Air Traffic Communication in a Second Language: Implications of Cognitive Factors for Training and Assessment

CANDACE FARRIS, PAVEL TROFIMOVICH, NORMAN SEGALOWITZ, and ELIZABETH GATBONTON
Concordia University
Montréal, Québec, Canada

This study investigated the effects of second language (L2) proficiency and task-induced cognitive workload on participants' speech production and retention of information in an environment designed to simulate the demands faced by pilots receiving instructions from air-traffic controllers. Three groups of 20 participants (one native-English-speaking group, two native-Mandarin-speaking groups of relatively high and low levels of English proficiency) played the role of pilots. Participants listened to, repeated, and responded to simulated air-traffic controller messages (in English) under conditions of low and high workload. In the high workload condition, participants performed a concurrent arithmetic task while repeating the messages. The dependent variables were message repetition accuracy and speech production (accentedness, comprehensibility, fluency, as perceived by 10 native-English-speaking raters). The native English speaker group repeated messages more accurately than both L2 groups, and the low-proficiency group repeated messages less accurately in the high workload condition than in the low workload condition. The native speaker and the low-proficiency groups were perceived as less fluent in the high than in the low workload condition, and only the low-proficiency group's speech was perceived as more accented in the high than in the low workload condition. Implications for language training and assessment for English for specific purposes are discussed.

As they plan curricula and design activities, language instructors often ask, "What is the most effective way for my learners to acquire the language skills they need?" In the English for specific purposes (ESP) classroom, this question becomes particularly pertinent because ESP learners often come equipped with clear and immediate objectives, usually determined by workplace requirements. However, despite the strong relationship between ESP and language use in the workplace, the cognitive challenges inherent in the learner's communicative workplace environment and the effect of those challenges on their ability to communicate in their second language (L2) are not often researched.

The communicative environment of pilots and air-traffic controllers (hereafter, controllers) provides an excellent example of challenges in the workplace. Some of the challenges specific to controller–pilot communications include L2 usage, high workload, and the inherent complexity of radiotelephonic communications (e.g., invisible and unfamiliar interlocutor, congestion due to high traffic, radiotelephonic frequency constraints). Controller–pilot miscommunications, particularly those related to L2 use in nonroutine, stressful, high-workload situations, can threaten air safety. The objective of the present research was therefore to determine how increased cognitive workload, a factor common in the controller–pilot work environment, affects controller–pilot communications conducted in an L2. Our goal was to explore the implications of cognitive factors for the training and assessment of professionals whose jobs involve high cognitive workload.

LINGUISTIC CHALLENGES IN CONTROLLER–PILOT COMMUNICATIONS

High-profile accidents in which hundreds of people lost their lives and in which miscommunications played a significant role have heightened awareness of the importance of L2 proficiency for controllers and pilots. Based on data from several accident-reporting databases, the International Civil Aviation Organization (ICAO) has identified controllers' and pilots' inadequate L2 proficiency as a major challenge to effective controller–pilot communications (ICAO, 2004, p. 1-1). In recognition of this challenge, ICAO has introduced language proficiency requirements to ensure that all controllers and pilots are proficient in the language(s) used in air–ground communications. In the international aviation context this language is often English, an L2 for many of the world's controllers and pilots.

The ICAO language proficiency requirements, to be applied to all languages used in radiotelephony, stipulate that English be made available in situations where the flight crew and the ground do not share the same language. Therefore, all pilots and controllers involved in flight operations where the use of English may be necessary are required to demonstrate an operational level of proficiency in English, as defined in the ICAO Language Proficiency Rating Scale (ICAO, 2004, A8–A9). As of March 2011, all ICAO contracting states will have had to comply with these new standards, and with this deadline fast approaching, the aviation community is faced with the task of training and testing thousands of pilots, controllers, and other personnel. One of the first steps in accomplishing this task involves understanding the controller–pilot communicative environment.

