Utilisation of Prosodic and Semantic Cues of Nonunderstanding during Readback/Hearback Verbal Transactions in Technical Environments

Jaime Auton
PhD/Master of Organisational Psychology Candidate
Macquarie University, Sydney, Australia

Bachelor of Arts (Honours, Psychology), University of Sydney

Submitted on the 14th day of October, 2014

Table of Contents

Table of Contents **3**

Declaration **7**

Acknowledgments **9**

Summary **11**

Overview of Introduction **13**

Introduction: Section 1 ..... **15**

Nonunderstandings in Operational Environments **17**

Communication and Grounding 17
Problematic Communication 19
Nonunderstandings and Radio Communication 21
Repair and Recovery from Nonunderstandings 25
Modes of Communication 27
Radio Communication 29
Functional Aspects of Prosody 29
Prosodic Cues of Uncertainty 30
Procedural Variations to the Full Readback Requirement 33
Semantic Cues of Nonunderstanding 35
Sender-Initiated Repairs of Perceived Concealed Nonunderstandings 36

Introduction: Section 2 ..... **37**

Situation Assessment in Operational Environments **39**

Theoretical Models of Situation Assessment 39
Empirical Support for Recognition-Primed Situation Assessment 50
Limitation of the Construct of Cue Utilisation Underlying Expert Performance 52

Introduction: Section 3 ..... **55**

Research Aims and Design **57**

Specific Research Questions 59
Measuring Cue Utilisation 60
Domains of Research Interest 62
Summary of Thesis Contributions 63

Paper 1: Prosodic Cues and Perceptions of Nonunderstanding **65**

Paper 1 Aim 65
Paper 1 Publication History 65

Paper 1 ..... **67**

Introduction 69
Method 75
Results 82
Discussion 84
Paper 1 References 89

Paper 1 Appendix 95

Paper 2: Relationship between Visual and Auditory Cue Utilisation **97**

Paper 2 Aim 97
Paper 2 Publication History 98

Paper 2 .....**99**

Introduction.....100
Method.....107
Results.....116
Discussion.....121
Paper 2 References 125
Paper 2 Appendices 133

Paper 3: Relationship between Visual and Auditory Cue Utilisation (Extended) **141**

Paper 3 Aims.....141
Paper 3 Publication History 142

Paper 3 .....**143**

Introduction.....144
Method.....153
Results.....168
Discussion.....172
Paper 3 References 177
Paper 3 Appendices 187

Paper 4: Prosodic and Semantic Cue Utilisation in Perceptions of Nonunderstandings.....**195**

Paper 4 Aims.....195
Paper 4 Publication History 196

Paper 4 .....**199**

Introduction.....201
Method.....209
Results.....222
Discussion.....230
Paper 4 References 237
Paper 4 Appendix 244

Discussion of the Research Presented **245**

Research Outcomes 245
Key Findings.....254
Prosodic Cues and the Recognition of Nonunderstandings During Readback Responses.255
Semantic Cues and the Recognition of Nonunderstandings During Readback Responses262
Typologies and Prosodic and Semantic Cue Utilisation.....263
Limitations and Future Directions 264
Research Strengths.....268
General Conclusion 269

Complete Reference List **271**

Appendix A: Journal of Communication (2013) Publication (Removed for Copyright Purposes) **299**

Appendix B: Human Factors and Ergonomics Society Proceedings Publication (2012) (Removed for Copyright Purposes) **301**

Appendix C: Numeric Digit Span Recall Task and Manipulation Check Description **303**

Appendix D: Paper 4 EXPERTise 1.0 Stimuli **321**

Appendix E: Human Factors and Ergonomics Society Proceedings Publication (2014) (Removed for Copyright Purposes) **331**

Appendix F: Cluster Analysis Procedure **333**

Appendix G: Ethics Approval Letters **337**

Declaration

This thesis has not been submitted for a higher degree to any other university or institution. The work is predominantly that of the PhD candidate. Components of the thesis that involved collaboration have been noted as such. The Macquarie University Ethics Committee approved the research reported in this thesis on the 22nd of March 2011 (Reference No: 5201000458D) and the 19th March 2013 (Reference No: 5201300099).

