

Errors in air-ground pilot communication: an experimental study

Erros do piloto na comunicação ar-terra: um estudo experimental.

Dominique ESTIVAL (Western Sydney University)¹
Brett MOLESWORTH (University of New South Wales)²

ABSTRACT

Understanding aviation communication has obvious implications for flight safety and for the training of pilots and Air Traffic Controllers (ATC). Its study also sheds light on the use of Aviation English as a type of English for a very specific purpose and on real-world communication constraints in a high-stakes environment. We present results from a study combining human factors and linguistics, in which pilots were recorded in flight simulator experiments probing communication accuracy under diverse conditions. An analysis of the audio recordings in terms of whether errors were omissions or mistakes, whether errors occurred with words or numerals, and which types of aviation information were affected, revealed complex interactions between language background, pilot qualification, workload and ATC speech rate. The differences between native English speakers (NES) and non-native English speakers (NNES) are complex and more nuanced than might be expected. These results not only are relevant for aviation safety, training and testing, but they also provide insights on communication under challenging conditions and on the relation between language proficiency and cognition, with implications for second language teaching and the use of English for Specific Purposes (ESP) in highly constrained environments.

Keywords: Aviation English, English as a second language, communication errors

RESUMO

Compreender a comunicação para aviação tem implicações evidentes para a segurança operacional do voo, e para a capacitação de pilotos e Controladores de Tráfego Aéreo (ATC). Seu estudo também ajuda a compreender o uso do Inglês para Aviação como um tipo de inglês para um fim muito específico e as restrições da comunicação no mundo real em ambientes de alto risco. Apresentamos os resultados de um estudo que combina fatores humanos e linguística, no qual foi gravada a interação de pilotos em experimentos de simuladores de voo para verificar a precisão da comunicação em condições adversas. A análise de gravações de áudio para averiguar se os erros foram omissões ou enganos, se os erros ocorreram com palavras ou números, e quais tipos de informações de aviação foram afetadas revelou interações complexas entre o arcabouço linguístico, a qualificação do piloto, a carga de trabalho e a velocidade de fala dos ATC. As diferenças entre os falantes nativos do inglês (NES) e os falantes não-nativos do inglês (NNES) são complexas e com mais nuances do que o esperado. Esses resultados são relevantes não apenas para a segurança operacional, a capacitação e a avaliação no âmbito da aviação, mas também para promover perspectivas sobre a comunicação em condições

¹ Western Sydney University, MARCS Institute, Australia. <https://orcid.org/0000-0002-6178-3825>;
d.estival@westernsydney.edu.au.

² University of New South Wales, School of Aviation, Australia. <https://orcid.org/0000-0001-5252-5130>;
b.molesworth@unsw.edu.au.

desafiadoras, e sobre a relação entre proficiência linguística e cognição, com implicações para o ensino de segunda língua e o uso de Inglês para Fins Específicos (ESP) em ambientes altamente restritivos.

Palavras-chave: *Inglês para aviação, inglês como segunda língua, erros de comunicação*

1. The language of aviation and communication errors

Aviation is one of the safest modes of transportation but incidents still occur, with miscommunication often cited as a contributing factor in both commercial and general aviation. For instance, when a McDonnell Douglas MD-87 collided during take-off with a Cessna Citation 525-A at Milano Linate airport in 2001, the Cessna pilot who entered the runway without clearance seems to have been confused by the instructions given by ATC (AGENZIA NAZIONALE PER LA SICUREZZA DEL VOLO, 2004). The study presented here focuses on the types of error found in air-ground communications, between pilots and Air Traffic Control (ATC), and more specifically in single pilot general aviation operations. While the term ‘aviation communication’ refers to any type of communication used in aviation, including light signals, marshalling gestures or navigation aids, the term *aviation language* refers to the use of a natural language – English among others – in aviation (ESTIVAL; FARRIS; MOLESWORTH, 2016; BOROWSKA, 2017). We are here interested in communication with language, but it is important to note that some information elements communicated in aviation can be transmitted via other means – for instance, the current altitude is also transmitted via the transponder. Information ‘items’ such as altitude, heading or call-sign are the basic building blocks of aviation communication.

A number of factors make communication more difficult in the aviation environment: noise, time pressure, quantity of information, and radio physical constraints. Because communication is crucial to aviation safety, aviation communication was standardised and codified very early to mitigate those factors and to avoid ambiguity. With the increased internationalisation of aviation, the international body responsible for aviation safety, the International Civil Aviation Organization (ICAO), established English as the international language of aviation in the late 1940s (BIESWANGER, 2016; MODER, 2013). Following the widely reported *communication clashes* of Cushing (1994), English Language Proficiency was mandated for all aviation professionals from 2003 (ICAO, 2016). Importantly, Aviation English covers not only “communications which involve ‘standard phraseology’ – a prescribed, highly constrained set of phrases to be used insofar as possible in all radiotelephonic communications between controllers and pilots – but also communications which involve the use of natural English or ‘plain

language” (ESTIVAL; FARRIS; MOLESWORTH, 2016; p. 2). Aviation English is thus one variety of English for Specific Purposes (ESP).

We do not attempt to present here a complete overview of the substantial research over the past decades investigating the issues of language-related miscommunication in aviation (see *inter alia* GOGUEN; LINDE, 1983; CARDOSI, 1993; MORROW; LEE; RODVOLD, 1993; MODER, 2013; PRINZO, 2008; MORROW; RODVOLD, 1998; BIESWANGER, 2013). From an ESP perspective what is important is that since 2003 many countries have implemented English language training and testing but that English proficiency can still be an issue. For instance, after a China Eastern Airlines Airbus A330-200, with 245 people on board, took off without clearance from Osaka Kansai in 2011, the airline announced it would improve the English training of its pilots, prompting speculations about the causes of the misunderstanding.³ In spite of ICAO’s efforts, the ability of pilots to use Aviation English is still questioned and has been the subject of a number of studies and reports (e.g. EUROCONTROL, 2006b; KIM; BILLINGTON, 2016; KIM; ELDER, 2009; TIEWTRAKUL; FLETCHER, 2010). Looking at commercial aviation, Tiewtrakul and Fletcher (2010) found that pilots failed to understand 79 out of 312 ATC messages at Bangkok International Airport, and that miscommunication between Thai ATC and non-native English speaking (NNES) foreign pilots was consistently higher than with Thai or native English speaking (NES) pilots. In General Aviation (GA), after a series of accidents and incidents at training aerodromes, the Australian Civil Aviation Safety Authority (CASA) undertook a review which found that a concerning number of these events had involved NNES pilots (CASA, 2009a, 2009b). The CASA review identified communication issues as a contributing factor, citing instances of NNES pilots not making required radio calls at certain aerodromes; our own research – presented in Estival and Molesworth (2009); Estival and Molesworth (2012); Molesworth and Estival (2015); and Estival and Molesworth (2016) – confirmed the prevalence of miscommunication in Australian GA. The study presented here, based on experiments with pilots in a flight simulator, provides support for these earlier findings but, more importantly, it contributes new findings about the communication errors made by both NES and NNES pilots and it adds nuance by considering the pilots’ flight qualification.

Just as ESP does not only deal with NNES issues, this paper is not primarily concerned with the differences between NNES and NES pilots. Rather we investigate the type of errors committed by both NES and NNES pilots to draw conclusions and propose recommendations for teaching Aviation English to all pilots. Indeed, while some of our earlier results showed that NNES pilots face different challenges than NES pilots, we show here that the differences in the errors they make diminish with greater level of experience and that some of the conditions examined revealed no differentiation between the two groups

³ <http://www.bbc.co.uk/news/mobile/world-asia-china-15980197>. Last accessed 17/12/2019.

of pilots. One important issue that this paper raises, though, is that NES need to learn the restricted code of Aviation English as much as NNES. Moreover, while ESP research has provided many insights on the structures and conditions of use of English in a wide variety of contexts, such as the legal, business or academic domains (PALTRIDGE; STARFIELD, 2013), the impact of factors such as noise, workload or communication channel has perhaps not been given as much prominence. This study contributes to a better understanding of communication in challenging environments, showing complex relationships between variables such as experience, language background, workload and communication factors. It provides insights on the relation between language proficiency and cognition, with implications for the use of ESP in highly constrained environments.

1.1. Aviation communication

Because only one person at a time can transmit on a radio frequency, dialogue turns between pilots and ATC must be clearly marked, with turns organised in sequences of transmissions (or *calls*). A call consists of a number of *items*, some of which must be read back (in a *readback*) or reported accurately by the pilot.⁴ Example (1)⁵ shows the expected call by a pilot who has been asked by ATC to contact Sydney Centre, and who reports the altitude and heading for aircraft with call-sign ‘ABC’.

(1) Pilot: Sydney Centre. ABC at 5500, tracking 190.

Example (1) contains four information items: name of the station being called (Sydney Centre), call-sign of the station calling (pronounced ‘Alpha Bravo Charlie’), altitude (5500 feet) and heading (190 degrees). Example (2) shows a dialogue turn, comprising an ATC instruction with three items and the pilot’s expected readback.

(2) ATC: ABC. Climb to 2500. Maintain runway heading until advised.
Pilot: Climb to 2500. Maintain runway heading. ABC.

The standard phraseology for aviation communication prescribes not only the vocabulary to be used (e.g. *negative* instead of *no*), but also the pronunciation of certain words (e.g. “TOU-sand” for *thousand*)⁶ and it imposes strict rules for turn-taking (ESTIVAL, 2016). In their *hear back*, ATC check that all the items to be read back are present and correct, and then ask for confirmation if necessary. In cases where no standard phraseology is available, *plain Aviation English* is used. Plain English is not to be understood as unconstrained general English, however, and communications should still be clear, concise and as close to standard phraseology as possible (ICAO, 2016).