Controller–Pilot Communicative Environment

In order to ensure accuracy (and, ultimately, air traffic safety), controller–pilot communications follow a collaborative scheme involving three phases: initiate, present, and accept (Morrow, Lee, & Rodvold, 1991). Generally, the pilot first initiates radio communications with the controller (*initiation phase*), after which the controller gives the pilot instructions (*presentation phase*). A major component of the pilot's role involves remembering the instructions given by the controller long enough to repeat them (*acceptance phase*) and to subsequently act on them (e.g., by navigating the aircraft according to the controller's instructions). The acceptance phase is important because it provides an opportunity for the controller to verify that the pilot has understood the instructions correctly (Morrow et al., 1991, p. 278).

Controllers and pilots work under varying workload conditions and may be required to perform several tasks concurrently, resulting in a high cognitive workload. *Cognitive workload* may be loosely defined as the amount of cognitive resources required for task performance. Many factors may contribute to it, such as the type of task, the number of tasks performed concurrently, and personal characteristics of the individual performing the task. Cognitive load theory (e.g., Sweller, 1994) and models of working memory (e.g., Baddeley, 2003), which hold that humans have limited resources for information storage and processing, have particular significance for the work of controllers and pilots, where information is exchanged and acted on, often under time constraints, in a concurrent multitask environment. For example, a pilot may be required to execute a checklist or perform a calculation while receiving controller instructions. Although these tasks may be routine and highly practiced, each requires attention nonetheless. When performed concurrently, even simple tasks place high demands on the pilot's limited-capacity working memory, where incoming information is stored and processed.

Workload Effects on Task Performance

The effect of concurrent tasks on native speakers' performance in an aviation context is relatively well documented. Concurrent tasks produce a detrimental effect on performance in a variety of tasks, including high-priority tasks (Loukopoulos, Dismukes, & Barshi, 2003; Raby & Wickens, 1994). In observations of pilots' behavior in the cockpit, Loukopoulos et al., for instance, noted that in the classroom tasks are practiced in a linear fashion, but in the cockpit these same tasks often have to be performed concurrently, thus increasing the risk of pilot forgetting or error. The well-known pilot task prioritization maxim "aviate–navigate–communicate" succinctly summarizes the concurrent task environment of pilots but does not necessarily reflect the complexity and interdependence of these tasks. The tasks associated with aviation and navigation (higher priority tasks) often depend on accurate communications with the controller (regarded as a lower priority task, according to the task prioritization maxim). For example, the controller provides the pilot with important navigational instructions, such as heading, speed, and altitude, which the pilot then repeats and carries out in navigating the aircraft. Therefore, effective controller–pilot communications are critical to air safety.

Although concurrent task performance involving communication creates challenges for all speakers and listeners, it may create special challenges when L2 communications are involved. In fact, there are no studies known to us that have investigated the effects of cognitive workload (i.e., workload resulting from a concurrent nonlinguistic task) on L2 speech production. Previous studies investigating this issue in L1 speech processing have found that high cognitive workload imposed by concurrent task performance leads to measurable changes to speech in comparison with speech produced without other accompanying tasks or under low cognitive workload (see, e.g., Dromey & Benson, 2003; Jou & Harris, 1992). Among these studies, at least one has examined listener reactions to L1 speech produced under high cognitive workload (Lively, Pisoni, Van Summers, & Bernacki, 1993). It appears that speakers' responses to cognitive workload vary considerably, and that robust changes to speech resulting from high cognitive workload are perceptible to listeners. If high cognitive workload affects the quality of L1 speech to an extent perceptible to listeners, then it is important to investigate the extent to which it does so with L2 speech, especially given the importance of accurate L2 communications to air-traffic safety.

THE CURRENT STUDY

Recognizing the paucity of research on cognitive workload in L2 communication and the practical implications of such research for work-related L2 training and assessment, we investigated the effects of task demands on L2 speech production in a simulated pilot navigation task. Forty L2 English speakers of two proficiency levels and 20 native English speakers played the role of pilot in this task. The participants first received recorded oral instructions, then repeated them, and finally carried them out by navigating grids on a computer screen. All participants completed this task under two conditions: while performing a concurrent mental arithmetic task (*high workload condition*) and without such a task (*low workload condition*). For the purposes of this article, we analyzed the participants' speech as they repeated the instructions, but excluded from consideration the data relating to navigation accuracy. We compared how accurately they repeated the instructions under high and low cognitive workload, and how accented, comprehensible, and fluent they sounded. Our objective was to determine how two factors (L2 proficiency and degree of workload) might influence measures of message repetition and speech production in a simulated pilot navigation task.

