases *at least once per flight* (2011: 13) and 52% of the 568 controllers report that they come across these *at least daily* (2011: 45). Other results of interest include the fact that 48% of the pilots operated to airports where ICAO standard phraseology is not used (2011: 17-18) and 11% of the pilots cited the use of local languages as a factor in decreasing situational awareness (2011: 28-30). We will see later that non-standard phrases and local languages are indeed a considerable problem for pilots.

---

[1] Communication problems due to medical reasons, e.g. noise-induced hearing loss, are not considered here.

[2] The results of many studies into pilot-controller miscommunications might not be fully applicable everywhere as most studies relate to North America. In contrast, there are few studies on this topic in Europe or other parts of the world.

3. METHOD

The methodology consists of three data analyses of (1) reports from CHIRP[3] (UK Confidential Human Factors Incident Reporting Programme), (2) aircraft accident/incident reports from the UK Air Accident Investigation Branch (AAIB), and (3) communication flight logs completed by British airline pilots. Only British data sources were used because, as far as could be ascertained, there are no studies that deal only with British pilot-controller communication issues (cf. also Footnote 1) and also to achieve homogenous sampling. The reason for using two types of report and flight logs is twofold. Firstly, we wanted to compare the number of issues mentioned in the reports and on the flights to those without such issues. Secondly, we were interested in the difference in number of miscommunications found in the two report types and wanted to know what communication issues occur during actual flights, the basis of which is the recent comment by IATA, IFALPA and IFATCA who noted that

> the use of non-standard phraseology, local accents, and the use of local languages in radio communication are infrequently reported as contributing factors to incidents and accidents. However, the vast majority of the survey's respondents stated that these factors were a concern and routinely caused misunderstanding. (2011: 7-8)

CHIRP reports
These reports are published in CHIRP feedback online publications dealing with air transport. Each publication contains confidentially submitted reports from controllers, pilots, engineers, relevant organisations, and cabin crew. CHIRP publications also include editorials and comments from readers in response to reports published in preceding publications. At the time of analysis, 86 air transport publications were available, of which the most recent 43 publications (from 2015 to 2000) were selected for analysis in this study. The 43 publications were then analysed with a view to identifying any reports relevant to this study. For reports to be considered relevant they had to (i) concern commercial flights and (ii) they had to be submitted by pilots or controllers, resulting in a total of 531 reports relevant to the study. All other reports were omitted, including reports in the form of comments since these refer to reports in preceding publications and would have led to duplication (or more) of reports. Each report was subsequently screened for citations of communication problems by means of an online search using ten search words and truncations[4]: *callsign*, *call sign*, *comm*, *congest, discipl*, *English*, *language*, *local*, *native*, *non-stand*, and *phrase*. The extracted problems were then allocated to the categories below, which were compiled specifically for the purposes of this study on the basis of the various communication issues found in the literature.

(1) *The use of local languages*
(2) *English language issues* (e.g. variations in English pronunciation/enunciation as well as in prosody, i.e. tempo, rhythm, pitch and loudness, by both native and non-native speakers)
(3) *Message issues* (e.g. incomplete, misunderstood, omitted transmissions, and so on)
(4) *Phraseology issues* (e.g. use of non-standard phrases; lack of global standardisation/harmonisation of ATC phraseology, resulting in various standards being in use in addition to the ICAO standard)
(5) *Frequency issues* (e.g. blocked, congested frequencies, and so on)

The aim of analysing the CHIRP reports was twofold: to establish the proportion of reports submitted that mention communication problems and the exact nature of these problems.

AAIB aircraft accident/incident reports
The same number of air accident/incident reports were analysed, i.e. 531, which are available in a searchable database on the AAIB website. As before, any reports considered relevant had to involve commercial flights. The AAIB database offers a variety of report types relating to an accident/incident, including full formal reports, summaries, and addenda (additional material relating to a report), and to avoid counting a report more than once only one of these report types was included in the total number of relevant reports. The reports were selected in reverse chronological order from 2015 to 2004 and the analysis of these 531 reports was conducted in the same manner as for the CHIRP reports, with the extracted communication issues allocated using the same categories. The aim of analysing the AAIB reports was to determine how many of the total number of accident/incidents examined in the reports involved communication problems and their exact nature.

Communication flight logs

[3] CHIRP reports were chosen because, unlike in the USA, ASRS reports in the UK are submitted to airlines and are confidential, hence requiring special permission for access.
[4] Truncated word forms were used to identify any occurrences which contained the specific string of characters.

