Aviation Safety: An Initial Exploration of the Feasibility of Using Language Engineering Technologies for Reducing Pilot-Air Traffic Control Miscommunications

Bettina Bajaj [a] *and Arnab Majumdar* [b]

[a] *Centre for Translation Studies (CenTraS)*
University College London
London, WC1H 0PQ, UK

[b] *Centre for Transport Studies*
Imperial College London
London, SW7 2AZ, UK

ABSTRACT

This paper describes some initial investigations into the possibilities of using state-of-the-art language engineering technologies to minimise miscommunications between pilots and controllers. Despite considerable efforts to remedy this situation by providing solutions that focus almost exclusively on new proposals for making air traffic control (ATC) messages clearer and easier to understand and on better ATC communication training strategies, communication issues persist. In order to demonstrate this, we discuss an aircraft accident and incident in which communication problems between pilots and ATC have been identified as contributory factors. The types of miscommunication are described in their situational and operational contexts. It is then argued that employing automatic speech recognition (ASR), machine translation (MT) and terminology extraction (TE) technologies would have the potential to reduce such miscommunications and hence might have contributed to preventing the accident and incident. This paper presents a snapshot of our initial work as well as thoughts on its future development, including a description of how an ASR-MT-TE communication system as an addition to voice and data link communications might be designed and implemented into flight decks and ATC workstations and how this system may impact on mental workload, situation awareness, and attention allocation of pilots and controllers.

Keywords: Attention Allocation, ATC Miscommunications, ATC Phraseology, Automatic Speech Recognition, Machine Translation, Mental Workload, Situation Awareness, Terminology, Terminology Extraction

INTRODUCTION

Like the general topic of communication, ATC communications is a complex phenomenon involving a number of different aspects, such as those shown below. Research into the area of ATC communications tends to focus either on individual aspects or combinations of these.

- linguistic aspects (e.g. Cushing, 1994; Barshi, 1997; IATA, 2011)
- cognitive aspects (e.g. Reason, 2008)
- psychological aspects (Mosier *et al*., 2013)
- communicative aspects (e.g. Linde, Goguen and Devenish, 1986)
- technical aspects (e.g. Cushing, 1994; O'Neil, 1999)
- human factors aspects (e.g. Hawkins, 2010)
- combinations of aspects (e.g. linguistic/cognitive; see Barshi and Farris, 2013)

ATC communications take place via voice communications. To this end, specific phraseologies and terminologies have been created in order to facilitate clear, unambiguous, concise, and efficient communications between pilots and controllers (cf. ICAO, 2007). ATC phraseology/terminology can be described as a restricted sublanguage with reduced vocabulary, clearly assigned terms, simplified syntax, and altered pronunciation. This controlled language aims at avoiding miscommunication by eliminating specific linguistic devices, such as those listed below.

- Homophones[1], e.g. *to* vs. *two*
- Homographs[2], e.g. *close* (which can mean *near* or *shut*, and so on)
- Homonyms[3], e.g. *go ahead* (meaning *to urge speaking* or *to move forward*)
- Synonyms, e.g. *runway holding point* and *runway holding position*

The International Civil Aviation Organization (ICAO) has developed and published the ICAO ATC phraseology standard (ICAO, 2007), which is the international standard that should be adopted world-wide. However, IATA, the International Air Transport Association, has recently highlighted that unambiguous radio communications between pilots and ATC are routinely impeded as both groups often depart from this standard (IATA, 2011), e.g. by using non-standard phrases such as *ready for take-off* instead of the ICAO phrase *ready for departure*. They also observe that the quality of English being spoken by both groups is characterised by more or less pronounced dialects and accents and that local languages are used frequently by pilots and controllers, although ATC communication regulations clearly stipulate that radio communications should be carried out in English.

Voice and Data Link Communications

Numerous studies have attempted to improve voice communications between pilots and controllers, which are carried out via very high frequency (VHF) and high frequency (HF) radio, and also occasionally via satellite.

Linguistic studies have led to a number of vital recommendations for improvements. Suggestions have, for example, been made regarding message lengths, message complexity and phraseology wordings (e.g. Barshi, 1997; Barshi and Farris, 2013). A large number of studies have put forward new methods and tools for improved training to help reduce errors in ATC communications (e.g. Cushing, 1994; Elliot, 1997; Robertson, 1997; Alderson, 2011).

Non-linguistic studies have addressed issues such as frequency congestion, noise reduction, and so on, and have also focused on alternative methods of getting ATC messages across to pilots and controllers. Text messaging as an additional means to voice communications has been examined in studies on controller/pilot data link communication (CPDLC) (e.g. Schneider, Healy, Barshi and Kole, 2011). At present, ATC messages can be sent and generated as text on displays to and from the flight deck and to and from the controllers' workspace. While CPDLC helps to reduce frequency congestion, it is nevertheless a limited service which is used only for en route navigation (cf. Nolan, 2011), i.e. in the cruising phase, and not during the terminal phases of flight (descents and approaches). According to Nolan (2011), current data link messages include four types of information: initial ATC aircraft contact, altimeter settings, communications transfer to the next sector, and various messages from an approved menu. Future plans are to expand data link transmissions while reducing the number of voice communications at the same time (cf. also Mosier *et al*., 2013). This will be concurrent with the introduction of digital communication methods, which will allow a much larger number of messages to be transmitted, including aircraft information, weather, traffic and approach information as well as routine pilot-controller communications (Nolan, 2011). Cardosi,

---

1 Words which are pronounced the same but are different in meaning (Crystal, 1992).

2 Words that have the same spelling but are different in meaning (Crystal, 1992).

3 Words with the same form but different in meaning (Crystal, 1992).

Falzarano and Han (1998) point out that proponents of the expanded CPDLC believe that this will eliminate all ATC communication errors. However, while the envisaged CPDLC system may get rid of some of the miscommunications, Cardosi *et al.* (1998) do not believe that all of the existing ATC communication issues can be removed. They also anticipate that other error types may start appearing, e.g. misread numbers, messages sent to wrong aircraft, and so on. According to them, it is important to design and implement the future data link system with close attention to human factors considerations so as not to add to the workload of pilots and controllers.
A number of studies appear to be in the process of exploring the use of data link messages also during high workload phases since they are examining the application of automatic speech/voice recognition (e.g. Geacăr, 2010; Cordero, Dorado and De Pablo, 2012) as well as auxiliary synthetic speech (e.g. Lennertz *et al.*, 2012; 2013) in addition to ATC voice communications during such phases. As we will show later, while there are similarities to our proposed communication system, the above approaches would not be able to deal with the issue of local languages.