Jaime Auton 14th October 2014

Acknowledgments

I would like to take this opportunity to acknowledge the individuals who have contributed in some capacity to my PhD ‘journey’ and now, thankfully, my submission. I would first and foremost like to thank my supervisor, Professor Mark Wiggins, for his significant contribution to this thesis. Throughout my entire candidature, Mark has been very generous with his time, expertise, feedback and overall support. His enthusiasm for my research topic, and psychological research and practice in general, has had an immeasurable influence on my development as a young researcher and future psychologist. I am very grateful to have had such a wonderful mentor throughout this process, which has resulted in a thesis I am proud to submit.

I would also like to thank my associate supervisor, Dr. Ben Searle, who provided considered and knowledgeable feedback to all components of my thesis. His assistance and advice was invaluable. I would also like to acknowledge my second associate supervisor, Dr. Nan Xu Rattanasone, whose linguistic expertise was instrumental in the development of the auditory stimuli described in this thesis. A big thanks to my good friend, colleague and co-author, Dr. Thom Loveday, who, in addition to providing sage advice on a variety of academic matters, always managed to make me see the light at the end of the tunnel.

Lastly, I would like to thank my friends and family. While the majority of them managed to escape the last few years without *really* knowing what I did all day, I could not have done this without their unwavering support. A special mention to my parents, Dad, Mum and SP, who have always been incredibly supportive of my academic pursuits; my beautiful sisters, Emilie and Stephanie, who will be thrilled that they will not have to hear about my “double degree” ever again; and to Aunty (Sister) Celine, who is sadly not finishing this journey with me, but has played a special role in my education. And finally, to my best friends, Maeve, Ana, and Charlotte, who provided so much love, support and laughter, which has helped me reach the end of this journey with a smile on my face.

Summary

The readback/hearback protocol is a communicative procedure used within many technical environments to minimise the risk of communication errors over the radio or telephone. This protocol requires the receiver of a verbal instruction to repeat or 'read back' the instruction to the sender to ensure that it has been heard correctly. It does not, however, ensure that it has been *understood* correctly. An instance where a receiver has failed to understand an instruction is referred to as a *nonunderstanding*. In high-risk, high-consequence environments, nonunderstandings pose the greatest risk when they are concealed by a receiver, as the sender is consequently unaware that an error has taken place. The primary aim of this program of research was to identify the prosodic and semantic cues that are used by senders to recognise nonunderstandings during readback/hearback exchanges.

An allied aim of the thesis was based on the reasonably consistent observation that experienced operators within technical environments have a greater capacity to extract and utilise visual cues in their environment to understand a situation, compared to their less experienced counterparts. This observation, however, has been limited to the investigation of the use of *visual* cues to the exclusion of those from other modalities. The current thesis, therefore, examined whether operators who demonstrated a relatively greater capacity for the utilisation of visual cues also demonstrated the same capability in the utilisation of auditory (prosodic and semantic) cues during assessments of nonunderstanding.

Within the context of electricity transmission control, Paper 1 examined whether the prosodic cues that listeners use to interpret uncertainty (namely intonation, inter-turn delay and filler), are also used to judge instances of nonunderstanding. Intonation and filler were identified as prosodic cues that are used to discern levels of nonunderstanding on the part of the receiver ('speaker'). Within the same domain, Paper 2 found no evidence for a relationship between auditory and visual cue utilisation.

Paper 3 sought to establish whether the use of prosodic cues to detect instances of nonunderstanding reflected a universal capability or whether it reflected expertise in the use of industry-specific phraseology. Using electrical rail control operators and naïve psychology students as participants, the results revealed that the use of prosodic cues during assessments of nonunderstanding reflected a universal capability, independent of industry-related expertise.