⁴ See Estival (2016) for a more detailed description of the language of aviation.

⁵ All the examples in this paper come from the flight simulator experiments described in Section 2.

⁶ The ICAO prescribes pronunciation for numbers, e.g. ‘FOW-er Tou-SAND’ for *4000*. To ensure readability of the examples, we use regular spelling when this does not affect the argument.

1.2. Communication errors

In spite of the highly codified nature of aviation communication, which aims to prevent ambiguity and confusion, errors sometimes occur. We categorise communication *errors* as either *omissions*, i.e. failing to transmit (part of) the intended information or *mistakes*, i.e. transmitting erroneous information. Excluding wilful negligence on the part of the pilot, an omission in a pilot transmission is assumed to result from either failing to hear or to understand an item or forgetting to include it in the readback or report. Failure to hear could result from workload, fatigue or distraction, or from problems with the transmission itself, such as noise interference (MOLESWORTH; BURGESS, 2013; JANG; MOLESWORTH; BURGESS; ESTIVAL, 2014) or transmission overlay (i.e. two radio calls at the same time). Failure to understand could be due to the same reasons but also to factors such as non-standard phraseology, accent or speech rate, or limited Aviation English proficiency (MORROW; RODVOLD, 1998; EUROCONTROL, 2006A; MOLESWORTH; ESTIVAL, 2015; WU; MOLESWORTH; ESTIVAL, 2019). A pilot's request for ATC to repeat a message is a clear indication of a failure to hear or to understand (part of) the message. By contrast, a mistake in producing a transmission is more likely to result from forgetting what was said, confusing information from different sources (cognitive interference), or even slip-of-the-tongue (CUTLER, 1982). Errors due to distraction, workload, confusion or memory lapse all originate in cognition, so analysing their occurrence may shed light on the cognitive aspect of miscommunication.

Multiple errors in a pilot's transmission are evidence of impaired communication. Example (3) from a high workload flight (Flight 3B, see below) shows: the ATC call to a pilot, the expected readback (with optional elements within parentheses) and the actual readback. Pilot 10 made 3 errors: 2 mistakes (Track *180* instead of *270*; Altitude *4500* instead of *5500*) and one omission (Transponder code *5033*).

- (3) ATC: ABC. Sydney Centre. You are identified at 4500 South-West of Wilton. Track 270 and climb to 5500. Squawk 5033 and contact Wilton 126.7.
 Expected: (Track) 270, (Climb to) 5500, (Squawk) 5033, (Contact) Wilton 126.7. ABC.
 Actual (P10): 4500 feet, 180 degrees, 126.7 at Wilton, we will be...tracking to Bathurst due to weather.

By looking at errors in pilots' transmissions under different conditions, our research aims to better understand the factors contributing to communication errors in aviation. This will allow us to propose ways to mitigate their impact in potentially dangerous situations, such as pilots misunderstanding clearances. More specifically, this paper addresses the following research questions:

Question 1: whether different conditions affect the nature (*omissions* or *mistakes*) of errors in pilots' transmissions to ATC;

Question 2: whether *words* and *numerals* give rise to different types of errors;

Question 3: whether different aviation communication items give rise to different types of errors.

2. Methodology

2.1. The flight simulator experiments

Farris, Trofimovich, Segalowitz, and Gatbonton (2008) investigated communication differences between NES and NNES under experimental conditions. Combining the fields of aviation human factors and linguistics, our research expands on this model with flight simulator experiments designed to assess pilots' communication accuracy by varying four factors expected to affect communication. Two of these factors directly affect radio communication and the ability of pilots to understand ATC transmissions: *increased speech rate* (MORROW; RODVOLD, 1998; BURKI-COHEN, 1995; TAYLOR; YESAVAGE; MORROW; DOLHERT; III; POON, 1994; SAID, 2011) and *greater information density* (CARDOSI, 1993; BARSHI; FARRIS, 2013). In their analogue-based studies, Barshi and Farris (2013) found that the ability of subjects (non-pilots) to understand a radio transmission was directly related to the number of items in the transmission: four or more items (instead of three or fewer items) double the number of communication errors. EUROCONTROL (2006b) found that rapid speech rate increases the likelihood of miscommunication in commercial aviation. Similarly for general aviation, we found that with increased ATC speech rate (from approximately 90 to 160 words per minute) pilots, particularly NNES pilots, produced more errors (MOLESWORTH; ESTIVAL, 2015). The other two factors are expected to affect communication less directly, by imposing additional cognitive demands rather than making communication more difficult: *increased pilot workload* and *increased radio congestion* (ORLADY; ORLADY, 1999; MORROW; LEE; RODVOLD, 1993; MORROW; RODVOLD, 1998). For each factor, pairs of flight scenarios with associated audio stimuli were created: one flight (A) served as the baseline and the other (B) provided the more challenging condition, as described in Table 1.

Table 1. Flight scenarios

<p>Scenario 1: Rate of speech of ATC</p> <ul style="list-style-type: none"> - Flight 1A – Baseline: pauses between items in ATC transmissions; average 89.5 words/minute - Flight 1B – Fast: no pauses between items in ATC transmissions; average 159.8 words/minute <p>Scenario 2: Information Density in ATC transmission</p> <ul style="list-style-type: none"> - Flight 2A – Baseline: fewer than 3 items per ATC transmission. - Flight 2B – High: 4 or more items per ATC transmission. <p>Scenario 3: Pilot workload</p> <ul style="list-style-type: none"> - Flight 3A – Baseline: normal navigation flight. - Flight 3B – High: navigation flight with in-flight fuel recalculation. <p>Scenario 4: Radio frequency congestion</p> <ul style="list-style-type: none"> - Flight 4A – Baseline: 1 other aircraft on the same frequency during approach and circuit to land at a training airport. - Flight 4B – Congested: more than 5 aircraft on the same frequency during approach and

circuit to land at a training airport.

The flight scenario details, including the flight plans and weather forecasts issued to the participants, were designed by the authors, both pilots familiar with the environment. All the flight scenarios would be familiar to the pilot subjects: scenarios 1, 2, and 3 consisted in standard navigation flights from a local aerodrome and scenario 4 involved an approach and landing at the local training airport. The authors also scripted the radio calls from ATC and other pilots that were played during the flights; these were recorded by aviation professionals. Each flight lasted about 8 minutes, with an opportunity for pilot transmission every 30 seconds, providing at least 15 pilot transmission opportunities per flight.

2.2. The participants

In contrast to Farris, Trofimovich, Segalowitz, and Gatbonton (2008), who employed analogue tasks and participants with no flying experience, the participants in the current study were all pilots recruited from local flight training schools, performing in a medium fidelity flight simulator. A total of 17 pilots (8 NES and 9 NNES) completed the set of 8 flights, in an average of 2 hours. Before the experiments, the pilots filled an informed consent form and a demographic questionnaire. Pilot experience ranged from 42 to 3,500 hours, with 7 pilots holding a Private Pilot Licence (PPL) or lower qualification, and 10 pilots a Commercial Pilot Licence (CPL) or higher (see Table 2).

Table 2. Distribution of pilots by language and qualification

Qualification \ Language	NES	NNES⁷
Commercial Pilot Licence (CPL) or higher	6	4
Private Pilot Licence (PPL) or lower	2	5
Total	8	9

These numbers are small, but the participants were all pilots and the spread of experience ensures we can make meaningful comparisons between NES and NNES pilots of high and low experience.⁸

2.3. Data Coding and Analysis

The experiments provided data for a total of 136 flights and 2,142 pilot transmission opportunities. The pilots' verbal responses were recorded with Audacity on one computer,⁹ and their

⁷ Regarding language background, the 9 NNES pilots gave their native language as: Cantonese (4), Chinese (1), Malayalam (1), Italian (1), Danish (1), and Russian (1); they reported to have been speaking English for between 2 and 35 years (average mean of 17.11, SD =11.96).

⁸ The sample size was sufficient to detect between-groups effects using Cohen's criteria (Cohen, 1988). With alpha at .05, and power at 0.95 (\bar{x} group 1 = 5 (SD=2), \bar{x} group 2 = 2 (SD=1)) the projected sample size needed (using G*Power 3.1) for this effect size is 14 (N).

⁹ Due to technical issues, the audio recordings for 6 flights were incomplete, resulting in 85 possible pilot transmissions (out of 2,142) missing from the data.

actions through the X-Plane flight simulator software on another computer. The X-Plane output files, giving altitude, heading and position at each point during the flights, will allow future analysis by comparing the pilots' actual behaviour with that expected from the scenarios and by comparing their verbal responses with their actions.

The audio data was coded by comparing the actual verbal responses with the transmissions expected from the scenarios and the regulations (as stated in the Australian Aeronautical Information Publication, AIP). Deviations from the expected transmission were categorised: a) as omissions or mistakes, b) as whether the error occurred with a lexical word or with a numeral, and c) for the locus of error (i.e. the type of item affected, e.g. Altitude or Heading). It is important to note that, even in non-routine situations where there may be no prescribed phraseology and pilots must resort to Plain English, it is possible to determine the number and types of items to be transmitted and thus to count and categorise errors. The flight simulator experiments did not include any non-routine situation requiring Plain English, but the pilots' transmissions contained many non-standard expressions. Deviations from the phraseology such as different ordering of the items or addition of extra words (as in example (4), also from the high workload Flight 3B) were not coded as errors.

- (4) Expected: (Track) 270. (Climb to) 5500. (Squawk) 5033. ((Contact) Wilton) 126.7. ABC.
Actual (P9): Climb to 5500, maintain track 180, contact Wilton 126.7. and...see ya.