METHOD

Participants

The original pool of participants included 62 engineering students (47 male, 15 female) from Montréal English-medium universities (mean age: 27; range: 19–41). Subsequently, the data from two participants (both male) were excluded. One was unable to perform some of the tasks; the data for the other were lost due to a malfunction in the recording equipment. The remaining 60 participants were divided into three groups. The native speaker group (henceforth, NS) included 20 native English speakers. The remaining two groups of 20 included native Mandarin speakers who had arrived in Canada as adults to pursue postsecondary education.

The L2 speakers were divided into two proficiency groups (henceforth, *high* and *low*) based on three sets of measures. The first proficiency measure was derived from a listening comprehension test, a diagnostic pretest used for TOEFL preparation (Phillips, 2005). In this test, the participants heard a simple conversation and a lecture and, following each one, responded to six multiple-choice comprehension questions (on paper). Each participant thus obtained a listening comprehension score out of 12.

The remaining two sets of proficiency measures were derived from an oral interview with each participant, a 2-minute monologue in response to a simple prompt (e.g., "Describe your recent trip/vacation"). Each participant's interview was first transcribed in order to obtain lexical and morphosyntactic error counts. *Lexical errors* were defined as incorrectly used words or phrases; *morphosyntactic errors* were defined as mistakes in sentence structure, morphology, or syntax. For each participant, a speaking accuracy score was defined as a proportion of errors, calculated by dividing the total number of lexical and morphosyntactic errors by the total number of words in the speech sample.

Finally, a brief excerpt (about 20 seconds) from each participant's interview was presented to a panel of judges for ratings of accentedness, comprehensibility, and fluency using nine-point Likert scales. The judges were 10 native English speakers (7 female, 3 male) from English-medium universities (mean age: 24; range: 19–33). For accentedness (1 = *heavily accented*, 9 = *not accented at all*), the judges were told to estimate the degree of foreign accent in the participants' speech, disregarding acceptable pronunciations typical of native regional varieties of English. For comprehensibility (1 = *hard to understand*, 9 = *easy to understand*), the judges were instructed to rate how difficult or easy it was to understand what the participants were saying. For fluency (1 = *not fluent at all*, 9 = *very fluent*), the judges were asked to rate the degree to which the participants' speech sounded fluent (i.e., spoken without undue pauses, filled pauses, hesitations, or dysfluencies such as false starts and repetitions). Accentedness, comprehensibility, and fluency scores were calculated for each participant by averaging the 10 judges' ratings of each speech sample.

For all sets of measures, one-way analyses of variance (ANOVAs) comparing the three participant groups yielded significant F ratios, $F(2, 57)$ values > 27.35, p values < 0.0001. Tukey honestly significant difference (HSD) posthoc tests showed that the three proficiency groups significantly differed from one another for all proficiency measures ($p < 0.05$). Mean values and standard deviations for all proficiency measures are presented for each group in Table 1.

Materials and Procedure

We used a simulated pilot navigation task adapted from previous studies of controller–pilot communications (Barshi, 1997; Barshi & Healy, 1998, 2002; Schneider, Healy, & Barshi, 2004). The task was modified to

TABLE 1
Means (*M*) and Standard Deviations (*SD*) for Each Proficiency Measure

	Group					
	Native speaker		High		Low	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
Listening comprehension	11.06	0.84	9.13	1.26	7.89	1.81
Speaking accuracy	0.01	0.00	0.14	0.05	0.19	0.06
Accentedness	8.53	0.37	3.65	1.14	2.55	0.57
Comprehensibility	8.66	0.56	5.15	1.21	3.86	0.76
Fluency	8.68	0.25	5.05	1.01	3.91	0.74

include a high workload condition. In the low workload condition, the participants listened to and repeated recorded messages spoken by a male native English speaker. These messages simulated authentic controller–pilot communications; they corresponded to the order of controller instructions for heading, altitude, and radio frequency but contained no aviation jargon (Barshi, 1997). The messages were one to three commands in length and contained instructions for navigation on a stack of four 4-by-4 grids displayed on a computer screen. Examples are as follows. One command: *Turn right two squares*; two commands: *Turn right one square, climb down one level*; three commands: *Turn left one square, climb up one level, move forward one step.* On hearing each message, the participants first repeated it in its entirety and then carried out the commands it contained by clicking on the appropriate squares in the grids.