The final study described in Paper 4, sought to establish whether there were differences in the use of prosodic cues depending upon the use of standard and non-standard phraseology amongst receivers. The results indicated that a filler acted as a prosodic cue during perceptions of nonunderstanding in the case of standard phraseology (i.e., a full readback response), and that filler and intonation acted as prosodic cues during non-standard responses (i.e., partial readback responses). It was also evident that the use of standardised phraseology provided operators with an important indicator as to the level of understanding of the receiver. Finally, greater levels of visual cue utilisation were associated with a relatively greater attention to filler and the use of standardised phraseology as indicators of nonunderstanding. The outcomes of the thesis have implications for the training and assessment of operators in environments that depend upon the accuracy of readback/hearback exchanges in reducing the potential for communication error.

Overview of Introduction

The introduction to this program of research is divided into three sections. The first section, *Nonunderstandings in Operational Environments,* draws on the communication literature to establish how incidences of concealed nonunderstandings can compromise the reliability of the readback/hearback protocol used in many operational environments. This section subsequently examines the types of auditory cues available to a sender ('listener') to assist in the recognition, and subsequent repair of concealed nonunderstandings on the part of a receiver ('speaker').

The second section, *Situation Assessment in Operational Environments*, draws on the expertise and naturalistic decision-making literature to describe the process by which experienced operators in high-risk, high-consequence domains are expected to extract and utilise environmental cues to understand, or assess the nature of a situation (such as a nonunderstanding). Although this literature has largely focused on how experienced operators use *visual* cues during situation assessment, it is posited that assessments of nonunderstanding can also be made on the basis of *auditory* cue utilisation. This section also discusses the well-established relationship between operator experience and visual cue utilisation performance during situation assessment. However, it is not clear whether this relationship exists between visual and auditory cue utilisation. Arguably, the empirical demonstration of such a relationship would provide greater support for cue utilisation as a mechanism underpinning expert situation assessment.

The final section, *Research Aims and Design*, draws together the two literatures discussed in the previous sections to provide support for the overall aims and research questions posited in the current thesis. Methodological considerations are also discussed, including the measurement of cue utilisation and domains of research interest. This section concludes by outlining the potential theoretical and practical outcomes of this thesis.

Introduction: Section 1

Nonunderstandings in Operational Environments

The greatest problem in communication is the illusion that it has been accomplished

- George Bernard Shaw

Communication and Grounding

Effective communication is central in assuring safety and productivity in many high-consequence, high-risk environments, such as aviation, power control, medicine and electrical rail control (Australian Transport Safety Bureau, 2007; Prinzo & Britton, 1993). Communication is a process that involves the transmission of a message from one communicator to another (Kanki & Smith, 2001). As a process, communication is only considered successful when the receiver[1] of a message has correctly understood the sender's intended meaning (Dascal & Berenstein, 1987; Kanki & Smith, 2001).

Communication is, inherently, a joint activity between communicative partners (Clark, 1994; Clark & Brennan, 1991). Models of dialogue structure, such as the collaborative model (Clark, 1994; Clark & Brennan, 1991; Clark & Schaefer, 1987, 1989; Clark & Wilkes-Gibbs, 1986), also referred to as the contribution model (Clark & Schaefer, 1987, 1989), emphasise that both communicative partners have a shared responsibility to ensure that information exchanged has been appropriately heard and understood before the conversation proceeds. A communicative exchange involves two stages: the presentation of content and the *grounding* of that content (Clark & Schaefer, 1987).