For each flight, we counted both the number of incorrect *calls* (containing at least one error) and the number of incorrect *items*, because the call and item levels reveal different aspects of communication. The percentage of incorrect calls per flight provides a measure of the potential impact of miscommunication on radio congestion, when ATC might need to ask pilots to repeat their calls, but it is a coarse measure of communication accuracy. The percentage of incorrect items, which reflects a pilot's comprehension of ATC transmissions and that pilot's production, is a finer-grained measure of communication accuracy; as shown in Table 3, it is usually lower than the percentage of incorrect calls, because most calls contain several items.

Table 3: Percentage of incorrect calls and incorrect items

	Speech rate		Information density		Workload		Congestion		Average
	F1A	F1B	F2A	F2B	F3A	F3B	F4A	F4B	
% incorrect calls	44	41	41	59	40	50	31	31	42
% incorrect items	25	23	28	31	25	36	21	18	26

Earlier research hypothesized that NNES pilots would demonstrate a lower level of communication accuracy in the more challenging conditions (fast ATC speech rate, high information density, high pilot workload, and radio frequency congestion) than NES pilots would, and that pilots with lower qualifications would be less accurate than pilots with higher qualifications. Actual results

were more nuanced and revealed a complex relationship between the variables. Molesworth and Estival (2015) found an average overall communication accuracy of 57%, when measured in percentage of correct *calls*, with 37% incorrect transmissions for NES pilots and 49% for NNES pilots, i.e. average 43% error rate. Estival and Molesworth (2016) however, showed that with the finer-grained measure of *items* correctly transmitted, differences between flight conditions that were significant at the call level were no longer significant at the item level and, conversely, that differences between groups of pilots that were not significant at the call level were significant at the item level. At both call and item levels, NES pilots made fewer errors than NNES pilots. Significantly, however, NES and NNES pilots found it equally easy when the task was routine, such as arriving at the local aerodrome (Flight 4 A/B), or when communication was particularly difficult because of high information density (Flight 2B). Interestingly, high information density in ATC transmission (Flight 2B) had little effect (no statistical significance) on the percentage of items pilots transmitted correctly (28% for Flight 2A and 31% for Flight 2B), but showed a statistically significant difference at the call level (41% for Flight 2A and 59% for Flight 2B, shown in red in Table 3), which led to the conclusion that when required to transmit a higher amount of information, most pilots managed to transmit the same proportion of correct items.

The present study examines the nature and category of errors committed by pilots and the results reported here reveal further contrasts between pilots due not only to language background but to level of qualification.

3. Results

3.1. Omissions and mistakes (Research Question 1)

To analyse the nature of the errors, a series of dependent samples *t* tests were employed, and pilots' transmission accuracy in terms of omissions and mistakes was compared across flight conditions, i.e. between the A and B flights. The first dependent variable was the percentage of items correctly transmitted, relative to the total number of items to be transmitted. The second dependent variable was the nature of the error (*omission* or *mistake*). There were four main results.

Result 1. Overall, irrespective of flight pairing, there was a higher percentage of omissions (22.94%) than mistakes (2.72%) (shown in red in Table 4 below).

Result 2. The nature of the error only varied for *Information Density* (Result 3) and *Speech Rate* (Result 4), both involving difficulties with communication.¹⁰ Although, as was expected (see BARSHI; FARRIS, 2013), increased *Pilot workload* adversely affected the number of items

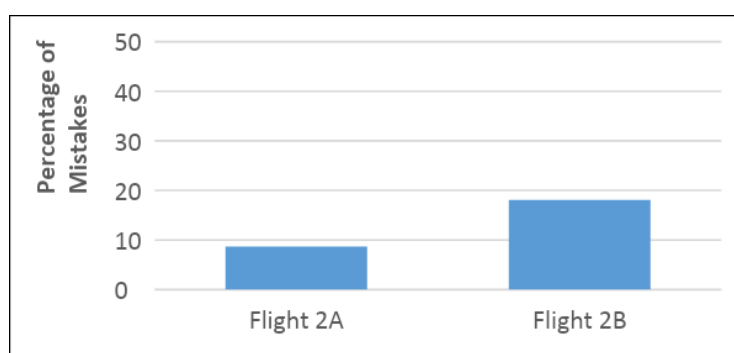
¹⁰ As expected, since no effect had been found at either the call or item level, increased radio congestion had no impact on the nature or category of errors.

correctly transmitted at a statistically significant level, there was neither a main effect nor any interaction between groups for the nature of errors. This means that, when pilot workload is greater but communication itself is not more difficult, the nature of communication error does not change, irrespective of pilot qualification or language background.

Result 3. For *Information Density*, there was both a main effect for flight condition (3.1) and pilot qualification (3.2).

3.1. Flight condition. As shown in Figure 1, the percentage of mistakes increased from 8.67% in Flight 2A (3 items or less) to 18.1% in Flight 2B (4 items or more).

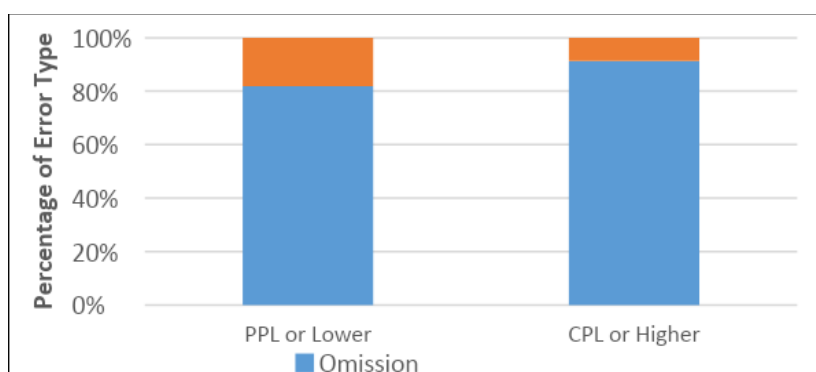
Figure 1. Percentage of mistakes (out of total errors) in the *Information Density* flights



Thus, when communication is more challenging, the nature of communication error changes and more mistakes are made.

3.2. Pilot qualification. As shown in Figure 2, CPL or higher licensed pilots made a smaller proportion of mistakes (8.65%) than PPL or lower licensed pilots (18.12%).

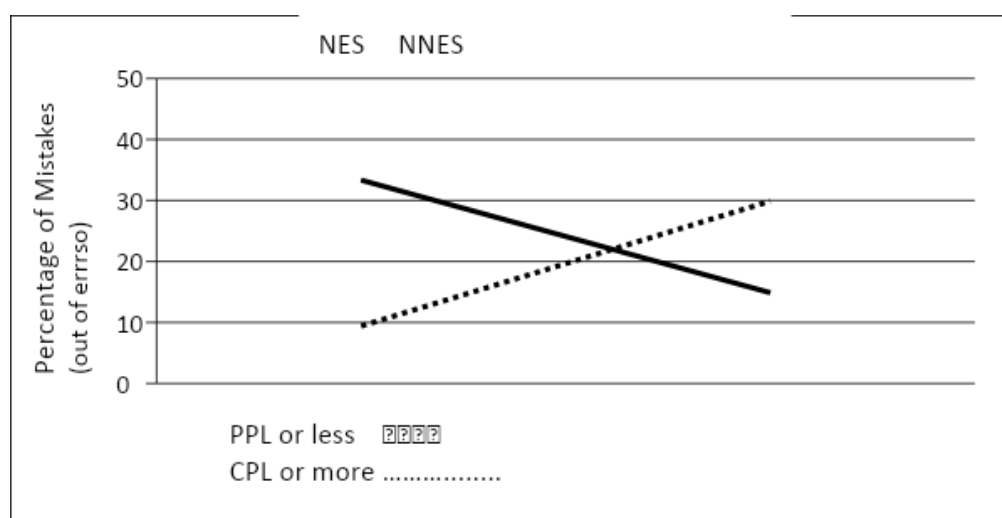
Figure 2. Omissions and mistakes for low and high pilot qualification, in the 'Information Density' flights



All pilots are less likely to make a mistake than an omission, but an error made by a CPL or higher licensed pilot is even less likely to be a mistake than an omission, showing that pilot training has a significant bearing on the nature of errors.

Result 4. For *Speech Rate*, the results are more complex. With increased ATC speech rate (Flight 1B), the proportion of mistakes and omissions varied according to both pilots' qualification and language background, both marginally significant. Although CPL or higher licensed pilots made fewer mistakes (19.65%) than PPL or lower licensed pilots (24.11%), when comparing the two license groups for NES against the two license groups for NNES, there was an interaction between language background and licence level for increased ATC speech rate as shown in Figure 3.

Figure 3. Mistakes in the fast *Speech Rate* flight, for pilot licence and language background



Low qualified NES pilots (left-hand side of the solid line) made a higher proportion of mistakes than high qualified NES pilots (left-hand side of dotted line), while low qualified NNES pilots (right-hand side of solid line) made a smaller proportion of mistakes than high qualified NNES pilots (right-hand-side of dotted line). One possible conclusion that can be drawn is that when communication itself is difficult, low qualified NNES pilots remain silent rather than risk making a mistake.

From this study of omissions and mistakes, we can already draw some practical conclusions. From Result 2, we would suggest that, when faced with omissions in a pilot's radio transmission, ATC should tailor their response in order to reduce that pilot's workload, e.g. by giving the pilot more time to respond. On the other hand, when ATC notice mistakes in a pilot's transmission (Result 3.2), they should aim to facilitate communication, e.g. by speaking more slowly or reducing the number of items in each call. From Result 4, apart from the recommendation that ATC slow their speech rate with NNES student pilots, we can make some suggestions for pilot training and for language teaching, to which we return in the final section. From Result 3.1, we can infer that the quality of pilot training, not flight hours nor language background, since these were not significant, is crucial for pilots' communication accuracy.