A total of 30 pilots working for several UK airlines were asked at random to complete four-week flight logs for the purposes of recording any communication issues during their flights. Each of the four-week flight logs consisted of four consecutive weeks, and all the four-week periods recorded were also consecutive. A total of eight pilots responded[5] (26.6%), including a female pilot (12.5%)[6], and of these there were four captains (including two training captains) and four senior first officers. Six of the respondents fly long-haul routes and two are on short-haul. The pilots were provided with a prepared flight log template accessible on their work-issued iPads. Five pilots were able to provide three four-week periods, while three recorded two four-week periods due to standby and leave, resulting in a total of 21 flight logs. Instructions included that the total number of sectors flown in each four-week period needed to be recorded in a table, in which the pilots also had to highlight those sectors with communication issues. For each sector with such problems, they also needed to give a brief description of the exact nature of the problems. The completed flight logs did not need to be subjected to the same analysis as the above reports since no filtering was necessary. The reported miscommunications were then allocated to the same categories as above. In addition, the data were also analysed per pilot in order to answer the following questions: (i) Do short-haul and long-haul routes produce different results? (ii) How many communication issues are there per pilot? (iii) Do all the pilots experience the same types of problem? The goal of analysing real-life flight logs was both to obtain an idea of the extent to which pilots are exposed to communication problems during their day-to-day flying schedule and the nature of these problems.

4. RESULTS

The data were analysed using descriptive statistics. Note that due to the small size of the samples and the fact that they refer to the UK, the results do not lend themselves to generalisation using inferential statistics.

Results of the analysis of CHIRP reports

Out of the 531 CHIRP reports that were analysed, 117 reports were found to contain communication issues (22%). Since some reports revealed multiple communication problems, a total of 170 individual communication issues were counted. These 170 issues were subsequently allocated to the five categories of communication problems that could be identified in the relevant literature (see Section 3), but as the results in Table 4.1 below show, four further categories need to be added in order to accommodate the types of problem found. As can be seen, the largest number of problems belong to the message/transmission category (43%), closely followed by phraseology and terminology problems (21.8%). In third place, with 17.1%, we find frequency issues.

Results of the analysis of aircraft accident/incident reports

33 out of the 531 accident/incident reports contained communication issues (6.2%). The subsequent analysis of these revealed a total of 58 individual communication issues. As was the case with the CHIRP reports, the number of categories had to be extended by a further four. In Table 4.1 below, it can be seen that, as with the case of the CHIRP reports, the largest number of problems belongs to the message/transmission category (34.5%) and the second largest category with 17.2% concerns phraseology and terminology. Also in second place is the category to which unspecified communication issues were allocated, e.g. if it was impossible to identify why there was an unclear communication situation. In third place, issues with radio communication units were found in 12.4% of the cases, which is a much higher percentage than in the case of the CHIRP reports (2.4%). This was closely followed by issues regarding the quality of English (10.3%), which were cited twice as often as in the CHIRP reports (5.3%). It should be noted that no mention of local languages was made and that frequency issues play a lesser role than in the CHIRP reports but, as will be seen, they are much more frequently cited in the flight logs.

5 Although the response rate appears low, it fares well in view of the comment made in CHIRP Issue 72: “In the survey of flight crew, ATCOs and Licensed Engineers, we received 1,790 completed responses, around 6% of the total number of forms sent out. This percentage return is within the range expected for a survey of this kind” (2004: 1).

6 212 (5.5%) out of 3,866 pilots employed by British Airways are women (Pilot 1, pers. comm., July 25, 2015).