During the first stage of a communicative exchange, the sender will provide information for the receiver to interpret and from which meaning can be derived. During the grounding stage, the sender and receiver collaborate to establish the mutual belief that the receiver has understood the sender's intended message (Clark & Brennan, 1991; Clark & Schaefer, 1987). For grounding to take place, the receiver must exhibit evidence of understanding to the sender,

[1]The generic terms 'sender' and 'receiver' will be used in this thesis to refer to the communicative participant who initiates an utterance and the communicative participant who is the recipient of said utterance, respectively. As such terms do not necessarily denote a speaking or listening role, the terms 'speaker' and 'listener' may be used in specific instances to avoid ambiguity.

and the sender must accept the validity of such evidence. This evidence may include the receiver proceeding with a relevant next-turn, an explicit acknowledgement of understanding, or it may consist of a non-verbal cue, such as a head nod (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1990).

Although a receiver may provide such evidence of understanding after hearing a sender's message, it is not always a reliable indication that it has been understood. For example, a receiver may not recognise that he/she has not understood an utterance and consequently, may still provide evidence of understanding. In contrast, a receiver may recognise that an utterance has not been understood, but deceptively conceal this lack of understanding (Clark & Schaefer, 1987; Clark & Wilkes-Gibbs, 1990). Any failures in understanding (whether outwardly expressed by the receiver or detected as such by the sender) are resolved ideally, during this grounding stage. When the process of grounding is successful, the utterance from the sender is referred to as *common ground* between the two communicative participants (Clark & Schaefer, 1987, p. 19).

The following linguistic example, derived from Clark and Schaefer's (1987, p. 20) corpus of telephone exchanges between operators and customers, illustrates how common ground is achieved through the mutual effort of two communicative partners (in which O stands for the female operator, and C for the male customer):

O: *Name of the people please?*

C. *Mrs. Lane*

O. *Sorry, would you say that again please?*

C. *Lane*

O. (spelling) *M A I?*

C. (spelling) *L A N E*

O. *N for Nellie A N E*

C. *No, L for London*

O. *Oh! Sorry, Lane, L for Leonard*

C. *Yes, 50 High Street.*

At the beginning of this exchange, the Operator poses a question to the Caller (presentation), which is understood by the caller, as evident by her relevant response *Mrs. Lane.* The meaning of the Operator's question is now considered common ground between the communicative participants. On receiving the Caller's response to her question, the Operator indicates that there is some difficulty hearing the Caller, by requesting the Caller to repeat her response. Both communicative partners spend the subsequent utterances grounding the Caller's initial utterance, *Mrs. Lane*, until the Operator provides evidence that the utterance has finally been understood (*Oh! Sorry, Lane, L for Leonard*). This evidence of understanding is accepted as valid by the Caller in the final utterance (where *Mrs. Lane* is now common ground), when the Caller proceeds to the new topical utterance (*50 High Street*). The presentation of content and the grounding of that content are referred to as a contribution in this model of dialogue structure, where conversations comprise contributions rather than utterances (Clark & Schaefer, 1987).

Problematic Communication

Although the objective of any communicative exchange is to ensure mutual understanding between all participants, problematic communication still occurs frequently. This is especially the case in high-risk environments, since these environments are characterised by uncertain and dynamic conditions and severe time pressure that create vulnerability to error (Orasanu & Connolly, 1993). Indeed, communication has been identified as a contributing factor in a significant proportion of errors within many high-risk industries (Kim, Park, Han, & Kim, 2010). For example, Billings and Cheaney (1981) noted that information transfer problems were a primary factor in over 70% of cases reported to the National Aeronautics and Space Administration (NASA) Aviation Safety Reporting System. Similarly, Greenberg et al. (2007) reported that errors in communication contributed to almost

a quarter of surgical malpractice claims that led to patient harm.

Miscommunication is the broad term used to encompass the various types of communicative problems that can occur in discourse (Skantze, 2005). However, there are two main types of miscommunication: *misunderstandings* and *nonunderstandings* (Allwood & Abelar, 1984; Hirst, McRoy, Heeman, Edmonds, & Horton, 1994). A misunderstanding takes place when a receiver interprets a message in a way that is not consistent with the intention of the sender (Hirst et al., 1994; Skantze, 2005). During misunderstandings, a receiver is under the false impression that the correct interpretation of the message has been obtained (Skantze, 2005).