3.2 Words and Numerals (Research Question 2)

Many aviation communication items include numbers which are crucial for operations (such as altitude or heading) and while the choice of lexical items is severely constrained by the phraseology (ESTIVAL, 2016), the range of possible numerical values is very large. Radio communication errors often involve confusions between numbers (BURKI-COHEN, 1995) and errors in our data frequently occurred in items containing numbers. These errors were coded as occurring either on a *word* (or phrase) or on a *numeral* and analyses were performed to compare flight pairings with respect to the omissions and mistakes for words and numerals. The hypothesis was that there would be more mistakes for numerals than for lexical words, not only because of the greater choice of possible numerals than lexical words, but because numerals are less predictable from context. The numerical items in our data are: Altitude (e.g. *4500 feet*), Heading (e.g. *150 degrees*), Radio Frequency (e.g. *124.55*) and Transponder Code (e.g. *1200*). These items also contain lexical words, e.g. *climb* or *maintain* for Altitude and Heading, *contact* for Radio Frequency, or *squawk* for Transponder Code, i.e. verbs denoting the instructions to be performed by the pilot. Some items which pilots must read back or report only involve words, e.g. *Looking (for traffic)* or *Clear(ed) to land*. For analysis of the errors, a complicating factor is that words and numerals do not give rise to the same opportunities for errors in the readback of numerical items, because only the numerical value is obligatory (AIP, 2005). For example, as shown in (5), the instruction *Climb 4500* gives rise to 3 possible errors: omission of the numeral (5.b); mistake on the numeral (5.c), or mistake on the instruction (5.d). Both *Climb 4500* (5.e) and *4500* (5.f) are correct readbacks for (5.a), as omission of the instruction is not an error.

- (5) a. ATC: Climb 4500.
 b. Pilot: *Climb.
 c. Pilot: *Climb 3500.
 d. Pilot: *Maintain 4500.
 e. Pilot: Climb 4500.
 f. Pilot: 4500.

Nevertheless, some pilots read back the complete instruction, thereby increasing the opportunities for mistakes on words. Example (6) shows an expected readback during Flight 3A (normal workload) and the actual readback from Pilot 1, classified as incorrect (2 mistakes, 1 omission).

- (6) ATC: ABC. Wilton. Continue climb to 5500 and maintain 180. Traffic is a Cessna at your 11 o'clock at 5000. Report sighted. Contact Sydney Centre 124.55
 Expected: (Continue climb to) 5500. (Maintain) 180. Looking for traffic.
 (Contact) Sydney Centre 124.55. ABC.
 Actual (P1): Maintain 5500. Track 180, and contact Sydney Centre 124.55. ABC.

In (6) both mistakes concern the verbs for the instruction: *maintain* instead of *continue climb* for Altitude, and *track* instead of *maintain* for Heading, showing confusion between the instructions *continue climb* and *maintain*. The omission is that of the phrase acknowledging traffic information.

In (7) from Flight 1B (increased ATC speech rate), the pilot made two mistakes in one single item (Altitude): on the instruction (*descend* instead of *climb*) and on the numeral (*4000* instead of *4500*).

- (7) Expected: (Track) 150. (Climb to) 4500. (When clear of the zone,) (contact Sydney Centre), 124.55. ABC.
Actual (P7): Track 150 degree. Descend to 4000 feet. Contact Sydney Centre 124.55. ABC.

Items involving a radio frequency are more complex, because a radio station is often referred to by the name of the station (e.g. *Sydney Centre* or *Canberra Approach*) in addition to a digit sequence (e.g. *124.55*) as in (7). Thus, an instruction to switch to another frequency provides 3 possibilities for error:

- a) word error on the instruction (i.e. *contact*), which is very unlikely and does not occur in our data; (note that the instruction itself does not have to be read back);
- b) word error on the optional station designator (e.g. *Canberra* instead of *Camden*) as in example (8);
- c) numerical error on the obligatory digit sequence, as in example (9).

In (8) from Flight 3B (increased workload), the pilot confused *Canberra* with *Camden* and gave a radio frequency which is almost that for Camden (*120.1*) instead of Canberra (*124.5*).

- (8) ATC: ABC. Melbourne Centre. When ready, descend to 4500. Contact Canberra Approach 124.5.
Expected: (Descend to) 4500. (Contact) (Canberra Approach) 124.5. ABC.
Actual (P7): Contact Camden, 120.5, and descend to 4500 feet, ABC

In (9) from Flight 2B (increased information density), the numerical error in the pilot's readback (Frequency *110.1* instead of *126.7*) indicates a confusion with another numeral in the ATC instruction (Track *110*).

- (9) ATC: ABC. You are approaching Wilton, identified at 4500 in Danger area. Contact Wilton 126.7. Track 110 and climb to 5500.
Expected: (Contact Wilton) 126.7. (Track) 110. (Climb) 5500. ABC.
Actual (P2): Climb to 5500 and contact 110.1. ABC.

The results given in the previous section showed that there were many more omissions than mistakes in our data, irrespective of flight condition. In answer to Research Question 2, the analyses revealed four results, with the first one (Result 5) regarding the overall distribution of mistakes and omissions for words and numerals, and the other three concerning the impact of the four experimental conditions.

Result 5. As expected, mistakes rarely occur with the limited lexicon of Aviation English phraseology (1%) but are more frequent with the numbers to be transmitted (5%), while omissions are more likely to occur with a word (36%) than with a numeral (11%).

Table 4 gives the percentage of items in pilot transmissions in which the error affected a word or a numeral, relative to the total number of opportunities for errors (i.e. all the items to be transmitted by all participants) and whether the error was a mistake or an omission. Table 4 also gives the breakdown for each flight condition, with statistically significant result coloured red.

Table 4. Omissions and mistakes for word and numeral items, relative to opportunity

	Speech rate		Information density		Workload		Congestion		Total
	F1A	F1B	F2A	F2B	F3A	F3B	F4A	F4B	
% Word omissions (SD)	39.61 (16.30)	32.61 (15.56)	41.03 (20.14)	57.88 (22.72)	36.81 (11.48)	36.81 (11.49)	19.57 (11.11)	20.92 (10.07)	35.66 (14.86)
% Numeral omissions (SD)	9.14 (9.11)	8.56 (6.86)	17.13 (9.74)	16.90 (8.22)	10.37 (7.03)	17.04 (7.38)	3.70 (4.06)	1.85 (2.34)	10.59 (6.84)
% Total omissions (SD)	24.52 (10.18)	20.59 (10.05)	29.08 (13.72)	37.39 (14.47)	22.00 (8.22)	26.92 (7.77)	11.63 (7.11)	11.39 (5.30)	22.94 (9.60)
% Word mistakes (SD)	0.63 (1.20)	0.38 (0.81)	0.38 (1.09)	1.63 (2.22)	0.53 (1.19)	0.80 (1.26)	0.38 (0.81)	0.00 (0.00)	0.59 (1.07)
% Numeral mistakes (SD)	3.01 (2.43)	5.39 (3.93)	3.94 (4.96)	9.55 (6.25)	6.67 (6.89)	9.94 (8.84)	2.32 (0.93)	0.06 (0.25)	5.11 (4.31)
% Total mistakes (SD)	1.82 (1.25)	2.88 (2.12)	2.16 (2.48)	5.59 (3.60)	3.60 (3.81)	5.37 (4.78)	0.30 (0.57)	0.03 (0.13)	2.72 (2.34)

To assess the impact of the four experimental conditions on the category of error, a series of dependent *t* tests was employed to compare the A and B flights. The dependent variables were the percentage of omissions and mistakes from the total number of items to be transmitted by all the participants, i.e. relative to opportunity. The first dependent *t* test compared the percentage of total omissions, the second dependent *t* test compared the percentage of numeral omissions, while the third compared the percentage of word omissions. The same pattern of analysis was employed for mistakes: total mistakes, mistakes with numerals, and mistakes with words. For these analyses, alpha was set at .025 (.05/2 Bonferroni correction) to control for familywise error, as a result of the repeated use of the data. There were three main results.

Result 6. Only the *Information Density* condition affected the category of error: there were no statistically significant differences between the A and B flights for *Rate of Speech*, *Pilot Workload* or *Radio Congestion*.

Result 7. For *Information Density*, while there was no statistically significant difference for numeral omissions between Flight 2A (3 items or less) and Flight 2B (4 items or more), there was a

statistically significant difference for word omissions: 41% in Flight 2A and 58% in Flight 2B. This indicates that increasing the number of items in ATC transmissions results in pilots omitting words rather than omitting numbers. Indeed, we would expect that pilots strive to read back or report numerical values, as these are the most important pieces of information for a flight.

Result 8. Conversely, while there was no statistically significant difference between Flight 2A and Flight 2B for word mistakes, there was a statistically significant difference both for total mistakes and for numeral mistakes: 2% total mistakes and 4% numeral mistakes in Flight 2A against 6% total mistakes and 10% numeral mistakes in Flight 2B. This indicates that increasing the number of items in ATC transmissions adversely affects the accuracy with which pilots manage to transmit numbers.

In summary, the analysis of omissions and mistakes for words and numerals provides evidence of the different impact of the conditions under examination. High information density, which affects communication directly, not only increases the number of errors (Result 3), i.e. the magnitude of communication impairment, it affects the accuracy with which numbers (Result 8) – but not words – are transmitted. That is, this condition also affects the nature of impairment. On the other hand, while pilot workload increased the number of errors (MOLESWORTH; ESTIVAL, 2015), it did not affect the nature or category of error (Result 5). This contrast suggests that different cognitive processes are engaged for communication than for other tasks (in this case performing fuel calculations). It would be interesting to further study whether increased workload with other flying tasks, such as controlling the aircraft, monitoring the instruments, or navigating, have similar effects, i.e. would affect the magnitude but not the nature of communication impairment.