Table 4.1: Results of the three data analyses

Categories of Communication Issues	CHIRP reports		Aircraft accident & incident reports		Communication flight logs	
	N = 170	*f* / *N* (100%)	*N* = 58	*f* / *N* (100%)	*N* = 232	*f* / *N* (100%)
1. Unspecified communication issues (lack of, ineffective, unclear, low quality of communications)	*f* = 9	5.3%	*f* = 10	17.2%	*f* = 10	4.3%
2. Problems with radio communication units (complete/temporary loss)	*f* = 4	2.4%	*f* = 7	12.1%	*f* = 1	0.4%
3. Quality of English issues (heavily accented English by non-native speakers, by native speakers, using English colloquialisms)	*f* = 9	5.3%	*f* = 6	10.3%	*f* = 51	22%
4. Use of local languages	*f* = 6	3.5%	N/A	N/A	*f* = 16	6.9%
5. Phraseology and terminology issues (ambiguous, non-standard, unclear, lack of international harmonisation)	*f* = 37	21.8%	*f* = 10	17.2%	*f* = 3	1.3%
6. Frequency issues (busy, congested, blocked, unreadable, static interference, mis-set, closed, confusion, unauthorised frequency changes, lack of frequency change instructions, frequency 'black holes', HF communications of inadequate quality)	*f* = 29	17.1%	*f* = 3	5.2%	*f* = 61	26.3%
7. Message/transmission issues (incomplete, incorrect, missed, misheard, misconstrued, unclear, conflicting, unrealistic, omission of, too complex, spoken too fast, split messages, unauthorised, at inappropriate times, numerals incorrectly pronounced, repeated several times over, call sign issues)	*f* = 73	43%	*f* = 20	34.5%	*f* = 87	37.5%
8. Unwillingness-to-communicate issues (deliberately ignoring transmissions, not clarifying instructions)	*f* = 1	0.6%	N/A	N/A	*f* = 3	1.3%
9. Other communication issues (due to O_2 masks, high ambient cockpit noise, side tone in controller headsets causing confusion, unreliable new RT systems for ATC)	*f* = 2	1.2%	*f* = 2	3.4%	N/A	N/A

Results of the analysis of communication flight logs

A total of 21 four-week flight logs was produced by eight pilots. In turn, these flight logs consist of a total of 240 sectors: 125 long-haul and 115 short-haul. The pilots reported communication problems in 85 out of the 240 sectors (35.4%). On short-haul, 26 out of the 115 sectors produced problems (22.6%), while on long-haul 59 out of 125 sectors generated issues (47.2%). For the 85 sectors with communication issues, a total of 232 individual problems were logged. Table 4.1 above shows that the message/transmission category is ranked highest (37.5%), followed by frequency issues with 26.3%. In third place, there is the quality of English category with 22%, and the use of local languages came fourth (6.9%). It can be seen that most problems also fall into the message/transmission issues category, but that, unlike in the two report types, phraseology/terminology issues were rarely reported. However, the quality of English and local languages categories were logged much more often than in the reports. It is noteworthy that on three occasions the pilots encountered an unwillingness to communicate by ATC, which was cited just once in the CHIRPs, but was absent in the accident/incident reports.

Table 4.2 below shows the results per pilot, which indicate that within the long-haul and short-haul groups the percentage of communication issues is somewhat similar. On long-haul, most pilots reported problems on 37% to 54% of their sectors flown, with the exception of Pilots 6 and 7 who logged problems in 93.8% and 21.1% of the sectors respectively. The two short-haul pilots produced much lower percentages, i.e. 20.9% by Pilot 5 and 29.2% by Pilot 8. In summary, the flight logs revealed a much higher percentage of communication problems (overall 35.4%; long-haul 47.2%; short-haul 22.6%) than in both the CHIRP (22%) and aircraft accident/incident reports (6.2%).

Table 4.2: Results by pilot

Pilot	Gender	Rank*	Fleet	Flight length**	Number of four-week periods	Number of sectors	Number of sectors with communication issues	*f* / *N* (100%)
1	M	TrngCPT	B747-400	LH, MH	3	24	$f = 13$	54.2%
2	M	CPT	B747-400	LH	3	20	$f = 9$	45%
3	M	TrngCPT	B747-400	LH, MH	2	24	$f = 9$	37.5%
4	M	SFO	B747-400	LH, MH	3	22	$f = 9$	40.9%
5	M	SFO	A321/20/19	SH, MH	2	91	$f = 19$	20.9%
6	M	CPT	A380	ULH, LH	3	16	$f = 15$	93.8%
7	F	SFO	B747-400	ULH, LH, MH	2	19	$f = 4$	21.1%
8	M	SFO	A320/19	SH, MH	3	24	$f = 7$	29.2%
Total					21	240	$N = 85$	35.4%

* TrngCPT = Training Captain; CPT = Captain; SFO = Senior First Officer
** ULH = Ultra long-haul sector, LH = long-haul sector, MH = medium-haul sector, SH = short-haul sector

5. DISCUSSION

The results from the three analyses[7] are clearly not at a similar level to the information transfer issues reported by Billings and Reynard (1981), who, as will be recalled, reported that 73.3% of the analysed ASRS reports included such issues. However, the lower number of reported issues is not unexpected as since 1981 many improvements have been made regarding ICAO phraseology, new training and testing methods in aviation English for non-native speakers were introduced, and human factors training intensified. Also, since 1981 fewer High Frequency (HF) frequencies have been in use, which has improved the quality of ATC transmissions. However, in view of the results from the flight logs, in which still relatively high numbers of communication issues have been reported, this explanation may not give us the full picture. It is necessary therefore to investigate the factors which may have influenced the three data analyses in this study.