Until CPDLC is used widely as part of the Next Generation Air/Ground Communication System (NEXCOM), voice communications will remain the main means of ATC communications (Nolan, 2011; Airbus, 2014), the unambiguity of which is hence of safety-critical importance. It is also noteworthy here that line pilots tend to consider the current data link messaging as time-consuming and cumbersome (personal communication with a UK Training Captain B747-400, 27 August 2013). Barshi and Farris appear to have a similar view as they note that data link messages have a clear disadvantage over voice communications since pilots react more easily to voice messages than to a text message (2013). However, as will be seen later, it is argued that this may be different when ATC messages are textually displayed using the communication system suggested here.

ATC MISCOMMUNICATIONS

While there have been many recommendations for making ATC communications less prone to error, it is nevertheless clear that there are still far too many barriers to safe communications between pilots and air traffic controllers – a situation which has obvious consequences for aviation safety. The difficulty lies not only in the large number of communication problems but also in the types of problem that can occur during ATC communications. In general, the issues of miscommunication can be categorised into five major groups[4]:

(1) The use of local languages during ATC communications.
(2) The variations in English pronunciation and enunciation by native and non-native speakers.
(3) The variations in prosody, i.e. tempo, rhythm, pitch and loudness, by native and non-native speakers.
(4) The use of non-standard phraseology.
(5) A lack of global standardisation and harmonisation of ATC phraseology, resulting in several standards being in use, including
  - the ICAO international phraseology standard;
  - the CAA phraseology standard in the UK, which intentionally contains some non-standard phraseology (Skybrary, 2013a);
  - the standards by several other European countries, which have also adopted some non-standard phrases (Skybrary, 2013a);
  - the FAA phraseology standard in the US, which also employs non-standard phraseology.

While there are numerous publications on the communication problems arising from the use of non-standard ATC

4 Another problem that can be experienced during ATC communications is not listed above as it does not seem to be mentioned in the relevant literature, yet it is often experienced by line pilots (personal communication with a UK Training Captain B747-400, 21 June 2013). This problem tends to occur when language beyond standard ATC phraseology has to be used. For example, when pilots (or controllers) who are English native speakers try to communicate with controllers (or pilots) whose first language is not English in a manner that is more appropriate for communicating with other English native speakers (e.g. by using idiomatic expressions, complex words, long sentences, and so on) breakdowns in communication often ensue. The English native speakers should instead use more easily understandable language, such as clear, simple and unambiguous English. This problem ties in with an observation made in the area of automatic speech recognition, which performs less well in situations when humans talk to humans whereas it performs much better when humans talk to computers since in such cases humans tend to speak more slowly, simply and clearly (Jurafsky and Martin, 2009).

phraseology and terminology, it is rarely discussed that there are differing ATC phraseology and terminology standards in existence. Exceptions are, for instance, the report from IATA (2011) and the Airbus Flight Operations Briefing Notes on effective pilot/controller communications (2014), in which it is noted that there are critical differences between, for example, the ICAO and the North American ATC phraseology standards. A typical example of such a discrepancy in phraseology would be the situation of a commercial airliner being in danger of running out of fuel during the landing approach. The ICAO phraseology taught to UK pilots stipulates that the pilot responsible for ATC communications has to declare an alert using *Pan-Pan Pan-Pan Pan-Pan* followed by the mention of the call sign together with the nature of the emergency, i.e. low fuel state, to ATC ground control. In contrast, in US airspace the term *minimum fuel advisory* has to be used, followed by the pilot's statement as to the number of minutes of fuel remaining. Further examples of such discrepancies are shown in Table 1 below.

Table 1: Examples of differences in ATC phrases and terms between ICAO, FAA and CAA

ICAO	**FAA (Federal Aviation Authority), US**	**CAA (Civil Aviation Authority), UK**	**Definition/Meaning of Concept**
vacate	*exit*	*vacate*	moving away from something
Pan-Pan Pan-Pan Pan-Pan	*minimum fuel advisory*	*Pan-Pan Pan-Pan Pan-Pan*	signalling emergency due to, for example, low fuel state
apron	*apron* *ramp*	*apron*	paved parking area excluding runway and taxiway
line up and wait	*taxi into position and hold*	*line up and wait*	clear to line up on the departure runway but not cleared to take off
descend now	*descend*	*descend now*	ambiguity between descending now or later
taxi to holding position	*taxi into position and hold*	*taxi to holding position*	taxi to, and hold at, at a point clear of the runway

As has been seen, the ICAO phraseology standard is the standard which ought to be used globally. However, we have also seen that, to varying degrees, several countries have adopted their own phraseology standards (ICAO, 2007) and these seem to have arisen from the fact that although ICAO emphasises that their standard ought to be adopted by all the stakeholders, they do in fact not have any powers to make it a prescribed standard and adoption of it is left to the countries participating in civil aviation. The reasons as to why individual countries are unwilling to adopt the ICAO standard might be due to cultural and political reasons. There may be a national reluctance in adopting a standard that is seen to be 'imposed' on them by an international organisation. In this context, the recent phraseology study by IATA highlights the 'hope' that their study "will provide momentum towards a greater harmonization of communications" (IATA, 2011: 54). However, it may be argued that the harmonisation of ATC phraseology and terminology is not something that should be left at the discretion of airlines and national aviation regulators but that it should be viewed as an indispensable tool for improving aviation safety.