Future research should investigate in more detail the context for numerical errors, for instance possible confusions between Heading (given as a 3-digit group, e.g. *1 8 0* for *180°*) and Radio Frequency (also given as separate digits, e.g. *1 2 4 5 5* for *124.55*)¹¹, between the altitude to climb and the reporting level (given as spelled out numbers, e.g. *four thousand five hundred* for *4500 feet*) or between QNH, Radio Frequency and Transponder Code, as in (8) and (9). Another question is the effect of several numbers in one call, and the order in which they are given. Example (10) from Flight 1A shows different ordering of items in ATC and pilot's transmissions, and raises the question whether wrong ordering of items correlates with more errors.

- (10) ATC: ABC. Climb 3500. Track 160. Traffic is now a Cherokee at your 2 o'clock at 3500. What altitude do you intend to fly to Wollongong?

¹¹ The word *decimal* (e.g. *one two four decimal five five*), although recommended by the AIP, is usually omitted.

Expected: (Climb) 3500. (Track) 160. Looking for traffic. 4500 to Wollongong. ABC
 Actual (P11): Altitude intend to Wollongong is 4500 and your altitude climb to was 3500 and tracking 160. ABC.

Relying on memory, especially the aural component of working memory, can prove fraught and a cause of errors. A practical suggestion for student pilots to guard against errors when several numbers are involved is to follow the professional pilots' practice of writing down numbers, preferably in a template with slots for the different types of information. Since the condition with the greatest negative impact on accuracy (especially for numbers) was increased information density in ATC transmissions, student pilots should be encouraged to ask ATC to repeat long instructions; better still, ATC should refrain from providing long instructions in the first place.

3.3. Locus of errors (Research Question 3)

To some extent, the more interesting question from a linguistic point of view is which item types are more likely to give rise to errors. This is also a crucial question from the point of view of aviation safety, as different item types carry different potential threats if read back incorrectly: a mistake on assigned altitude or heading may result in traffic conflict and therefore a potential mid-air collision; an incorrect radio frequency may result in an aircraft being out of contact and unable to hear further calls from ATC or other aircraft and therefore in potential traffic conflicts; a seemingly innocuous error, such as call-sign omission, may increase ATC workload and radio congestion. Table 6 lists the types of information items in the data, ranked by percentage of errors, with illustrative examples of correct calls and possible errors.

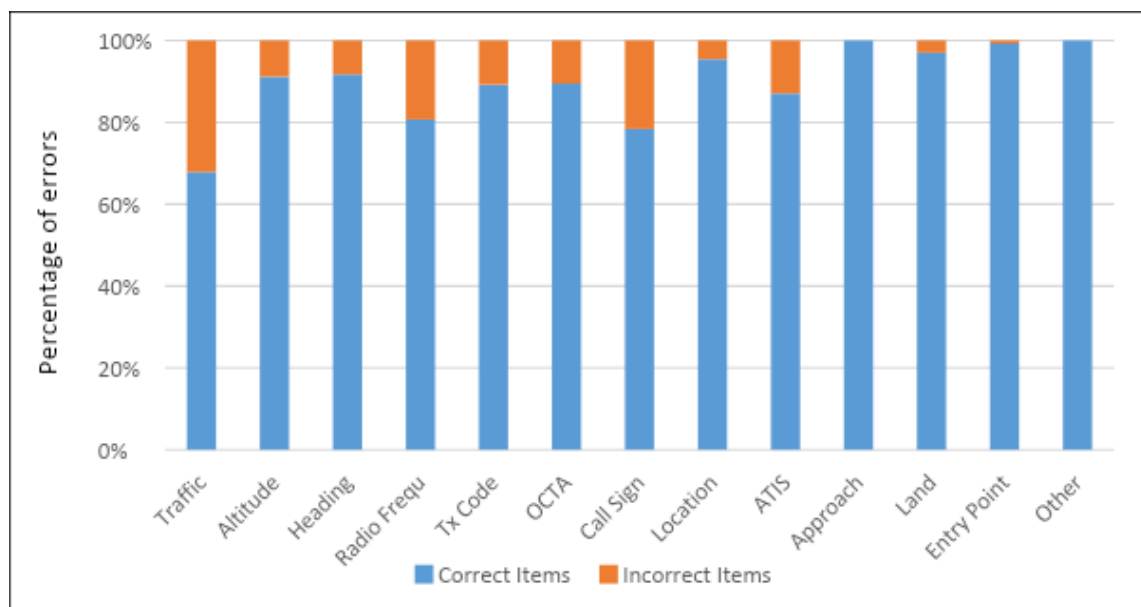
Table 6. Aviation communication items with examples,

Item Type	Correct calls	Incorrect responses	% errors
Traffic	<i>Traffic sighted</i> <i>Looking for traffic</i>	Failure to respond to traffic information	32.18
Call-Sign	<i>ABC (Alpha Bravo Charlie)</i> <i>YWD (Yankee Whiskey Delta)</i>	Omitting call-sign in readback or report.	21.59
Radio Freq	<i>124.55</i> <i>Canberra Approach</i>	Not saying the frequency; giving wrong frequency.	19.30
ATIS	<i>Information Hotel</i>	Failure to state received information; giving wrong information code.	12.91
Tx Code	<i>3000</i> <i>1200</i> <i>5033</i>	Not reading back the transponder code; giving wrong code.	10.79
OCTA	<i>Remain outside controlled</i> <i>airspace</i>	Failure to read back instruction to remain outside controlled airspace	10.52
Heading	<i>180 (one eight zero)</i> <i>270 (two seven zero)</i>	Not reading back the heading; giving the wrong heading.	9.20
Altitude	<i>4500 (four thousand five hundred)</i> <i>3000 (three thousand)</i>	Not reading back the altitude; giving the wrong altitude; failure to report passing altitude.	8.84
Location	<i>2RN (two R N)</i> <i>Crosswind</i>	Not giving location; giving wrong location;	6.53

	<i>Wilton</i>	failure to report at location.	
Land	<i>Cleared to land</i>	Failure to read back ' <i>Cleared to land</i> '	2.99
Entry Point	<i>Join downwind 29 Right</i>	Failure to read back the entry point	1.52
Approach	<i>Cleared visual approach</i>	Failure to read back ' <i>Cleared visual approach</i> '	0.00
Other	<i>Confirm inbound</i>	Not confirming inbound. Other errors.	0.00

There was no expectation that experimental conditions would affect the locus of error and indeed the data showed no statistically significant difference for item type, neither within a flight pair (A and B flights) nor across the set of 8 flights. Of interest, however, was the type of items which give rise to errors and the relative frequency of these errors, relative to opportunity, as shown in Figure 4. Two results for the locus of error are worth mentioning: Result 9 (see Figure 4) and Result 10 (see Figure 5).

Figure 4. Errors per information item type



Result 9. By far the most common error in the data was the failure to respond to traffic information.

While other items correspond, even in a flight simulator, to actions required to fly the aircraft (Altitude, Heading) or settings to be changed (Radio Frequency, Transponder Code), traffic information is less realistic so the prevalence of this error may be an artefact of the experimental setup, as the pilots would be less concerned about the consequences of a mid-air collision than in a real flight. Nevertheless, pilots are trained to respond *Looking for Traffic* when they are alerted by ATC to traffic in their vicinity and have not seen the other aircraft, so we would expect this reply. Even more of a concern was the high proportion of pilots answering *Traffic sighted*. Given there never were any other aircraft in the

display, if this is an automatic answer given by pilots in real flight situations when they have not seen another aircraft, this would have serious implications for air safety and should be investigated further.

Result 10. The second common error was not including the aircraft call-sign as required.

In fact, 42% of word omissions were the aircraft call-sign, confirming earlier results by Cardosi (1993) and Morrow and Rodvold (1998). Call-sign mistakes can and do occur in real flights (the pilots of Malaysian Airlines flight number MH370, which went missing in 2014, stated their call-sign as MAS377 instead of MAS370 when requesting clearance to taxi)¹², but none occurred in our data. Pilots are used to changing call-signs when they fly different aircraft and our participants made no mistake when giving their realistic (for Australian aircraft) call-sign *ABC*, but they often omitted it. Interestingly, it was sometimes the only item pilots managed to transmit correctly. This would suggest that the call-sign might be the easiest information item for a pilot to transmit and that it can serve as fall-back when no other information is readily available.

The next most frequent locus errors concerned the four numerical items (Altitude, Heading, Radio Frequency, Transponder Code) and the word item for the ATIS (automated information for an aerodrome). Numerical items are arguably the most critical pieces of information and they must always be read back. Figure 5 shows the relative prevalence of omissions and mistakes for words and numerals in the four numerical items, relative to opportunity.