The high workload condition differed from the low workload condition in that participants performed a mental arithmetic task while repeating each message. More specifically, a number between 11 and 99 would appear randomly in one of six boxes surrounding the navigation grids 0.5 seconds after the message was heard (e.g., 57). The participant would then mentally reverse the digits and add the original and reversed number while repeating the message (e.g., 57 + 75). The participant would then utter the solution to the arithmetic problem immediately after repeating the message (*Turn left one square. Climb up one level. Move forward one step. Answer: 132*). The arithmetic task was therefore concurrent with the speaking task.

The experiment involved a within-subjects design. That is, all participants performed the task under both workload conditions (low and high), with the order of conditions counterbalanced across participants. Each condition consisted of 36 messages (12 of each length) preceded by 12 practice trials (4 of each length). The entire testing session was audio recorded for later analysis.

To analyze listener reactions to participants' message repetitions, speech samples were excised from the participants' recordings and were subsequently played to 10 raters who rated them on nine-point Likert scales for perceived accentedness, comprehensibility, and fluency (Derwing & Munro, 1997). These raters (2 males, 8 females), who were university students with no language teaching experience (mean age: 25; range: 20–35), were different individuals from those who participated in the language proficiency rating described earlier. However, the same criteria were used as before. For accentedness, the raters estimated the degree of foreign accent in each repeated message. For comprehensibility, they judged how difficult or easy it was to understand each repeated message. For fluency, they judged the degree to which each message was repeated without undue pauses, hesitations, or dysfluencies. A total of 12 speech samples per participant (two for each message length in each workload condition) for a total of 720 samples were presented to the raters in 8 blocks of 90 samples each. The rating data were collected in separate sessions, and the order of presentation of blocks was counterbalanced. The raters were allowed to take frequent breaks.

Data Analysis

We analyzed the participants' accuracy in repeating messages (henceforth, *repetition accuracy*) and their accentedness, comprehensibility, and fluency as a function of English proficiency group (*NS, high, low*) and workload condition (*high, low*). For this report, we did not analyze the data separately for messages of different lengths; thus, the measures of repetition accuracy, accentedness, comprehensibility, and fluency represent averaged scores for messages of one to three commands. Repetition accuracy was treated as a measure of the ability of those role-playing pilots to retain the information contained in the simulated controller messages. Accentedness, comprehensibility, and fluency were chosen as dependent measures because these aspects of speech production potentially have an impact on the accuracy and efficiency of controller–pilot communications.

The participants' repetition accuracy was scored using a strict method adopted in a previous study of controller–pilot communications (Schneider et al., 2004). For each participant, a repetition accuracy score was calculated based on whether the words essential for accurate navigation were repeated for each message. For example, in the command *Turn left one square,* the words *left* and *one* are essential for accurate navigation because they carry information that is critical for enacting the command. By contrast, the words *turn* and *square* carry nonessential, largely redundant information. If the participant repeated all essential words in each command, this participant received a point for that message. With 36 messages repeated per condition, a total score out of 36 was calculated for each participant in each workload condition.

Each participant's accentedness, comprehensibility, and fluency ratings in each workload condition represented averages across the 10 raters' ratings for the six message repetitions in the low workload condition and the six message repetitions in the high workload condition.