ATC communications is a highly safety-critical area in which the use of local languages and non-standard ATC phrases and terms have already led to several fatal aircraft accidents (e.g. accident of KLM and PanAm in Tenerife 1977; Avianca accident at New York in 1990), and hence it seems paradoxical that not only local languages are spoken during ATC communications, but also that pilots and controllers do not always adhere to standard phraseology and terminology and that differing ATC phraseology and terminology standards should exist in the first place. Variations in pronouncing English and speech idiosyncrasies also remain a significant impediment to ATC communications as even among English native speakers' variations in English pronunciation and idiosyncrasies, such as speaking fast, can cause problems. While the latter problem can be improved upon to a large degree by speech training, the problems of using local languages and non-standard phrases are more difficult to solve.

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2097-8

Pilots, in particular, get very confused when they are, for instance, on the approach to a foreign airport and they can hear ATC speaking to another aircraft on the same flight path in a local language. Those pilots who are unable to speak the local language will not know what this particular aircraft is doing in terms of altitude, speed, heading, and so on (personal communication with a UK Chief Pilot Boeing Fleet and a UK Training Captain B747-400, 22 July 2013). The non-adherence to standard phraseology by pilots and ATC is seen as equally as dangerous as this can and has often led to confusion and fatal accidents, in particular in the case of line-up and take-off clearances, as was the case with KLM and PanAm in Tenerife in 1977. The existence of several phraseology standards brings with it the well-known psychological phenomenon that when facing situations in which human beings have to perform under stress (e.g. emergencies) they may revert back to old behaviour (cf. Reason, 1988; Bourne and Yaroush, 2003; Covelli, Rolland, Proctor, Kincaid and Hancock, 2010), i.e. to what they know best or have learnt first. This can even be the case despite regular emergency training.

Since voice communications are beset with many problems and so far the CPDLC system does not seem to have improved ATC communications significantly, it is vital that pilots and controllers be given an additional communication method which has the potential of reducing all of the above-mentioned communication issues. The solution that we propose intends to use automatic speech recognition (ASR), machine translation (MT) and term extraction (TE) technology.

THE POTENTIAL USE OF LANGUAGE ENGINEERING TECHNOLOGIES FOR REDUCING ATC MISCOMMUNICATIONS

Automatic Speech Recognition

The goal of automatic speech recognition (ASR) is to transfer speech to text. Spoken words are converted into written words by decomposing a word into smaller units of speech. These then form the basis for speech recognition algorithms which change 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 is still difficult, but lately progress has been made for ASR to work well in a variety of areas, e.g. human-computer interaction, such as the BMW voice control system (2014), dictation, but also in the area of real-time/live subtitling using respeaking technology (e.g. Romero-Fresco, 2011). ASR systems for use in aircraft have already been investigated for a number of years but these studies have mainly focused on applying ASR in voice input systems (e.g. Simpson, Mccauley, Roland, Ruth and Williges, 1985; Baber and Noyes, 1996; Lennertz *et al*., 2012; 2013). 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). More complex ASR technology, and of particular interest here, is employed for training controllers (Karlsson, 1990; Klie, 2010; Cordero *et al*., 2012), e.g. the FAA uses ASR in their ATC simulators.

The ideal performance of an ASR system would be if it recognised everything a speaker said without any time delay (real time factor, RTF) and with a 0% word error rate (WER) irrespective of vocabulary size, prosody, accents, dialects and channel/noise issues. In reality, however, the existence of these bottleneck areas means that the performance levels in terms of accuracy and speed will be affected, but it is nevertheless possible to raise them substantially by applying certain constraints. For example, by using smaller vocabulary sizes, reduced noise, high-quality channels, and training the systems on individual speakers[5] – all of this enables high accuracy. Since ATC phraseology is a sublanguage with reduced vocabulary and with altered pronunciation, and since the ASR system would be trained specifically by each user, the performance levels of ASR would be expected to be high.

Machine Translation

Machine translation (MT) performs translation from one language to another without human intervention (Forcada, 2010). According to the European Association for Machine Translation (EAMT), “today a number of systems are available which produce output which, if not perfect, is of sufficient quality to be useful in a number of specific

[5] Modern ASR systems are trainable as they are based on statistics (e.g. hidden Markov model, maximum entropy model).

domains” (2013:1). Nowadays there is also increasing demand for integrating MT technology with other language engineering technologies, for instance, with speech input/output systems – an integration of technologies which we would also like to achieve in order to remedy the continuing ATC communication issues.

Like ASR, the quality of MT depends on many problems inherent in language and speech. MT performance increases the more a language is reduced in terms of 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). ATC communications constitute a similarly restricted language and delimited domain. 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. One of the most well-known hybrid machine translation engines is Systran, which combines a rule-based approach to machine translation (RBMT) with a statistical approach to machine translation (SMT). The RBMT part enables high-performance consistency and predictability of translations as well as high adherence to terminology and phraseology, which would be a salient requirement for ATC communications. The SMT approach would suit the machine translation of ATC messages because they constitute grammatically reduced and relatively short textual units. The SMT method is also particularly suitable for training MT systems. The main advantage of this combined approach is that within a delimited domain the quality of translations is likely to be high.

Terminology Extraction

Terminology/term extraction can be defined as the automatic identification of term candidates included in a text or text corpus in specialist domains. Traditionally, term extraction software has been developed for specific linguistic needs, including glossary creation, knowledge representation, and aiding human and machine translation (cf. Vivaldi and Rodríguez, 2007). Approaches to term extraction include linguistic methods (e.g. terms are identified by their morpho-syntactic patterns) as well as statistical methods (e.g. terms/phrases having a certain frequency-based statistic higher than a given threshold are extracted; n-gram-based approaches). More recent tools employ hybrid methods (linguistic and statistical). Term extraction can take place on its own but also automatically as part of the machine translation process and, as has been mentioned above, the performance levels are likely to be high when applied to delimited domains in which restricted language is used. This would be the case in ATC communications.