A likely reason behind such a discrepancy between Billings and Reynard's 73.3% and the CHIRP result of 22% could be the number of ASRS reports they analysed, i.e. 12,373, all of which were submitted in a period of two years. In contrast, in the UK considerably fewer CHIRP reports were filed during a two-year period as only 531 relevant reports (the number of omitted reports is roughly the same) were counted from 2015 to 2000. The reason for this could be fourfold. Firstly, far more flights take place per day in the US compared to the UK. Secondly, while ASRS reports are anonymous, CHIRPs are confidential but not anonymous, which means that pilots and controllers may be reluctant to file reports for fear of repercussions. Thirdly, the low number of reported issues to CHIRP may also be due to a lack of interest in reporting them. Finally, the legal implications of ASRS reporting in the USA cannot be discounted as an explanation for the high level of reporting.

The low figure of 6.2% of cited communication problems in the aircraft accident/incident reports ties in with the observation by IATA, IFALPA, and IFATCA that non-standard phrases, local languages, and accents are rarely cited as being contributory to accidents, but that for the large majority of pilots and controllers these issues are a cause of concern and regularly result in misunderstandings (2011: 7-8). They fail to provide any explanation for this, but it may be speculated that perhaps, the specific personality traits of pilots play a role here. Commercial pilots are a fairly homogenous group in terms of stress resistance, multi-tasking abilities, ability to find solutions fast, willingness and ability to adapt, to cope with problems under pressure, and they have a "can-do" attitude when faced with solving problems (cf. Green *et al*. 1996, The Air Pilot's Manual 2013). Hence, in view of the relatively high number of problems reported in the flight logs, in particular by Pilot 6 who cited issues in 93.8% of his sectors, it may indeed be the personality traits of pilots which prevent communication issues from developing into an accident. The same could perhaps be said of the personality traits of controllers which appear to be similar.[8] This is an area that *could warrant further investigation*. However, it can be argued that this reliance from regulators and airlines on pilots and controllers being able to cope with and resolve communication problems on a continual basis means that this is an accident waiting to happen. Concerns are also expressed by controllers: "What did he say? Asking "Say again", often leaves us none the wiser. Whilst we manage on a day-to-day basis, we are concerned that we would not understand them in the event of an emergency" (CHIRP Issue 80 2006: 3).

---

7 Limitations of this study can be seen in the small sample sizes and that only pilot flight logs but no ATC communication logs could be obtained. The accident/incident and CHIRP reports though include inputs from both pilots and controllers.

8 For instance, cf. https://www.eurocontrol.int/articles/skills-required-be-air-traffic-controller (Courtesy of EUROCONTROL).

The pilots reported in their flight logs that the most frequent communication issues belong to the message/transmission category, in which they, for example, listed incomplete, incorrect, missed, misheard, and too complex messages. This was closely followed by frequency issues, where the major problems reported are congested, blocked, and busy frequencies, and also the inadequate quality of HF frequencies over Africa and South-East Asia, which means that they regularly fly long distances without any ATC contact. Both categories overlap, however, inasmuch as some message/transmission issues are brought on by frequency issues, e.g. by low-quality HF, by blocked, or congested frequencies. The pilots often reported that they had to spend a lot of time discussing with each other what the controller had said, and they also mentioned how often messages needed to be clarified, e.g. Pilot 1 reported that one message had to be repeated 4-6 times. The third most cited problem was the quality of English category, which also overlaps with the message/transmission category as instructions given in heavily accented English often need to be repeated several times. Another frequent problem was the use of local languages, which the pilots unanimously reported as causing a decrease in situational awareness. A surprising issue was logged by Pilot 7 who said that in a particular airspace, ATC would not speak to her because, in her opinion, she was female.[9] Other reported refusals to communicate stem from political differences between two states. Due to the small sample, we cannot know whether such refusals, particularly in the case of the female pilot, are just isolated occurrences or whether this could be indicative of a bigger problem. The discrepancy between short-haul and long-haul results can be explained by the different countries and regions that are flown to or over. The problems on long-haul are much more characterised by frequency issues, the quality of English spoken and the use of local languages than is the case for the predominantly European short-haul flights.