A nonunderstanding, by contrast, occurs when a receiver fails to obtain a complete understanding of a message (Allwood & Abelar, 1984; Skantze, 2005). A nonunderstanding can transpire when a receiver has either failed to obtain any meaningful interpretation of a message or cannot choose between two or more interpretations of the message (Hirst et al., 1994; Skantze, 2005). Some theorists make a distinction between instances where a receiver's understanding is non-existent (i.e., nonunderstanding) and instances where a receiver's understanding is incomplete (i.e., partial understanding; Grimshaw, 1980; Zaefferer, 1977). In contrast, other theorists use the term nonunderstanding to label instances where a receiver fails to find any "interpretation at all for some or all of the utterance" (Hirst et al., 1994, p. 215). The latter definition conceptualises nonunderstanding as a gradual phenomenon, which can vary from a 'less complete' to a 'total lack' of understanding.

Nonunderstandings are likely to occur in three contexts, the first of which occurs when receivers are unaware of the meaning of a token within a particular utterance (Allwood & Abelar, 1984). For example, if a pilot is the recipient of the following Air Traffic Controller (ATC) instruction, "*Cessna 223, Cross Seton* (a navigation reference point) *at 4000 feet*" (Morrow, Rodvold, & Lee, 1994, p. 238), a nonunderstanding may be experienced where the location of Seton is not known by the receiver.

A nonunderstanding can also occur when there is no means to associate incoming information with the information retained in memory (Allwood & Abelar, 1984). In the example above, the pilot may be aware of the location of Seton, but the route may be unclear. Finally, it could also be conceived that, even in instances where the information has been understood, a receiver may experience confusion and hence, fail to understand the correct interpretation of the message when the information presented by a sender is perceived as erroneous, incomplete, or mistimed (Morrow et al., 1994). For example, a pilot may question the altitude command with, "*I don't think we can descend that fast*" and thereby not understand the appropriate way to proceed (Morrow et al., 1994, p. 238).

The main distinction between a misunderstanding and a nonunderstanding is the extent to which the receiver is aware that a communicative problem has occurred (Hirst et al., 1994). During a misunderstanding, a receiver is under the impression that the message has been interpreted accurately and therefore, the communication error is not recognised immediately (Hirst et al., 1994; Skantze, 2005). In contrast, a nonunderstanding will be realised immediately by the receiver who has failed to understand one or more components of a message (Hirst et al., 1994).

There are two types of nonunderstandings that can occur in discourse. A *disclosed* nonunderstanding occurs when a receiver fails to understand a message and alerts the sender that the message has not been understood. In contrast, a *concealed* nonunderstanding occurs when the receiver fails to understand a message, but deceptively conceals this state of nonunderstanding from the sender (Weigand, 1999).

Nonunderstandings and Radio Communication

Nonunderstandings can occur in many different types of communicative exchanges, including everyday social interactions. However, they can also threaten the effectiveness of the readback/hearback protocol, which is a communications procedure that specifies, in high risk, high consequence environments, how verbal instructions are to be delivered and

confirmed over the radio/telephone. The protocol is used in many different technical environments, including aviation, electricity transmission control, hydroelectric power generation, electrical rail control and more recently, in some healthcare settings (Hanna, Griswold, Leape, & Bates, 2005; Prinzo & Britton, 1993).

This is a four-step protocol, the first of which involves a sender transmitting a verbal message to a receiver. For example, an ATC might send the following instruction to a pilot; "*Aircraft X, descend to flight level 210*". The receiver then actively listens to the instruction. During the third step, the receiver repeats the instruction verbatim (known as the *readback*), so that a pilot might respond to the ATC with "*Cleared to descend to 210, Aircraft X*". The sender is then expected to listen to the readback to ensure that it has been repeated correctly, which is referred to as the *hearback* portion of the exchange (Cushing, 1994, p. 40; Prinzo & Britton, 1993; Prinzo, Hendrix, & Hendrix, 2006).