The Proposed ASR-MT-TE Communication System

Key findings in the areas of ASR and MT show that the performance levels of both types of system depend on how issues such as dialects, accents, prosody, channel, noise, vocabulary size, domain delimitation, grammar differences, and linguistic devices are dealt with. Consequently, the task of creating an ASR-MT-TE communication system which produces results of the highest quality will not be easy. However, as we have seen, both technologies as well as TE perform well with standardised and restricted sublanguages, to which ATC phraseology belongs. We hope that the proposed system may be able to reduce the aforementioned ATC communication issues as follows:

Addressing Problem (1)	Whenever local languages are used by pilots and controllers the MT system will enable translations into English in real-time.
Addressing Problems (2), (3)	State-of-the-art ASR systems are trainable at source on individual speakers, which should address pronunciation, enunciation, and prosody issues.
Addressing Problems (4), (5)	The ASR system will convert spoken English ATC messages into text in real-time using speech recognition technology in order to enable pilots and ATC to spot non-standard phrases. Safety-critical terminology will be highlighted specifically as high-priority information using TE technology.

When designing the proposed communication system, it has to be borne in mind that the system is intended as an additional, i.e. back-up, communication system to voice communications, thus contributing to the level of system redundancy in flight decks (e.g. a Boeing B777-200 has three autopilots and three artificial horizons) and ATC workstations. The new system should provide pilots and controllers with textual versions of transmitted voice messages using data link. For example, for every voice ATC message to a specific aircraft a transcribed text version would be displayed in real-time on a screen on the flight deck of that aircraft. At the same time, any other aircraft which is in the same airspace on the same radio frequency (or happens to be on the ground at the same airport) and which has the ASR-MT-TE communication system installed would also receive the text version of that particular ATC message via data link on their displays, just like they can hear and listen to ATC messages intended for other aircraft. If, for example, a local language is used during ATC communications, the MT system would get activated as well and pilots and controllers would receive the English translations of what has been transmitted in the local language on their respective displays in real-time. In addition, any safety-critical terms would be highlighted in a suitable manner via the TE system. Possibilities here include intensity or colour coding (cf. Wickens and McCarley 2008). Pilots from other aircraft would therefore always be able to understand ATC transmissions in a local language between, for example, a neighbouring aircraft and ATC and be in the picture about what that specific aircraft is doing. In other words, pilots and controllers would benefit from improved situational awareness.

The ASR-MT-TE Communication System on the Flight Deck

It will be up to the pilots to decide whether they need to look at the transcribed and/or translated ATC message displayed on the screens or not. For instance, as long as a pilot dealing with ATC communications (usually the pilot not flying) hears ATC phraseology in English as well as in standardised form, s/he will likely not need to look at the texts on the display. However, if local languages are used on the radio, the pilot not flying will probably want to read the ATC instructions that have been translated into English on the display. This may even be the case during high workload phases of flight in order to find out what is going on, for example, in the airspace around him/her, or on the ground when taxying, and especially any time near the runways. Similarly, even if ATC communications are conducted in English it may be that the pilot not flying detects some non-standard phrases transmitted by ATC or other pilots. Such non-standard phrases can then be double-checked on the display.

In fact, it is particularly important during times of high workload (which usually constitute safety-critical situations) that ATC communications are accurate and, in order to ensure this, that the pilot not flying has the option to read the textual transcripts/translations of a transmitted message if deemed necessary. The ASR-MT-TE communication system may therefore have the potential to reduce situations of high workload since pilots would have more time to aviate and navigate (cf. Owens, 2013; Barshi and Farris, 2013) rather than having to spend time to double-check ATC messages with the respective controllers or having to waste valuable time by trying to discuss among each other what the controller has said. When workload increases, communication tends to disintegrate (cf. Billings and Cheaney, 1981; Barshi and Farris, 2013). For example, it has been reported that pilot-controller communications decrease in length under high workload (Raby and Wickens, 1994), which may be seen as a result of restrictions on a human being’s time-sharing ability that underlie complex performance (Jennings and Chiles, 1977). How would this time-sharing ability then get affected by the additional transcripts/translations in the flight deck? According to Wickens’ 4-D multiple resource model (2008), the perception of auditory and visual information takes place 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 ATC voice messages as well as their transcripts/translations in the flight deck would be an improved situation for pilots. Since the terminal phases of flight are those with the highest workload, the fact that ATC messages are textually displayed in addition to voice communications may lower the workload in such phases because the pilot not flying 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. It goes without saying that appropriate experiments to substantiate the above claims will have to be carried out.

As mentioned earlier, the potential use of automatic speech recognition/input in ATC communications is in the process of being explored in a number of studies. For example, 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. Geacăr points out, just as we have done earlier, that the use of speech input would mean building in a level of redundancy into ATC communications since so far “communications were one of the few aircraft systems without redundancy” (2010: 1). He also explores speech input for producing voice commands to the data link system with the intention of lowering pilot and controller workload since text could be entered without using a keyboard. Introducing yet a further method of communicating ATC messages, Lennertz *et al*. (2012, 2013) have experimented with data link messages that are accompanied by corresponding synthetic speech outputs in addition to traditional ATC radio communications with a view to reducing head-down time during single-pilot operations.

In contrast to the communication system that we propose, it is clear that the above approaches would be unable to deal with local language use. 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. Although Lennertz *et al*. conclude that adding synthetic speech commands to textual data link messages did not "introduce additional complications" (2012: 31), it does not seem to be well motivated to add new spoken commands to existing voice communications. Since their experiment involved single-pilot operations, this combined speech-text method may have its benefits there. However, it remains to be seen how synthetic speech outputs of data link text messages in addition to the usual voice communications will fare in two-pilot operations, which are far more complex in terms of pilot tasks and aircraft systems. Moreover, in two-crew aircraft, CRM (Crew Resource Management) requires pilots to communicate with each other to a significant extent, and interruptions from a synthetic speech output could potentially disrupt their workflow and add to their workload, e.g. when reading checklists.