In terms of mental workload and situational awareness, it was in particular the long-haul pilots that were adamant that communication issues had a negative effect. Since situational awareness is part of a pilot's decision-making process, which in turn is affected by high stress/workload situations, it is clear that communication issues leading to such a situation are undesirable since the loss of situational awareness could lead to accidents. Hence, although the percentages of 35.4% overall and 47.2% long-haul problems were much lower than Billings and Reynard's 73.3%, these figures can nevertheless be considered to be too high. It is thus surprising that despite continuous human factors training, improved training and testing methods for aviation English, the ongoing process by ICAO to improve ATC phraseology, and the decrease in operational HF frequencies, such a high number of problems should still exist in today's flight decks and ATC workstations.

Therefore, it is pertinent to ask why such communication issues have not been reduced even further. There may be two obvious answers and one less obvious one to this question. Firstly, the human factor will always be present in interpersonal communications and this makes an eradication of errors unlikely. Secondly, technology has not advanced enough yet to solve frequency problems once and for all, and the planned expanded CPDLC via satellites will only solve communication issues to some extent but cannot solve the use of local languages and the quality of English. Thirdly, and less obviously, it may be argued that in pilot-controller communications, there is an inconspicuous concept involved that so far seems to have resisted most human factors efforts inasmuch as that many pilots and controllers are still not sufficiently mindful of what we may call *communication awareness*[10]. This often manifests itself in "*us-and-them*" communication situations instead of "*we*" situations. Indeed, as far back as in 1981, Billings and Cheaney surmised that a lack of awareness by pilots and controllers of the exact nature of communication could partly be responsible for causing communication issues. Hence, it can be argued that while some issues have been reduced, too many still occur and pilot-controller communication awareness has not yet improved sufficiently.

6. POTENTIAL SOLUTION

Since the results of the communication flight logs have shown that voice communications remain problematic and routinely cause misunderstandings and since there still appears to be a considerable lack of communication awareness among pilots and controllers, it is vital that such communication issues be minimised. Below, we explore the idea of developing an additional communication system using language-engineering technologies in automatic speech recognition, machine translation, and term extraction; however, due to limitations of space, the individual technologies will only be outlined briefly.

---

9 Pilot 7 also reported that she had similar experiences as a short-haul pilot.

10 Taking Billings and Cheaney's quote (cf. Chapter 2) as the basis, we propose the following working definition for the concept of communication awareness: It is vital to know exactly the nature of the communication situation we are in, to know what needs to be communicated, how it should be communicated, and when exactly. It also means that we are as fully informed as possible about the other person's communicative environment and that we are *willing* to communicate appropriately. For example, it is important to communicate slowly with the awareness that the other person may take longer to understand the information due to the environment and situation they are in.

Automatic speech recognition, machine translation, and term extraction

The goal of automatic speech recognition (ASR) is to transfer speech to text on the basis of speech recognition algorithms which transform a sequence of acoustic waves into a sequence of written text (Jurafsky and Martin 2009). The problem of speech being automatically recognised irrespective of any surrounding conditions poses problems, but recently substantial performance improvements have been achieved in the application of Deep Neural Networks (DNN) into ASR technology (e.g. Maas *et al.* 2012). Performance levels can also be raised if ASR is applied to *controlled languages* (characterised by standardised phrases, reduced/disambiguated terms, limited syntax, and repetitive information) and ATC phraseology fulfils the requirements of such a language. Like ASR, the quality of machine translation (MT) depends on many problems inherent in language and speech. MT performance also increases the more a language is reduced in terms of its grammatical structures, linguistic devices, and if a domain is delimited. For example, high-quality results have been achieved in the domain of meteorology (Gotti, Langlais and Lapalme 2013), which is a field characterised by standardised weather reports, reduced and disambiguated terminology, controlled language (e.g. limited use of syntax), and repetitive information (cf. Jurafsky and Martin 2009). The state-of-the art technology used in MT systems is hybrid, which brings together rule-based as well as statistical and example-based methods. The main advantage of this combined approach is that within a delimited domain, the quality of translations is likely to be high. As part of the machine translation process, terminology extraction (TE) often takes place simultaneously (cf. Vivaldi and Rodríguez 2007) and automatically identifies term candidates from spoken or written specialist texts. Approaches to term extraction include linguistic and statistical methods, but more recent tools are hybrid. Like MT, performance levels are likely to be high if used in delimited domains with controlled languages.