RESULTS

Repetition Accuracy

We first analyzed repetition accuracy scores in the low and high workload conditions to determine if L2 proficiency and workload were con-

TABLE 2
Means (*M*) and Standard Deviations (*SD*) for Repetition Accuracy in the High and Low Workload Conditions

	Workload			
	Low		High	
Group	*M*	*SD*	*M*	*SD*
Native speaker	33.35	1.02	32.50	1.10
High proficiency	30.65	1.54	28.95	1.78
Low proficiency	30.88	1.25	27.80	1.59

tributory factors (Table 2). Our analyses consisted of five planned orthogonal contrasts (Bonferroni corrected $\alpha = 0.01$). To determine the effect of L2 proficiency, we first computed two of these orthogonal contrasts: one between the NS group and both L2 groups combined (high and low), the other between the high and the low groups. These comparisons revealed that the NS group repeated messages more accurately than both the high and low groups, $t(57) = 4.25$, $p < 0.0001$, r (effect size) = 0.49, but that the high and low groups did not differ in their repetition accuracy, $t(57) = 0.53$, *ns*. To determine the effect of workload, we then computed the remaining three orthogonal contrasts, comparing each group's scores under the high and low workload conditions. These comparisons revealed that only the low group repeated messages significantly less accurately in the high than in the low workload condition, $t(57) = 3.50$, $p < 0.001$, $r = 0.42$.[1]

Accentedness, Comprehensibility, Fluency

As in the previous analysis, we analyzed accentedness, comprehensibility, and fluency ratings as a function of L2 proficiency and workload (Table 3). We first computed two orthogonal contrasts (similar to those described earlier) for each speech measure to determine the effects of L2 proficiency. These comparisons revealed that for all measures the NS group received higher ratings than both the high and low groups, $t(57)$ values > 21.92, p values < 0.00001, r values = 0.95–0.96, and that the high group received higher ratings than the low group, $t(57)$ values > 3.81, p values < 0.001, r values = 0.45–0.53. We then computed the remaining three orthogonal contrasts (similar to those described earlier) for each speech measure to determine the effects of workload. These compari-

[1] According to Field (2005), an effect size of 0.30 represents a difference of a medium magnitude, and an effect size of about 0.50 or above represents a difference of a large magnitude.

TABLE 3
Means (*M*) and Standard Deviations (*SD*) for Speech Ratings in the Low and High Workload Conditions

	Accentedness				Comprehensibility				Fluency			
	Low workload		High workload		Low workload		High workload		Low workload		High workload	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Native speaker	8.40	0.39	8.35	0.51	8.37	0.35	8.30	0.51	8.36	0.41	8.01	0.68
High proficiency	3.69	1.18	3.61	1.25	4.76	1.09	4.61	1.07	4.88	0.98	4.71	0.93
Low proficiency	2.87	0.55	2.69	1.19	3.75	0.88	3.54	0.73	4.22	0.82	3.91	0.75

sons revealed that the NS group received significantly lower fluency ratings in the high workload condition than in the low, $t(57) = 3.45$, $p < 0.001$, $r = 0.42$, and that the low group received significantly lower accentedness and lower fluency ratings in the high workload condition than in the low, $t(57)$ values > 2.89, p values < 0.001, r values = 0.36–0.38.

DISCUSSION

Summary of Findings

We investigated the effects of cognitive workload on L2 speakers' repetition accuracy and speech production (as judged by listeners) in a simulated pilot navigation task. Results revealed that the NS group repeated messages with greater accuracy than both L2 groups regardless of workload condition, and that the group with the lowest level of L2 proficiency was the one most affected by high cognitive workload. This finding suggests that L2 communications with controllers may be more challenging for pilots when they perform one or perhaps even more concurrent cognitive tasks.

Results also revealed that the NS group sounded less accented, more comprehensible, and more fluent than both L2 groups, while the high group, in turn, received higher ratings for all these measures than the low group. In addition, high workload led to lower fluency ratings for the NS group and lower accentedness and fluency ratings for the low group than did low workload. With respect to the fluency ratings, our findings suggest that high workload is associated with the production of dysfluencies such as undue or long pauses, false starts and repetitions, to an extent perceptible by listeners. Although the additional cognitive demands of the high workload condition did not affect repetition accuracy (at least for the NS group), these demands did affect speech fluency, suggesting that fluency measures may be good indicators of the impact of cognitive workload, even when repetition accuracy is stable. With respect to accentedness ratings, the findings suggest that low-proficiency L2 users depart even more from native-like, unaccented speech under high cognitive workload, although this increased workload may not necessarily make their speech less comprehensible.