Figure 5. Omissions and mistakes for information items containing both words and numerals

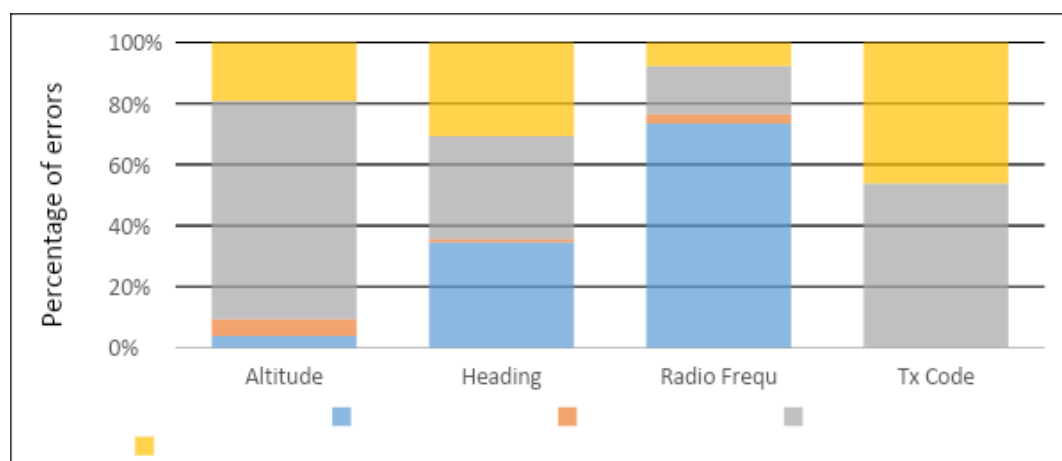


Figure 5 illustrates both Result 1, that there were more omissions than mistakes overall, and Result 5, that there were more mistakes with numerals than with words. This holds true for both Altitude and Heading; for Transponder Code, errors only concerned the numerals (i.e. there were no word errors) - with slightly more omissions than mistakes. In the case of Radio Frequency, which can be referred to by a phrase (e.g. *Sydney Centre*) in addition to a digit string (e.g. *124.55*), there were more overall errors

¹² See Estival (2016) for other real examples.

(mistakes and omissions) on words than on numerals but again more mistakes with numbers than with words (such a rare word mistake could be *Camden* instead of *Canberra* as in example (8)), that is there were more omissions with the name of the radio station than with the digit string (which is obligatory).

Figure 6. Word omissions and mistakes for information items containing only words, relative to opportunity

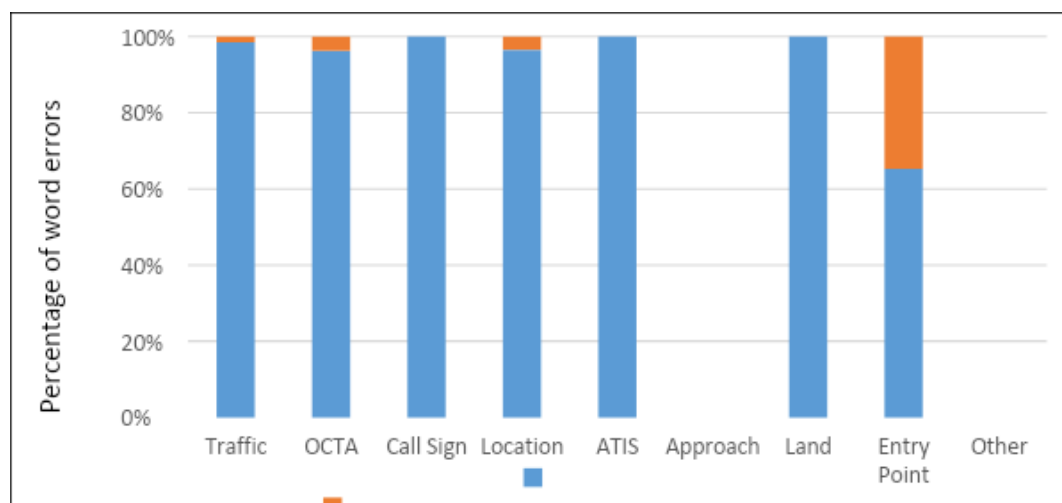


Figure 6 shows the relative prevalence of omissions and mistakes in the non-numerical items (i.e. containing only words), relative to opportunity. With a vocabulary as limited as that of Aviation English, words are certainly more memorable than the almost infinite range of numbers required for numerical items. Even where the numerical range for an item (e.g. a maximum of 360 degrees for heading) is limited, the number of possible and plausible numbers is much larger than the choice of words (e.g. ‘Turn’, or ‘maintain’). The required pronunciation for numbers (e.g. ‘one two four five five’ for 124.55) helps with number recognition and avoids confusion but does not help with memory. The conclusion however cannot be to replace all numerals with lexical items. In a technical environment such as aviation, numerical values are necessary and more precise. Although it might be tempting to conclude from Figure 5 that radio station names should be used rather than digit strings, the numerical value is still required to dial the correct station and correct repetition of the name is no guarantee that the correct number would be dialled. Of course, correctly reading back an assigned altitude or heading is no guarantee the correct action will follow either, but correct readback at least ensures that the correct information has been shared.

3.4. Multiple errors

Multiple errors in a call indicate some communication impairment. Table 5 gives the number of calls with two or more errors and the mean number of errors per call for each flight (all pilots combined). While there was no significant difference between the A and B flights for *Speech rate* and

Congestion, a paired samples *t*-test attested to the significant difference for *Information Density* (average mean error of 4.69 (SD = 3.03) for Flight 2A and 7.24 (SD = 2.77) for Flight 2B; ($t(17) = 2.660, p = .018$)) and for *Workload* (average mean error of 3.87 (SD = 1.88) for Flight 3A and 5.82 (SD = 2.32) for Flight 3B; ($t(17) = 3.084, p = .008$)). This confirms results at call and item levels for the negative impact of *Information Density* and *Workload* on pilots' communication ability (MOLESWORTH; ESTIVAL, 2015; ESTIVAL; MOLESWORTH, 2016).

Table 5. Radio calls with more than 2 errors

	Speech rate		Information density		Workload		Congestion	
	F1A	F1B	F2A	F2B	F3A	F3B	F4A	F4B
Number of calls with \geq 2 errors	64	56	75	123	58	99	32	29
Mean errors per call	3.76	3.50	4.69	7.24	3.87	5.82	2.00	1.71

4. Discussion

The flight simulator experiments initially aimed to assess the impact of four conditions (speech rate, information density, workload and radio congestion) on the communication accuracy of pilots with different language backgrounds (NES/NNES) and varying flight training levels. The new results reported here, however, confirm complex interactions that go beyond language background, between factors such as cognitive load, speech rate, and information density and the nature (omissions or mistakes) and category (words or numerals) of errors. As expected, mistakes with numerals were more common than with lexical words (Result 5). The most important finding was that increased information density in ATC transmissions led to an increase in mistakes with numerals in pilots' readbacks (Result 8), whereas increased pilot workload led to an overall increase in errors but had no effect on the nature (Result 2) or category (Result 6) of error. As expected, mistakes with numbers increased when several numbers were involved in one dialogue turn and when these numbers could be confused, for instance a Heading of *110* (degrees) and a Radio Frequency starting with *11*, such as *118.7*.

The analysis of the locus of errors revealed a high number of call-sign omissions (Result 10). This may not seem an important error, but call-signs allow ATC to distinguish between aircraft. Additional ATC requests for identification increase radio congestion, potentially preventing other aircraft from making urgent or even emergency calls, and causing confusion or distraction for other pilots. Call-sign mistakes did not occur in our data, but confusion of call-signs is recognised as a major cause of errors in spite of the phonetic alphabet and specified pronunciation for letters and numbers (ATSB, 2009; EUROCONTROL, 2006b; CARDOSI, 1993; MORROW; RODVOLD, 1998).

Language background, flight hours and pilot qualifications were independent variables for this study. An important result is that the pilots' language background was not correlated with either the category of error (words or numerals) or the locus of error (item type); the only differences between NES and NNES pilots (Result 4) concerned the nature of error (omissions or mistakes) in one condition (increased ATC speech rate). Another important result is that the number of flight hours was not significant, but the level of qualification was significant (Result 3.2). Together, these findings present important implications from a cognitive perspective and for aviation training and language teaching.

4.1. Cognitive implications

The finding that the nature of error only varied with increased information density (Result 2) suggests that different cognitive processes are engaged for communication and for other flying tasks. Pilots obviously can and do perform several tasks simultaneously; however, when the cognitive resources required for these tasks exceed the resources available, performance degrades. In Baddeley's Multicomponent Model of Working Memory (BADDELEY, 2000) tasks such as fuel calculation involve the executive function of working memory while communication tasks involve the phonological loop component of working memory. Such a model would explain why the increased workload of performing fuel calculations while continuing to fly the aircraft and communicate with ATC affected the frequency of communication errors but not the nature of those errors in Flight 3B.

4.2. Implications for aviation language training

This study presents implications for how communication is taught in flight training. The overall poor communication for GA (average accuracy of 58% in terms of correct calls, as shown in Table 3) seems to indicate that aviation radio communication is very difficult or that GA pilots are not adequately trained, or both. Indeed, communicating over the radio is challenging and learning a restricted code such as Aviation English requires paying attention to turn-taking protocol, unusual phraseology and pronunciation. For NNES pilots, it is more challenging than learning English for everyday conversation and more difficult than speaking over the telephone. All student pilots rehearse their calls on the ground and, when flying or navigating becomes difficult, they often ask their instructor to make the calls for them. The significant impact of qualification level, rather than flight hours, on communication accuracy would indicate that in order to reduce the number of communication errors, the quality of training needs to be improved. The finding that all pilots are less likely to make a mistake than an omission, but that higher qualified pilots (CPL or higher licence) are even less likely to make mistakes (Result 3.2) also suggests that quality of training is important for quality of communication. Future research should examine which parts of the CPL curriculum could account for this difference and might be applied at the

lower level. The insistence on professionalism early in the training could be a factor that might translate into language teaching contexts.