The proposed Advanced Intelligent Communication System (AICSys)

ASR systems for use in aircraft have been investigated for several years with a view to applying ASR in voice input systems in cockpits (e.g. Baber and Noyes 1996; Lennertz *et al*. 2012) and for training controllers (e.g. Cordero, Dorado and de Pablo 2012). For example, direct voice input by pilots is already used in the Eurofighter Typhoon (2014) and in the F-35 Lightning II Joint Strike Fighter (Schutte 2007). Similarly to our study, Geacăr (2010) investigates speech input as a back-up solution to voice communications, which means that any voice message is transcribed into text and transmitted as such via data link. In contrast, Lennertz *et al*. (2012) study data link messages that are accompanied by synthetic speech outputs in addition to voice communications with a view to reducing head-down time during single-pilot operations. It is clear, however, that the above approaches would be unable to deal with local language use as well as English accents and dialects. Geacăr's suggestion to use ASR for transcribing all ATC radio messages and to transmit these using data link seems to correspond to the ASR phase described in our system (see below), but he does not mention how local-language messages or messages spoken with accents or in dialects would be dealt with. Although Lennertz *et al*. claim that adding synthetic speech commands to textual data link messages did not "introduce additional complications" (2012: 31), it does not seem to be well justified. Since their experiment involved single-pilot operations, the combined synthetic speech-text method may have its benefits in such a situation. It remains to be seen, however, how synthetic speech outputs of data link text messages, in addition to the voice communications, will fare in two-pilot operations, which are more complex in terms of pilot tasks and aircraft systems. Moreover, in two-crew operations, Crew Resource Management (CRM) requires the crew to communicate with each other to a large extent and interruptions from synthetic speech outputs could disrupt workflow resulting in increased workload.

Key findings in the areas of ASR and MT show that the performance levels of both technologies depend on the application of the latest artificial intelligence techniques and on how issues such as dialects, accents, prosody, channel, noise, vocabulary size, domain delimitation, grammar differences, etc. are dealt with. The task of creating an intelligent communication system which produces results of the highest quality will be challenging. However, as we have seen, all three technologies should perform well with the controlled language of ATC phraseology. Consequently, the proposed system should be able to reduce many of the communication issues mentioned in this paper as follows:

- During the first phase, ASR technology would convert spoken English messages into text in real-time. This allows pilots and controllers to see the message on a display, including any non-standard phrases, which could then be queried with ATC if necessary. Since state-of-the-art ASR systems are trainable at source by individual speakers, this should also address pronunciation, enunciation, and prosody issues for both native and non-native speakers.
- The simultaneously running TE phase would highlight safety-critical terminology as high-priority information.
- If local languages are used, in the second phase MT technology is triggered into action and would enable translations into English in real-time.

- Since we propose to use data link methods, frequency issues should not affect the transcribed/translated messages.

The AICSys on the flight deck and in ATC workstations

The system is intended as an additional communication system to voice communications, thus contributing to the level of system redundancy in ATC workstations and in particular in flight decks, where thus far voice radio communications are without redundancy.

The proposed system would provide pilots and controllers with text versions of voice messages. For example, for every voice message to an aircraft, a transcribed text version would be sent via data link and be shown in real-time on a display on the flight deck. Simultaneously, any other aircraft in the same airspace which has the system installed would also receive the text version of that particular message on their display, so that pilots would also be able to read messages intended for other aircraft. If a local language is used on the radio, the MT system would get activated and pilots and controllers would receive English translations of the foreign-language message on their respective displays in real-time. Pilots would therefore always be able to understand messages in a local language and hence be in the picture of what nearby aircraft are doing and their intent. It would, however, be up to the pilots to decide whether they need to look at messages on the displays or not. For instance, as long as the pilot dealing with ATC communications hears transmissions in English using standard phrases, s/he would probably have no need to look at the display. However, if the message is garbled or if local languages are used, or if the aircraft is close to any runways, the pilot would want to know what has been said and look at the display.

Similarly, during times of high mental workload, controllers may filter messages from aircraft (Airbus 2014), hence not all communications addressed to them may be listened to or acknowledged if they are, for example, talking to another aircraft. The additional transcripts/translations on a suitable display in the workstation would therefore give them the chance to read a message should the need arise. Consequently, the role of the AICSys in avoiding miscommunications on the ATC side is evident. Using such a system, it should be possible to avoid ambiguity in critical information regarding, for example, clearances, and pilots and controllers might be prevented from mishearing call signs, information about flight levels or speed, irrespective of dialects, etc. Given the fact that controllers deal with more than one aircraft at a time, design and implementation considerations will be crucial in terms of how to separate the various transcripts/translations from individual aircraft in such a way that they are easily distinguishable (possibly separated from each other in space and by colour).