The ASR-MT-TE Communication System in the ATC Workspace

ATC workstations are operated by various types of controller (cf. Nolan, 2011). *Flight data controllers* aid other controllers in a tower and carry out clerical duties. Their main duties include the passing on of information to other controllers, e.g. IFR (Instrument Flight Rules) departure clearances, weather information, and so on. Among the duties of *clearance delivery controllers* is the reception, passing on, and amending of departure clearances to pilots. *Ground controllers* give instructions to aircraft taxiing to and from runways, whereas *local controllers* are responsible for sequencing arriving and departing aircraft while observing appropriate runway separation. What these types of controller have in common is that they all work with similar equipment, including communication, surveillance, and ancillary equipment, and so on. As in flight decks, there is a certain level of system redundancy built into ATC workstations, e.g. with regard to the communication system, server, and displays. Hence, the proposed ASR-MT-TE system would also add another level to this redundancy.

During times of high workload, controllers may filter messages from aircraft (Airbus, 2014), meaning that not all communications addressed to them may be listened to or acknowledged if they are, for example, talking to another aircraft. Hence, the additional transcripts/translations on a suitable display in the workstation would give them the chance to read a message when they are finished speaking to the other aircraft. Hence, the role of ASR-MT-TE systems 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 altitude or speed, irrespective of dialects, and so on. This in turn should improve the situation awareness of controllers and enable the workload to be managed appropriately.

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) and that they will not add to the controller's workload but rather reduce it.

Examples of Miscommunications

In ATC communications, errors between controller and pilots can cause numerous safety occurrences. The types of errors most prevalent in ATC are:

1. Readback/Hearback errors Type I – For instance, the pilot reads back the clearance incorrectly and the controller fails to correct the error (expectation plays a crucial role in hearbacks, cf. Hawkins 2010).
2. Readback/Hearback errors Type II – The controller fails to notice his/her own error in the pilot's correct readback or fails to correct critical erroneous information in a pilot's statement of intent.
3. No pilot readback.

Figure 1 below shows the pilot-controller communication loop which, as long as pilots and controllers follow it, will help to avoid miscommunications.

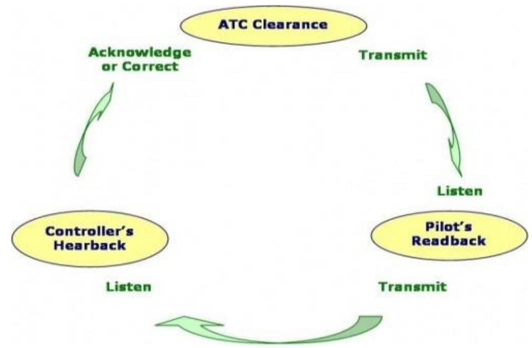

Figure 1. Pilot-Controller Communication Loop (taken from Skybrary, 2013b; Airbus, 2014)

In what follows, examples of ATC miscommunications are presented and discussed with a view to seeing whether they could have been prevented by the use of an ASR-MT-TE system on the flight deck and in ATC workspaces. The data were obtained from two different reports – one reporting on an accident and the other on a serious incident.

(1) Flying Tiger Line, FT66, 19 February 1989, accident with fatalities (Aviation Safety Network, 1989).
During the descent to Kuala Lumpur, the crew was cleared to route directly to the Kayell (KL) beacon. ATC transmitted to the crew, “Tiger 66, descend two four zero zero”, meaning that the pilot should descend to an altitude of 2400 feet amsl (above mean sea level). The captain, who had understood “descend to four zero zero”, read back “Okay, four zero zero”. This meant that he was planning to descend to 400 feet amsl – 2000 feet too low. The controller did not pick up on the wrong altitude in the captain’s readback. Also, in his transmission, the controller had not used standard ATC phraseology. Instead of transmitting “descend two four zero zero”, he should have said “descend and maintain two thousand four hundred feet”. In addition, the captain also used non-standard ATC phraseology when replying “okay, four zero zero” whereas the correct standard phrase should have been “Roger, descend and maintain four-hundred feet”, which should have allowed the controller to notice the incorrectly read back altitude. The Boeing 747-249F eventually descended below minimum altitude and crashed into a hillside at ca. 600 feet amsl just before reaching the Kayell beacon, where minimum descent height has to be 2400 feet. There were no survivors. The cockpit voice recorder revealed several other communication errors made by the pilots before the miscommunication about the descent altitude occurred as well as confusion regarding the acronym ‘KL’ used to refer to the Kayell beacon. This code was also employed by local ATC to refer to Kuala Lumpur instead of using the full expression ‘Kuala Lumpur’. There were also other errors committed by the flight crew, such as completely ignoring the continuous warnings of the on-board Ground Proximity Warning System (GPWS). The investigators concluded that this accident was mainly caused by the crew’s lack of abiding by the correct instrument approach procedure, by the captain’s inadequate crew resource management, and the pilots’ diminished situation awareness. However, the fact that non-standard ATC phraseology by both the pilots and local ATC was used was also seen as a major contributing factor.