The finding that low qualified NES pilots made more mistakes than high qualified NES pilots, whereas low qualified NNES pilots tended to remain silent (Result 4) requires further interpretation. A plausible explanation could be that early NNES pilots have a greater awareness of the limitations of their communication skills, resulting in a reluctance to speak when not certain how to respond. This difference between NNES and NES pilots suggests different approaches to their training: NNES need to be encouraged to use the radio rather than chastised for their mistakes, but NES need to be taught not to press the push-to-talk button too quickly and start to speak too early. NES need to be reminded to think before they speak until they have mastered both flying and aviation communication skills; indeed, pilots are taught to prioritize tasks and, as summarised in the often-used adage ‘Aviate, Navigate, Communicate’, communication must come last. When NNES have acquired greater aviation skills, however, they need to be reminded to be cautious and not to feel overconfident about their radiotelephony skills

An important question for Aviation English teachers, flight and ATC instructors and examiners, is what level of English proficiency is required and whether Aviation English with its specific phraseology for radiotelephony is sufficient. Mell (2006) showed that spontaneous unscripted native English is not appropriate in an NNES aviation environment while Drury and Ma (2002) found that ‘native English speakers should be taught how to communicate simply, slowly and precisely with their non-native English speaking colleagues’. The logical consequence that Aviation English training should be explicitly provided to NES as well as to NNES pilots and that testing of NES pilots should include the ability to communicate with NNES pilots and ATC has become an important topic for researchers (ESTIVAL; FARRIS; MOLESWORTH, 2016; KIM, 2012). Although advocated by the regulator (ICAO, 2010) it is not yet general practice in the industry (ESTIVAL, 2018, 2019). Our study shows that NES and NNES pilots make similar errors and thus supports the key argument in recent ELF research (SEIDLHOFER, 2011) that the onus of adjusting to ensure successful communication has to be shared by NES and NNES, and the suggestion by Clark (2017) that viewing Aviation English as a lingua franca would “help to remove the ‘ownership’ that native English speakers feel over English”.¹³

Finally, we must take into account the effect of the noisy cockpit environment (80db for Commercial Aviation and up to 95db for General Aviation). Noise makes it more difficult for any listener to understand speech but is especially detrimental for NNES listeners (SHIMIZU;

¹³ As pointed out by one anonymous reviewer, though, one aspect of ELF that cannot apply in a high-stakes context is the suggestion that if you do not understand something, you should ‘let it pass’ until you come across information you do understand (FIRTH, 1996, p.243)

MAKISHIMA; YOSHIDA; YAMAGISHI, 2002; GAT; KEITH, 1978). Radio transmission, resulting in a degraded speech signal even with high-quality noise-cancelling headsets (MOLESWORTH; BURGESS; GUNNELL; LOFFLER; VENJAKOB, 2014; JANG; MOLESWORTH; BURGESS; ESTIVAL, 2014) further compounds aviation communication problems and may increase the confusability of sounds in ATC instructions, as between ‘maintain’ and ‘vacate’ or, even with the required pronunciation, between ‘NI-ner TOU-sand’ (9000) and ‘FI-fe TOU-sand’ (5000). This points to the importance of context and physical environment in language training for specific purposes.

4.3. Implications for language teaching

Training organisations often assume that communication skills develop as a result of exposure, but we have shown that the number of flight hours was not as important as pilot qualification (indeed not a significant factor). We propose that this is due to the more rigorous training and the amount of studying required for higher pilot qualifications. For language teaching, this would suggest that the sheer number of hours of practice in a language lab or a classroom, or even of immersion in a language environment, is not sufficient to guarantee proficiency. Two other factors, quality of teaching and the student’s personal efforts, may have a greater effect on a student’s achievements. The better communication of pilots with professional qualifications also suggests that students’ motivation and personal goals are important factors, and we would expect that the motivations for studying a language have a similar effect on learning outcomes (see DÖRNYEI; USHIODA, 2009). The finding that low qualified NNES pilots tended to remain silent (Result 4) also has implications for language training and the assumptions instructors may make about quiet students’ abilities, as silence could indicate an avoidance strategy (KLEINMANN, 1977) rather than ignorance.

Conclusions and Future Research

This paper builds on previous results concerning the frequency of errors in pilots’ transmissions during communication with ATC and presents new results on the nature (omissions or mistakes), category (words or numerals) and locus (item types) of those errors. Besides the importance of these findings for air safety, there are wider implications for language production and language processing and for cognition under challenging conditions, such as increased workload and difficult communication. From a linguistic point of view, words and numerals offer different opportunities for errors, and give rise to different types of errors. From a cognitive point of view, omissions and mistakes seem to reflect different processes, with different implications for training and for managing workload during operations. Finally, from an operational perspective, a different locus of error (e.g. call-sign or

altitude) has different implications for aviation safety (frequency congestion or potential mid-air collision).

The findings presented here allow us to make some practical suggestions: for pilots to guard against errors by writing down information and requesting repetition of long instructions; and for ATC who notice omissions or mistakes in a pilot's transmissions to give that pilot more time to respond, or slow down and reduce the amount of information. We also found that although, as expected, NES pilots usually made fewer errors than NNES pilots, the NNES/NES distinction was confounded by other factors: the challenges posed by higher information density and greater pilot workload offset the native language advantage, while ATC speech rate had a significant impact only on low qualified NNES pilots.¹⁴ This specific finding strongly supports existing practices, and the ICAO recommendation that the speech rate of radiotelephone operators not exceed 100 words per minute (ICAO, 2001). Our findings provide further support for the recommendations of Clark (2017), based on an analysis of incident reports, regarding complexity and length of ATC messages and ATC speech rate, and especially the suggestion to “pursue research into an ‘ideal’ speech rate for [UK] controllers speaking with pilots with ICAO Level 4 proficiency”.

Following our analysis of the effect of four conditions on pilots' communication accuracy, we still need to investigate how these variables interact, e.g. high pilot workload and fast ATC speech rate or high pilot workload and high information density. Another question is how the accent of pilots and ATC might affect communication and the possible interaction between accent and workload. Indeed, Farris, Trofimovich, Segalowitz, and Gatbonton (2008) found that low proficiency NNES pilots repeated messages less accurately and that their speech was perceived as more accented in a high workload condition, while Estival and Molesworth (2012) reported that pilots' perception of communication difficulties changed with the arrival of an ATC speaking an outer circle English variety.

While a possible limitation of the experimental setup is the well-known observer paradox (LABOV, 1972), it is less of a factor for flight simulator experiments, as pilots always know they are under observation, both by ATC and by other pilots on the same frequency. Another limitation might be that the flight simulator only approximates real flying conditions, with little risk of adverse consequences for the pilots and no possibility of repetition by ATC. The lack of physical risk may indeed explain some of the deviations found in the responses to traffic information. As pointed out by Cardosi (1993) however, an analysis of real live communications “cannot provide a suitable data base for extensive error analysis, since the frequency of errors is small relative to the total number of

¹⁴ This does not mean that ATC speech rate never has any effect on highly qualified NNES commercial pilots, only that in our data it was only significant for low qualified NNES pilots.

transmissions.” Indeed the 43% error rate reported for GA was at the call level for all types of errors, including single call-sign omissions (MOLESWORTH; ESTIVAL, 2015). Controlled experiments in a flight simulator provide useful data for the analysis of errors under different conditions and reveal errors which would not be so apparent in the wild (HUTCHINS, 1995), i.e. the operational environment. This study is an example of a carefully designed experiment to identify what goes wrong and why in a very specific domain of language use. Such experiments are not commonly undertaken but would provide a strong evidence-base for ESP teaching. The results of this particular study could also inform the kinds of tasks and task conditions which could be included in the assessment of language proficiency for radiotelephony communication.

Acknowledgements

The authors acknowledge the support of the MARCS Institute, whose grant to Estival made this research possible. We would like to thank Mr David Maddock for the time spent recording the ATC messages, Mr Raymond Jang for assisting with the collection of data, and Ms Annie Zhou and Mr Dimuth Seneviratne for their help transcribing the communication recordings. We would also like to thank those pilots who volunteered their time to complete the research. Finally, we thank A/Prof Catherine Elder, Dr Hyejeong Kim, Prof Denis Burnham, and Dr Julia Trippe for insightful comments on earlier versions of this paper, and two anonymous reviewers.

References

- AGENZIA NAZIONALE PER LA SICUREZZA DEL VOLO. 2004. Accident involved aircraft Boeing MD-87, registration SE-DMA and Cessna 525-A, registration D-IEVX Milano Linate airport October 8, 2001. Rome, Italy.
- AIP. 2005. Aeronautical Information Publication. edited by Airservices Australia. Canberra, Australia: Airservices Australia.
- ATSB. 2009. Perceived threats, errors and safety in aerial work and low capacity air transport operations. Canberra, Australia: Australian Transport Safety Bureau.
- BADDELEY, A.D. 2000. The episodic buffer: a new component of working memory? *Trends in Cognitive Science* 4.11: 417-423.
- BARSHI, I.; C. FARRIS. 2013. *Misunderstandings in ATC Communication*. Ashgate Studies in Human Factors for Flight Operations. Farnham, England: Ashgate.
- BIESWANGER, M. 2013. *Applied Linguistics and Air Traffic Control: Focus on Language Awareness and Intercultural Communication*. In HANSEN-SCHIRRA, S.; K. MAKSYMYSKI (Eds.). *Aviation Communication: Between Theory and Practice*. Frankfurt/Main, Germany: Peter Lang Verlag. p. 15-32.
- , 2016. *Aviation English: Two distinct specialised registers?* In SCHUBERT, C.; C. SANCHEZ-STOCKHAMMER (Ed.). *Variational text linguistics: Revisiting register in English*. Berlin: DeGruyter, p. 67-85.
- BOROWSKA, A. 2017. *Avialinguistics: The study of language for aviation purposes*. Frankfurt am Main: Peter Lang.