Human Factors Considerations

It is vital during the high workload phases of flight that communications are accurate and not time-consuming. The proposed system may therefore have the potential to reduce high workload since pilots would have more time to aviate and navigate rather than having to spend time double-checking messages with ATC or by figuring out among each other what was said. When workload increases, communication tends to disintegrate (cf. Barshi and Farris 2013), for example, messages decrease in length under high workload (e.g. Raby and Wickens 1994) as a result of restrictions on a human being's time-sharing abilities that underlie complex performance (Jennings and Chiles 1977). How would this time-sharing ability be affected by the additional transcripts/translations in the flight deck? According to Wickens' 4-dimensional multiple resource model (2008), the perception of auditory and visual information occurs in different parts of the brain, which means that time-sharing between tasks that use different resources should be less conflicted. However, he points out that if both the auditory and the visual tasks need processing at a higher level, which is what the comprehension of spoken words and written text would require, then this "will still compete for common perceptual resources (and may also compete for common code-defined resources […])" (2008: 450). We nevertheless argue that the provision of transcripts/translations on a display would be an improved situation for pilots since in phases of high workload the pilot monitoring knows that if s/he has not been able to hear or understand a message over the radio, it will be available as text on the display. Through the help of the displayed text, it is argued that mental resources in the pilots' brains are freed.

For controllers, considerations of mental workload, situational awareness, and attention allocation are equally as important. The use of the proposed communication system could provide improved situational awareness for controllers as the lack of ambiguity in communications will greatly enhance the accuracy of the information provided. This being the case, then the certainty of information will in addition enable controllers to manage their workload in a fashion considerably more enhanced than at present.

7. CONCLUSIONS

This impetus behind this work was twofold. The first aim was to establish the current level of communication problems between pilots and controllers. The results from the data analyses showed that although the number of issues has fallen considerably since Billings and Reynard's seminal research in 1981, problems during pilot-controller communications are still numerous as was shown in particular by the analysis of the flight logs compiled by British commercial pilots. As a result of the ongoing problems, the second aim was to propose and explore a potential solution in the form of an additional communication system in order to minimise such problems. This paper presented a description of how such a system may improve pilot-controller workload and situational awareness. Further research will focus on creating a prototype of the system, while gaining a better understanding of what factors directly affect pilot-controller communications, as part of which the concept of communication awareness will be developed further.

REFERENCES

Airbus. 2014. Flight operations briefing notes: Human performance – Effective pilot/controller communications. http://www.airbus.com/fileadmin/media_gallery/files/safety_library_items/AirbusSafetyLib_-FLT_OPS-HUM_PER-SEQ04.pdf (accessed July 14, 2015).

Alderson, J. C. 2011. The politics of aviation English testing. *Language Assessment Quarterly 8 (*4): 386-403.

Baber, C. and J. Noyes. 1996. Automatic speech recognition in adverse environments. *Human Factors* 38 (1): 142-155.

Barshi, I. and C. Farris. 2013. *Misunderstandings in ATC communication.* Farnham, UK: Ashgate Publishing.

Billings, C. E. and E. S. Cheaney. 1981. The information transfer problem: Summary and comments. In *Information transfer problems in the aviation system,* ed. C. E. Billings and E. S. Cheaney, 85-93. NASA Technical Paper 1875. Moffett Field, CA: NASA Ames Research Center.

Billings, C. E. and W. D. Reynard. 1981. Dimensions of the information transfer problem. In *Information transfer problems in the aviation system,* ed. C. E. Billings and E. S. Cheaney, 9-14. NASA Technical Paper 1875. Moffett Field, CA: NASA Ames Research Center.

Cardosi, K. M. and P. W. Boole. 1991. Analysis of pilot response time to time-critical air traffic control calls. Cambridge, MA: National Transportation Systems Center.

Cordero, J. M., Dorado, M. and J. M. de Pablo. 2012. Automated speech recognition in ATC environment, ATACCS'2012, Research Papers, 46-53. www.hala-sesar.net/sites/default/files/documents/p46-cordero.pdf (accessed July 14, 2015).

Cushing, S. 1994. *Fatal words: Communication clashes and aircraft crashes*. Chicago: University of Chicago Press.

Elliot, G. 1997. English in aviation safety: Testing and training solutions. In *Selected Proceedings of the 1997 Symposium 'Aviation Communication'*, 9-11 April 1997, ed. P. Quigley and P. McElwain, 21-23. Prescott, AZ: Embry-Riddle Aeronautical University.