(2) Runway incursion during take-off at Paris-CDG, 10 January 2010, serious incident (BEA, 2007).
The incident occurred between a British-registered airliner (Airbus A321) and a French-registed airliner (Airbus A340). A third Airbus A320 had made an emergency landing on runway 27L at Paris-CDG. It had exited the runway via a taxiway and was being inspected by emergency vehicles. Meanwhile, the British A321 had landed on the parallel runway 27R and was instructed to vacate this runway via a taxiway. Simultaneously, the French A340 was at the threshold of runway 27L awaiting take-off clearance. The controller then gave take-off clearance to the A340 which was issued in French. At the same time the British A321 was cleared to cross runway 27L – the same runway on which the French A340 was in the process of taking off. Since the take-off clearance was given in French, the British A321 crew was unaware that they had now been put into a runway incursion situation. The automated airport collision alarm sounded and the controller realised his mistake and ordered the French A340 to abandon its take off. The French A340 was able to stop in time to avoid a collision. Apart from various other factors that contributed to this incident, such as that the controller was distracted by the emergency landing of the A320, the way strips boards were managed by ATC, how resources were managed in the ATC team, and so on, it was noted that the use of French during ATC communications between flight crews and ATC does not allow crews who are unable to understand and speak French to have adequate situational awareness of the traffic situation close to and on the runway or allow them to able to spot any ATC

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2097-8

errors. It was also mentioned by the French air accident investigation bureau that “the French-speaking crews were less attentive to messages given in English, because they were not addressed to them” (BEA, 2007: 4). Figure 2 below illustrates the positions of the aircraft relative to each other just before the incursion occurred.

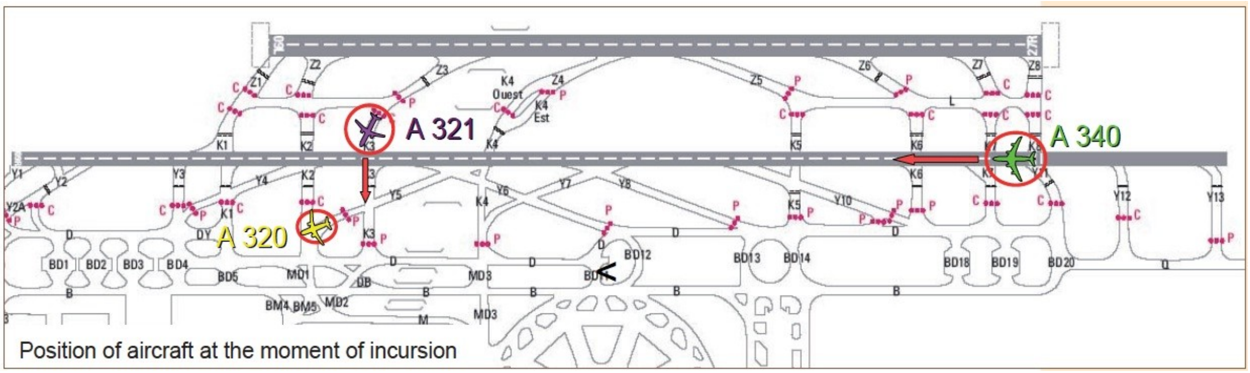

Figure 2. Paris-CDG runway chart (taken from BEA, 2007)

Would the Proposed Communication System have made a Difference?

Had the ASR-MT-TE system been implemented in the flight deck of the Flying Tiger aircraft as well as on suitable screens in ATC, it might have been possible for both the crew and the controller to notice that non-standard phraseology was used. The first phrase used by the controller “Tiger 66, descend two four zero zero” would have likely been rendered correctly by a trained ASR system. The system would have shown this message via speech recognition technology on a suitable screen in the flight deck, which would have given the crew an opportunity to spot that the intended descent altitude would have been 2400 feet, and not 400 feet, as the captain had understood. The discrepancy between what the captain had heard and what he would have seen on the screen might have shifted the attention of the crew and given them the chance to double-check the altitude with ATC. In case the crew had ignored this message on the screen, there would have been a second chance to pick up on the altitude error during the second transmission when the captain read back “Okay, four zero zero” since this message would have been shown on an ATC screen and the controller might have picked up on the wrong altitude when looking at the display.

There is obviously no guarantee that the Flying Tiger pilots would have looked at the textual transcripts as they were likely overloaded with the situation in the flight deck, which is indicated by the fact that the continuous alarms from the Ground Proximity Warning System were completely ignored. However, it is argued that the availability of ATC message transcripts on screens in the flight deck would have at least provided the pilots with additional opportunities to regain their situational awareness. Also, the controller would have received the messages on his/her screen. The fact that the system is meant to work simultaneously in both workspaces, i.e. both the pilots and the air traffic controllers have textual transcripts at their disposal, may mean that this type of communication system has the potential to break the chain of causative events during miscommunications.

In the case of the runway incursion, the MT engine would have been activated in addition to the ASR system and would have supplied the pilots of the British A321 aircraft with English translations of the French ATC messages on a screen. Moreover, the TE technology would have also kicked in and highlighted the safety-critical terms within the transcript, hence providing the pilots with two further means of noticing what is going on around them. This would have instantly increased the British pilots’ situational awareness around runway 27L and allowed them to question ATC regarding the appropriateness of their runway crossing clearance. The controller would have also received the message on his/her screen and could have noticed the error this way. This is particularly crucial since many airports do not have automated airport collision alarms. Even if the translated ATC message had not been grammatically and semantically perfect, it is likely that the key terms contained in the ATC message would have been translated and highlighted correctly. In this example, the safety-critical terms *take off* and *runway 27L* would have alerted the pilots in the British A321 of a potential conflict and given them the chance to query ATC.

HUMAN FACTORS CONSIDERATIONS

Flight decks and ATC workstations are environments in which very high workloads can be experienced at times, e.g. during terminal phases of flight, certain times of day, and emergency situations. It is therefore vital that the proposed ASR-MT-TE communication system is integrated into both work environments in such a way that it will not negatively impact on the mental workload, attention allocation, and situation awareness of both groups of professionals. For example, today's highly automated cockpits are designed to reduce workload but pilots are still fully responsible for the automated tasks, procedures and processes (Wickens and McCarley, 2008). As a result of this automation, pilots may not allocate much attention to the automated tasks and procedures with the effect that this may cause reduced situational awareness. As Wickens and McCarley point out, such a situation could have disastrous consequences when the automated processes stop working. At this stage of the research we can only speculate on how the communication system presented here will impact on the pilot's highly automated workplace and the controllers work environment in terms of mental workload, situation awareness and attention allocation.