- BURKI-COHEN, J. 1995. Analysis of tower ground controller-pilot voice communication. Washington DC, USA: Federal Aviation Administration.
- CARDOSI, K.M. 1993. An analysis of en route controller-pilot voice communications. Cambridge, MA, USA: Federal Aviation Administration.
- CLARK, B. 2017. Aviation English Research Project: Data analysis findings and best practice recommendations. Civil Aviation Authority (Civil Aviation Authority). Available at: <<https://publicapps.caa.co.uk/modalapplication.aspx?catid=1&pagetype=65&appid=11&mode=detail&id=7802>>.
- COHEN, J. 1988. Statistical power analysis for behavioral sciences. New York, USA: Lawrence Erlbaum.
- CUSHING, S. 1994. Fatal words: Communication Clashes and Aircraft Crashes. Chicago, IL, USA: University of Chicago Press.
- CUTLER, A. (Ed.) 1982. Slips of the tongue and language production. Vol. 19-7/8 [Special issue of *Linguistics*]: Mouton.
- DÖRNYEI, Z.; E. USHIODA (Eds.) 2009. Motivation, language identity and the L2 self. Bristol, UK.: Multilingual Matters.
- DRURY, C.G.; J. MA. 2002. Language Error Analysis - Report on Literature of Aviation Language Errors and Analysis of Error Databases. Report prepared for the FAA. University of Buffalo, State University of New York, USA.
- ESTIVAL, D. 2016. Aviation English: A linguistic description. In: ESTIVAL, D.; C. FARRIS; B. MOLESWORTH (Ed). Aviation English: A lingua franca for pilots and air traffic controllers. New York: Routledge, p. 22-53.
- _____. 2018. What Should We Teach Native English Speakers? International Civil Aviation English Association (2018) Conference, Daytona Beach, Florida, USA. Available at: <<https://commons.erau.edu/icaea-workshop/2018/proceedings/>>.
- _____. 2019. Aviation English Training for Native English Speakers: Challenges and Suggestions. International Civil Aviation English Association (2019), Tokyo, Japan. Available at: <<https://commons.erau.edu/icaea-workshop/2019/proceedings/1/>>.
- ESTIVAL, D.; C. FARRIS; B.R.C. MOLESWORTH. 2016. Aviation English: A lingua franca for pilots and air traffic controllers. In: PALTRIDGE, B.; S. STARFIELD. (Ed.). Routledge Research in English for Specific Purposes. London, UK: Routledge.
- ESTIVAL, D.; B.R.C. MOLESWORTH. 2009. A Study of EL2 Pilots Radio Communication in the General Aviation Environment. Australian Review of Applied Linguistics. 32.3: 24.1-24.16.
- ESTIVAL, D.; B.R.C. MOLESWORTH. 2012. Radio Miscommunication: EL2 Pilots in the Australian General Aviation environment. Linguistics and the Human Sciences. 5.3: 351-78.
- ESTIVAL, D.; B.R.C. MOLESWORTH. 2016. Native English speakers and EL2 pilots : an experimental study. In Aviation English: A Lingua Franca for Pilots and Air Traffic Controllers, In Routledge Research in English for Specific Purposes. London, UK: Routledge, p. 140-181.
- EUROCONTROL. 2006a. European Action Plan for Air Ground Communications Safety, Edition 1.0, May 2006. Brussels, Belgium: EUROCONTROL.
- _____. 2006b. Air-ground Communication Safety Study: Causes and Recommendations (Report # DAP/SAF 2006-09). Brussels, Belgium: European Organisation for the Safety of Air Navigation.
- FARRIS, C.; P. TROFIMOVICH; N. SEGALOWITZ; E. GATBONTON. 2008. Air Traffic Communication in a Second Language: Implications of Cognitive Factors for Training and Assessment. TESOL Quarterly. 42.3: 397-410.
- FIRTH, A. 1996. The discursive accomplishment of normality: on 'lingua franca' English and conversation analysis. Journal of Pragmatics. 26.

- GAT, I.B.; R.W. KEITH. 1978. An effect of linguistic experience. Auditory word discrimination by native and non-native speakers of English. *Audiology and Neurotology*. 14.4: 339-345.
- GOGUEN, J.; C. LINDE. 1983. *Linguistic Methodology for the Analysis of Aviation Accidents*. Ames Research Center, Moffett Field, California, USA: NASA Contract Report 3741.
- HUTCHINS, E. 1995. *Cognition in the Wild*. Cambridge, MA, USA: MIT Press.
- ICAO. 2001. *Aeronautical Communications. Annex 10 to the Convention on International Civil Aviation*. Montreal, Canada.
- _____. 2010. *ICAO Doc 9835. Manual on the Implementation of ICAO Language Proficiency Requirements, 2nd Edition*. Chicago, USA: International Civil Aviation Organization.
- _____. 2016. *Annex 10 to the Convention on International Civil Aviation: International Standards and Recommended Practices and Procedures for Air Navigation Services, 7th edition*. Montreal, Canada: International Civil Aviation Organization.
- JANG, R.; B.R.C. MOLESWORTH; M. BURGESS; D. ESTIVAL. 2014. Improving Communication in General Aviation through the use of Noise Cancelling Headphones. *Safety Science*. 62: 499-504.
- KIM, H. 2012. Exploring the construct of aviation communication: A critique of the ICAO language proficiency policy. Ph.D., Linguistics and Applied Linguistics, The University of Melbourne.
- KIM, H.; R. BILLINGTON. 2016. Pronunciation And Comprehension In English As A Lingua Franca Communication: Effect Of L1 Influence In International Aviation Communication. *Applied Linguistics*: 1-25. Available at: <<https://doi.org/10.1093/applin/amv075>>.
- KIM, H.; C. ELDER. 2009. Understanding aviation English as a lingua franca: Perceptions of Korean aviation personnel. *Australian Review of Applied Linguistics*. 32: 23.1-23.17.
- KLEINMANN, H.H. 1977. Avoidance behaviour in adult second language acquisition. *Language Learning*. 27: 93-107.
- LABOV, W. 1972. *Sociolinguistic Patterns*. Philadelphia, PA, USA: University of Pennsylvania Press.
- MELL, J. 2006. *Emergency Calls - Messages out of the blue*.
- MODER, C.L. 2013. Aviation English. In Paltridge, B.; S. Starfield (Ed.). *The Handbook of English for Specific Purposes*, Malden MA, USA: Wiley Blackwell, p. 227-242.
- MOLESWORTH, B.R.C.; M. BURGESS. 2013. Improving intelligibility at a safety critical point: In flight cabin safety. *Safety Science*. 51: 11-16.
- MOLESWORTH, B.R.C.; M. BURGESS; B. GUNNELL; D. LOFFLER; A. VENJAKOB. 2014. The effect on recognition memory of noise cancelling headphones in a noisy environment with native and non-native speakers. *Noise and Health*. 16.17: 240-247.
- MOLESWORTH, B.R.C.; D. ESTIVAL. 2015. Miscommunication in general aviation: The influence of external factors on communication errors. *Safety Science*. 73: 73-79.
- MORROW, D.; A. LEE; M. RODVOLD. 1993. Analysis of problems in routine controller-pilot communication. *International Journal of Aviation Psychology*. 3.4: 285-302.
- MORROW, D.; M. RODVOLD. 1998. Communication Issues in Air Traffic Control. In SMOLENSKY, M.W.; E.S. STEIN. (Ed.). *Human Factors in Air Traffic Control*. San Diego, USA: Academic Press, p. 421-456.
- ORLADY, H.W.; L.M. ORLADY. 1999. *Human factors in multi-crew flight operations*. Aldershot, USA: Ashgate.
- PALTRIDGE, B.; S. STARFIELD (Eds.) 2013. *Handbook of English for Specific Purposes*. Malden MA, USA: Wiley-Blackwell.
- PRINZO, O.V. 2008. The computation and effects of air traffic control message complexity and message length on pilot readback performance. *Measuring Behavior 2008*, Maastricht, The Netherlands, August 26-29, 2008.
- SAID, H. 2011. *Pilots/Air Traffic Controllers Phraseology Study*. Montréal, Québec: IATA.
- SEIDLHOFER, B. 2011. *Understanding English as a Lingua Franca*. Oxford, UK: Oxford University Press.

- SHIMIZU, T.; K. MAKISHIMA; M. YOSHIDA; H. YAMAGISHI. 2002. Effect of background noise on perception of English speech for Japanese listeners. *Auris Nasus Larynx*. 29: 121-125.
- TAYLOR, J.L.; J.A. YESAVAGE; D.G. MORROW; N. DOLHERT; J.O.B. III; L.W. POON. 1994. Effects of information load and speech rate on young and older aircraft pilots' ability to read back and execute simulated ATC instructions. *Journal of Gerontology: Psychological Sciences*. 49: 191-200.
- TIEWTRAKUL, T.; S.R. FLETCHER. 2010. The challenge of regional accents for aviation English language proficiency standards: A study of difficulties in understanding in air traffic control-pilot communications. *Ergonomics*. 53.2: 229-239.
- WU, Q.; B.R.C. MOLESWORTH; D. ESTIVAL. 2019. An investigation into the factors that affect miscommunication between pilots and air traffic controllers in commercial aviation. *The International Journal of Aerospace Psychology*: 53-63. Available at: <https://doi.org/10.1080/24721840.2019.1604138>.

Dominique Estival holds a PhD in Linguistics from the University of Pennsylvania (USA) and a Commercial Pilot License. She is a current Flight Instructor for general and recreational aviation in Australia and was approved as an assessor for General English Language Proficiency (GELP) and Level 6 AELP by CASA (the Australian Civil Aviation Safety Authority). She is a researcher at the MARCS Institute, Western Sydney University, Australia, and the co-author of Aviation English: A Lingua Franca for Pilots and Air Traffic Controllers (2016).

Brett Molesworth is an Associate Professor in the School of Aviation at UNSW Sydney. He is a Human Factors specialist with qualifications in Psychology (PhD and is a registered Psychologist in Australia) and in Aviation (BAV Hons and Commercial Pilot Licence). Brett's research focuses on understanding human performance in complex socio-technical environments.