EUROCONTROL. 2006. Air-Ground Communication Safety Study Causes and Recommendations. Edition 1.1. http://www.skybrary.aero/bookshelf/books/162.pdf (accessed July 14, 2015).

Eurofighter Typhoon. 2014. Technology: Cockpit. http://typhoon.starstreak.net/Eurofighter/cockpit.html (accessed: July 14, 2015).

Forcada, M. L. 2010. Machine translation today. In: *Handbook of translation studies*, Volume 1, ed. Y. Gambier and L. Van Doorslaer, 215-223. Amsterdam/Philadelphia: John Benjamins Publishing.

Geacăr, C.-M. 2010. Reducing pilot/ATC communication errors using voice recognition. In *Proceedings of the 27th International Congress of the Aeronautical Sciences, ICAS 2010*. www.icas.org/ICAS_ARCHIVE/ICAS2010/PAPERS/441.pdf (accessed July 14, 2015).

Goguen, J. A., Linde, C. and M. Murphy. 1985. Crew communication as a factor in aviation accidents. NASA Technical Memorandum 88254. Moffett Field, CA: NASA Ames Research Center.

Gotti, F., Langlais, P. and G. Lapalme. 2013. Designing a machine translation system for Canadian weather warnings: A case study. *Natural Language Engineering* 1(1): 1-36.

Green, R. G., Muir, H., James, M., Gradwell, D. and R. L. Green. 1996. *Human Factors for Pilots.* 2nd edition. Aldershot, UK: Avebury Aviation.

IATA, IFALPA and IFATCA. 2011. *Pilots and air traffic controllers phraseology study*. Montreal: International Air Transport Association.

Jennings, A. E. and W. D. Chiles. 1977. An investigation of time-sharing ability as a factor in complex performance. *Human Factors* 19: 535-547.

Jurafsky, D. and J. M. Martin. 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.* 2nd internat. edition. Upper Saddle River, NJ: Pearson Prentice Hall.

Kanki, B. G. and H. C. Foushee. 1989. Communication as group process mediator of aircrew performance. *Aviation, Space and Environmental Medicine* 60 (5): 402-410.

Lennertz, T., Bürki-Cohen, J., Sparko, A. L. et al. 2012. NextGen Flight Deck Data Comm: Auxiliary Synthetic Speech – Phase I. In *Proceedings of the Human Factors and Ergonomics Society* 56: 31-35.

Maas, A. L., Le, Q. V., O'Neil, T. M. et al. 2012. Recurrent Neural Networks for Noise Reduction in Robust ASR. http://www.isca-speech.org/archive/archive_papers/interspeech_2012/i12_0022.pdf (accessed: July 14, 2015).

Monan, W. P. 1983. Addressee errors in ATC communication: The call sign problem. NASA Contractor Report 166462. Moffett Field, CA: NASA Ames Research Center.

Morrison, R. and R. H. Wright. 1989. ATC control and communication problems: An overview of recent ASRS data. In Proceedings of the Fifth International Symposium of Aviation Psychology, Volume 2, ed. R. S. Jensen, 902-907. Columbus, OH: Ohio State University.

Prinzo, O. V. and T. W. Britton. 1993. ATC/Pilot voice communications – A survey of the literature. Oklahoma City: FAA Civil Aeromedical Institute.

Raby, M. and C. D. Wickens. 1994. Strategic workload management and decision biases in aviation. *The International Journal of Aviation Psychology* 4: 211-240.

Schneider, V. I., Healy, A. F., Barshi, I. and J. A. Kole. **2011. Following navigation instructions presented verbally or spatially: Effects on training, retention, and transfer.** *Applied Cognitive Psychology* 25: 53-56.

Schutte, J. 2007. **Researchers fine-tune F-35 pilot-aircraft speech system. Human Effectiveness Directorate.** http://www.afmc.af.mil/news/story.asp?id=123071564 **(accessed July 14, 2015).**

The Air Pilot's Manual. 2013. *Human Factors & Pilot Performance*, Volume 6. Shoreham Airport, UK: Pooleys-Air Pilot Publishing.

Vivaldi, J. and H. Rodríguez. 2007. Evaluation of terms and term extraction systems: A practical approach. *Terminology* 13 (2): 225-248.

Wickens, C. D. 2008. Multiple resources and mental workload. *Human Factors* 50 (3): 449-455.