Controller–Pilot Communications and L2 Processing

The finding that workload affects the amount of information retained and influences listener perceptions of speech (especially in the L2) is compatible with existing L2 processing research. For example, this finding is in accordance with conceptualizations of the role of automaticity in language processing. Such conceptualizations hold that well-practiced skills (e.g., L1 perception and production) are more highly automatic and require fewer attentional resources than newly acquired skills, such as L2 perception and production for low-proficiency L2 users (see Segalowitz & Hulstijn, 2005). Low-proficiency speakers thus appear to have greater difficulty than high-proficiency speakers in using their L2 perception and production skills in an efficient, automatic manner. When low-proficiency learners' attentional resources are distributed across several tasks, these learners appear to engage in a nonautomatic, effortful form of processing. The result is that less information is accurately retained and more accented and less fluent speech is produced.

IMPLICATIONS FOR L2 TRAINING AND ASSESSMENT

The findings have implications for L2 training and assessment, particularly in an ESP context. The current study revealed performance differences as a function of language proficiency, but the distinction between high and low proficiency was only relative here. Clearly, the notion of L2 proficiency needs to be clarified in practical terms, as it applies to thousands of professionals who will be tested according to ICAO's language proficiency requirements. The existing ICAO Language Proficiency Rating Scale is a globally recognized instrument reflecting six language skills (*pronunciation, structure, vocabulary, fluency, comprehension,* and *interactions*) and six proficiency levels (*pre-elementary, elementary, preoperational, operational, extended,* and *expert*). It may be important to continue fine-tuning this scale, validating it using a large population of pilots and controllers under conditions of varying workload or psychological stress typical of the controller–pilot workplace. It may also be useful to develop more objective measures of L2 proficiency in relation to the ICAO language proficiency requirements, because these measures may prove useful in ensuring the reliability of the ICAO Language Proficiency Rating Scale, and the quality of instructional materials and curricula used in language training for controllers and pilots.

Other implications of the findings are practical in nature. For training and assessment purposes, especially in the ESP context, learners may benefit from practicing their L2 skills under workload conditions similar to those they might face in the workplace. Instructors may wish to use scaffolding techniques, involving simple-to-complex sequencing of subtasks within a whole-task paradigm (see van Merrienboer, Kirschner, & Kester, 2003) in order to reduce cognitive load for the learner when learning complex tasks. This and other pedagogical interventions can often be accomplished without much specialized equipment. For example, to simulate a concurrent task environment that is similar in its cognitive demands to that of pilots, learners could solve a nonlinguistic puzzle or do an arithmetic task while communicating with a partner or in a group. Another example of increasing the cognitive demands of a language task may be to simulate the constraints of radio communications, such as monitoring and filtering for relevant information while listening to the communications of others and waiting for an opportunity to speak.

Similarly, teachers might design paired communicative activities in which interlocutors do not see one another, as in real controller–pilot communications. In setting up listening activities, teachers could also vary the regional variety of English and expose learners to English spoken by speakers of different language backgrounds, thus simulating the linguistic diversity which characterizes aviation English. Although the technical requirements may be greater, instructors could set up activities that demonstrate the effects of radio frequency constraints on phonetic perception (e.g., showing that /f/ is often indistinguishable from /s/ in radiotelephonic communications). Teachers might also simply increase task demands, for example, by reducing the amount of preparation time available or by setting a time limit to complete a listening or speaking task (Skehan & Foster, 1997).

Whatever pedagogical decisions ESP instructors make, they need not become absolute experts in the learners' field. A mere familiarization with the cognitive challenges and the communicative environment of the learners' workplace would go a long way in helping instructors to make sure that learners can cope with the constraints and challenges of real-life communications. Ultimately, this will ensure that learners meet their objective—achieving language proficiency adequate for their workplace.