As we have seen, in the case of pilots, it is not uncommon for them to be uncertain about the content of ATC messages and quite a considerable amount of time is spent discussing their meanings with each other, which invariably requires querying ATC again. The textual transcripts may thus enable pilots to save valuable time since, as has been mentioned earlier, being able to read ATC messages on a screen in the flight deck is a task that is perceived by pilots as being faster than having to ask the controller again for clarification. Hence, the intention to give pilots under heavy workload the option to read text, thereby allowing them to focus on aviating and navigating, seems well motivated. As a direct result of having ATC messages at their disposal at all times, a pilot's situation awareness would be increased throughout, in particular in situations where ATC or pilots from other airlines use local languages. It should be noted again that even if the translation of a foreign-language ATC message may be stilted or slightly confusing, or even if a non-standard ATC phrase results in a rough or incomplete transcript, such results can nevertheless be seen as an improved communication situation. Pilots would get at least an idea as to what was said by looking at the transcripts, which is more than they would have without them. For example, the idea that in such cases it would be useful to have any safety-critical terms that are contained in an ATC message highlighted comes from pilots themselves. According to them, the highlighted terms alone (e.g. *runway*, *holding position*, *take-off*, *descend*, *climb*, *turn*, *heading*, *landing clearance*, and so on) would already allow them to gain better situation awareness (personal communication with a UK Training Captain B747-400, 01 June 2013). In terms of attention allocation, pilots may have to be alerted to the transcripts of incoming ATC messages if any of these contain specific safety-critical terminology. In order to make sure that the usual order of attention scanning is interrupted so that the new and important information is processed instantly (Wickens and McCarley, 2008) the use of a cue may have to be considered for directing the attention of pilots to the respective screen in a manner that is as little disruptive as necessary. For example, depending on what works best in a busy and noisy cockpit environment visual or audible cues may need to be implemented into the design of the proposed system (cf. Wickens and McCarley, 2008).

In the air traffic controller's workstation, considerations of workload, situation awareness, and attention allocation are equally as important. The use of the ASR-MT-TE communication system will provide improved situation awareness for controllers as the lack of ambiguity in communications will greatly enhance the accuracy of the information provided to controllers. 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. The potential of the technology is to reduce ATC communication errors.

CONCLUSIONS AND FUTURE DEVELOPMENTS OF THE WORK

In this paper we presented a snapshot of our initial work as well as thoughts on its future development, including a description of how an ASR-MT-TE communication system might be designed and implemented into the flight deck and ATC workspaces and how this system may impact on mental workload, situation awareness, and attention allocation of pilots and air traffic controllers respectively. Nevertheless, there are many more questions that need to be asked. Among these are, for example, how this system would function in the adverse environment of a cockpit in terms of high noise, vibration levels, or when stressful speech is present. The question of what type of experiments would be best in order to test the human factors impact of this system would also have to be addressed. Another question is whether to use existing commercial ASR and MT technologies or whether to design these from scratch, and also to look into integrating artificial neural networks into ASR.

This research idea is currently in its development stages. At present, there are considerations to break down the creation of the ASR-MT-TE communication system into smaller constituent systems, for example, to develop the speech recognition system for English ATC messages first, which would at least catch non-standard phraseology, before the MT and TE parts are envisaged. We are also in the process of establishing partners to the research and preparing a submission for funding for a pilot study which may include the creation of a prototype system. The pilot study will examine ATC miscommunications in depth, establish what solutions pilots and controllers would like to see implemented, gauge their thoughts on the proposed communication system, examine current studies which already consider the use of speech input in flight decks, and survey the existing technologies for ASR, MT, and TE.

It is hoped that this research into an ASR-MT-TE system, or AICSys (**A**dvanced **I**ntelligent **C**ommunication **Sys**tem) as we may call it, will help to reduce the currently known types of ATC communication errors. In fact, the development of such a system would be crucial for this purpose since present and future data link communications technology will be unable to deal with local languages during ATC communications. We also believe that the envisaged voice input from pilots into the future data link communication system in addition to voice communications (although this will be reduced to a large extent) could be distracting and time-consuming for pilots.

REFERENCES

Airbus (2014), Flight operations briefing notes: Human performance – Effective pilot/controller communications. Available at: http://www.airbus.com/company/aircraft-manufacture/quality-and-safety-first/safety-library/. Accessed: 30 March 2014.

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

Aviation Safety Network (1989), “Accident description – Flying Tiger Line, Flight 66”. Available at: http://aviation-safety.net/database/record.php?id=19890219-0. Last accessed: 28 March 2014.

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

Barshi, I. (1997), *Effects of linguistic properties and message length on misunderstandings in aviation communication* (Doctoral dissertation). Boulder: University of Colorado.

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

BEA – Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (2007), Incidents in air transport: Runway incursions. November, No. 8. Le Bourget, France: BEA.

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

Bourne, L. E., Jr., Yaroush, R. A. (2003), Stress and cognition: A cognitive psychological perspective. Final Report, National Aeronautics and Space Administration, Grant Number NAG2-1561.

BMW (2014), BMW technology guide: Voice control system. Available at: http://www.bmw.com/com/en/insights/technology/technology_guide/articles/voice_control_system.html. Last accessed: 30 March 2014.

Cardosi, K., Falzarano, P., Han, S. (1998), Pilot-controller communication errors: An analysis of aviation safety reporting system (ASRS) reports. Available at: www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA356655. Last accessed: 30 March 2014.

Cordero, J. M., Dorado, M., De Pablo, J. M. (2012), Automated speech recognition in ATC environment, ATACCS’2012, Research Papers, London. pp. 46-53. Available at: www.hala-sesar.net/sites/default/files/documents/p46-cordero.pdf. Last accessed: 30 March 2014.

Covelli, J. M., Rolland, J. P., Proctor, M., Kincaid, J. P., Hancock, P. A. (2010), Field of view effects on pilot performance in flight. *International Journal of Aviation Psychology*, 20 (2), 197-219.

Crystal, D. (1992), *A dictionary of linguistics and phonetics*. 3rd edition. Oxford, UK: Blackwell Publishers.

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: P. Quigley and P. McElwain (Eds.), *Selected Proceedings of the 1997 Symposium ‘Aviation Communication’*, 9-11 April 1997, Prescott, AZ: Embry-Riddle Aeronautical University. pp. 21-23.

Eurofighter Typhoon (2014), Technology: Cockpit. Available at: http://typhoon.starstreak.net/Eurofighter/cockpit.html. Last accessed: 29 March 2014.

Forcada, M. L. (2010), Machine translation today, in: Y. Gambier and L. Van Doorslaer (Eds.), *Handbook of translation studies*, Vol. 1, Amsterdam/Philadelphia: John Benjamins Publishing. pp. 215-223.

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. Available at: www.icas.org/ICAS_ARCHIVE/ICAS2010/PAPERS/441.pdf. Last accessed: 30 March 2014.

Gotti, F., Langlais, P., Lapalme, G. (2013), Designing a machine translation system for Canadian weather warnings: A case study. *Natural Language Engineering*, 1(1), 1-36. Available at: http://rali.iro.umontreal.ca/rali/sites/default/files/publis/DesigningWMT-NLE.pdf. Last accessed 27 March 2014.

Hawkins, F. H. (2010), *Human factors in flight*. 2nd edition. Harry W. Orlady (Ed.). Farnham, UK: Ashgate Publishing.

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

ICAO (2007), *Manual of radiotelephony Doc 9432 AN/925*. Montreal: International Civil Aviation Organization.

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

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

Karlsson, J. (1990), The integration of automatic speech recognition into the air traffic control system. FTL Report R90-1, Flight Transportation Laboratory Department of Aeronautics and Astronautics, M.I.T. Cambridge, MA. Available at: http://dspace.mit.edu/bitstream/handle/1721.1/68122/FTL_R_1990_01.pdf?sequence=1. Last accessed: 28 March 2014.

Klie, L. (2010), Speech takes flight among pilots and tower personnel. *Speech Technology*, May/June 2010, 12.

Lennertz, T., Bürki-Cohen, J., Sparko, A. L., Macchiarella, N., Kring, J., Coman, M., Haritos, T., Alvarado, J. (2012), NextGen Flight Deck Data Comm: Auxiliary Synthetic Speech – Phase I, in: Proceedings of the Human Factors and Ergonomics Society, 56, pp. 31-35. Available at: http://pro.sagepub.com/content/56/1/31.full.pdf+html. Last Accessed 30 March 2014.

Lennertz, T., Bürki-Cohen, J., Sparko, A. L., Macchiarella, N., Kring, J., Coman, M., Haritos, T., Alvarado, J. (2013), *NextGen Flight Deck Data Comm: Auxiliary Synthetic Speech – Phase I*, Full project report, 248 pages, US Department of Transportation/Federal Aviation Administration/Embry-Riddle Aeronautical University. CreateSpace Independent Publishing Platform.

Linde, C., Goguen, J. A., Devenish, L. (1986), Aircrew communicative competence: Theoretical and pragmatic aspects of training design. Moffett Field, CA: NASA Ames Research Center.

Mosier, K. L., Rettenmaier, P., McDearmid, M., Wilson, J., Mak, S., Raj, L., Orasanu, J. (2013), Pilot-ATC communication conflicts: Implications for NextGen. *The International Journal of Aviation Psychology*, 23 (3), 213-226.

Nolan, M. S. (2011), *Fundamentals of air traffic control*. 5th edition. Delmar/Cengage Learning.

O'Neil, K. (1999), Improving the safety of communications. Available at: http://www.aatl.net/publications/antiblocking.htm. Last accessed: 30 March 2014.

Owens, D. (2013), The golden rules for pilots: Moving from PNF to PM. *Safety First: The Airbus Safety Magazine*, Issue 15, 5-7.

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

Reason, J. (1988), Stress and cognitive failure, in: S. Fisher and J. Reason (Eds.), *Handbook of life stress, cognition and health*. Chichester: John Wiley & Sons Ltd.

Reason, J. (2008), *The human contribution: Unsafe acts, accidents and heroic recoveries.* Farnham, UK: Ashgate Publishing Ltd.

Robertson, F. (1997), Aviation English as a foreign language for pilots: English language training practice for French professional pilots, in: P. Quigley and P. McElwain (Eds.), *Selected Proceedings of the 1997 Symposium 'Aviation Communication'*, 9-11 Apr. 1997, Prescott, AZ: Embry-Riddle Aeronautical University. pp. 45-49.

Romero-Fresco, P. (2011), *Subtitling through speech recognition: Respeaking*. Manchester/New York: St. Jerome.

Schneider, V. I., Healy, A. F., Barshi, I., Kole, J. A. (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. Available at: http://www.afmc.af.mil/news/story.asp?id=123071564. Last accessed: 30 March 2014.

Simpson, C. A., Mccauley, M. E., Roland, E. F., Ruth, J. C., Williges, B. H. (1985), System Design for Speech Recognition and Generation. *Human Factors*, 27 (2), 115-141.

Skybrary (2013a), Non-standard phraseology. Available at: http://www.skybrary.aero/index.php/Non-Standard_Phraseology. Last accessed: 30 March 2014.

Skybrary (2013b), Pilot-controller communications. Available at: http://www.skybrary.aero/index.php/Pilot-Controller_Communications_(OGHFA_BN). Last accessed: 30 March 2014.

Vivaldi, J., Rodríguez, H. (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.

Wickens, C. D., McCarley, J. S. (2008), *Applied attention theory*. Boca Raton/London/New York: CRC Press.