ACKNOWLEDGMENTS

This research, based on Farris's master's thesis, was made possible through a grant to Pavel Trofimovich from the Fonds Québécois de la Recherche sur la Société et la Culture (FQRSC) and the support of the Centre for the Study of Learning and Performance. We extend a special thank you to Immanuel Barshi of NASA-Ames Research Center for his helpful comments and for lending the software for the navigation task. We also thank Randall Halter, Alice Healy, Vicki Schneider, Andrea Kneeland, Tarek Nocharli, and Graham Elliott for their advice and help along the way, and all of the participants and raters, without whom this work would not have been possible. We are grateful to John Field and two anonymous reviewers for their helpful feedback on earlier drafts of this manuscript.

THE AUTHORS

Candace Farris completed her master of arts degree in applied linguistics at Concordia University in Montréal, Québec, Canada, and she is now pursuing a doctorate at McGill University in Montréal. Her research focuses on cognitive factors in language assessments for pilots and controllers. She has extensive experience teaching English for specific purposes in aviation and other industries.

Pavel Trofimovich is an associate professor of applied linguistics at the TESL Centre in the Department of Education at Concordia University in Montréal, Québec, Canada. His research and teaching focus on cognitive aspects of second language processing, second language phonology, sociolinguistic aspects of second language acquisition, and computer-assisted language learning.

Norman Segalowitz is a professor of psychology at Concordia University, Montréal, Québec, Canada. His research focuses on the cognitive processes underlying adult second-language acquisition and use. In particular, he addresses the roles that automatic and attention-based processes may play in fluency development and the implications these roles may have for language instruction.

Elizabeth Gatbonton is an associate professor of applied linguistics at the TESL Centre in the Department of Education at Concordia University in Montréal, Quebec, Canada. Her research and teaching focus on sociolinguistic aspects of bilingualism, teacher training, teaching methodology, second language phonology, and the role of formulaicity in the acquisition of fluency.

REFERENCES

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36,* 189–208.

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

Barshi, I., & Healy, A. F. (2002). The effects of mental representation on performance in a navigation task. *Memory & Cognition, 30,* 1189–1203.

Barshi, I., & Healy, A. (1998). Misunderstandings in voice communication: Effects of fluency in a second language. In A. F. Healy & L. E. Bourne (Eds.), *Foreign* *language learning: Psycholinguistic studies in training and retention* (pp. 161–192). Mahwah, NJ: Erlbaum.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. Evidence from four L1s. *Studies in Second Language Acquisition, 19,* 1–16.

Dromey, C., & Benson, A. (2003). Effects of concurrent motor, linguistic, and cognitive tasks on speech motor performance. *Journal of Speech, Language, and Hearing Research, 46,* 1234–1246.

Field, A. P. (2005). *Discovering statistics using SPSS (Second Edition).* London: Sage.

International Civil Aviation Organization. (2004). *Manual on the implementation of ICAO language proficiency requirements.* (Doc 9835 AN/453). Montreal, Quebec, Canada: Author. Available from http://caa.gateway.bg/upload/docs/9835_1_ed.pdf

Jou, J., & Harris, R. J. (1992). The effect of divided attention on speech production. *Bulletin of the Psychonomic Society, 30,* 301–304.

Lively, S. E., Pisoni, D. B., Van Summers, W., & Bernacki, R. H. (1993). Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences. *The Journal of the Acoustical Society of America, 93,* 2962–2973.

Loukopoulos, L., Dismukes, R. K., & Barshi, I. (2003, April). *Concurrent task demands in the cockpit: Challenges and vulnerabilities in routine flight operations.* Paper presented at the 12th International Symposium on Aviation Psychology, Dayton, OH, United States.

Morrow, D., Lee, A., & Rodvold, M. (1991, April). *Collaboration in pilot–controller communication.* Paper presented at the Sixth International Symposium of Aviation Psychology, Columbus, OH, United States.

Phillips, D. (2005). *Longman preparation for the TOEFL test: Next generation iBT.* New York: Pearson/Longman.

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

Schneider, V. I., Healy, A., & Barshi, I. (2004). Effects of instruction modality and readback on accuracy in following navigation commands. *Journal of Experimental Psychology: Applied, 10,* 245–257.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll. & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). New York: Oxford University Press.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1,* 185–211.

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4,* 294–312.

Van Merrienboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off the learner's mind: Instructional design for complex learning. *Educational Psychologist, 38,* 5–13.