The purpose of the readback/hearback protocol is to verify that the receiver has heard the instruction accurately, so that errors can be identified and corrected by the sender (Prinzo & Britton, 1993; Prinzo et al., 2006; Schneider, Healy, & Barshi, 2004). If a receiver is given an instruction that is not understood, this can be indicated to the sender in lieu of a readback (i.e., disclosed nonunderstanding), allowing the information to be re-communicated until the instruction becomes common ground between the communicative participants (Streitenberger, Breen-Reid, & Harris, 2006).

Although the readback/hearback protocol is an effective strategy for the identification of misunderstandings, it does not necessarily protect against concealed nonunderstandings, since a nonunderstanding can be concealed, despite an accurate readback (Jones, 2003). Concealed nonunderstandings on the part of the receiver in a readback/hearback exchange may also go undetected by a sender for various reasons. For example, senders (of any communicative exchange) have a tendency to overestimate the extent to which they have successfully communicated the intended meaning of a message (Fay, Page, & Serfaty, 2010; Keysar &

Henly, 2002). Therefore, they are not anticipating a lack of understanding. However, there is also an assumption in many high-risk environments, including aviation, that the receipt of an accurate readback demonstrates to the sender that the message has been correctly understood by the receiver and will be executed as requested (Morrow et al., 1994; Wickens, 2007). Yet, the delivery of an accurate readback from a receiver does not necessarily provide the sender with adequate evidence during the grounding process to indicate that mutual understanding (common ground) has taken place.

In high-consequence, high-risk environments, such as aviation and defence, *standard phraseology* is used to reduce instances of nonunderstanding and misunderstanding (Alderson, 2009; Garzone et al., 2010). Standard phraseology involves the use of highly regulated and specialised vocabularies, and is designed to reduce errors in understanding that could potentially arise from the ambiguity of particular words and phrases. However, despite the use of regulated language, "phraseology defects" still exist (Jones, 2003, p. 239). These include synonymous utterances and multiple meanings for words and phrases, thereby allowing 'understanding errors' to persist, even within high-risk communicative exchanges (Grayson & Billings, 1981; Jones, 2003). Cushing (1994, p. 1) notes how the collision between two Boeing 747s in Tenerife in 1977 was, in part, due to the misunderstanding of the meaning of the phrase *at takeoff*. This phrase was used by the flight crew to specify that they were "in the process of taking off", but was interpreted by the tower controller as signalling that the flight crew were "at the takeoff *point*"[2].

Compared to instances of disclosed nonunderstandings, those instances where a receiver conceals a nonunderstanding during a readback/hearback exchange pose the greatest risk of error, since the sender remains unaware that a failure has occurred and is provided no opportunity to establish common ground. Despite the implementation of the readback/hearback communication protocol, in addition to standard phraseology, nonunderstandings remain a common cause of communication problems within technical

[2] This terminology has since been corrected.

environments (Morrow et al., 1994).

One documented incident in the aviation context occurred in June 2007, between the crew of a LOT Polish Airlines Boing 737-500 (LOT 282) and British air traffic controllers (Air Accidents Investigation Branch, 2008). Shortly after departure from London Heathrow (LHR), LOT 282 encountered technical difficulties with the electronic flight displays while operating in instrument meteorological conditions. Although the aircraft landed safely back at LHR shortly after take-off, the crew struggled to effectively communicate with the British air traffic controllers (Air Accidents Investigation Branch, 2008).

During the communicative exchange between the air traffic controller and the Captain of LOT 282, there were several instances where it was clear that the commander did not understand the instructions he was being given, and yet he did not request clarification. For example, at one point, the controller instructed the Captain to "*Fly a heading of zero five zero degrees that'll be a right turn of approximately 90 degrees"*. Although the Captain acknowledged the instruction correctly, he subsequently failed to adhere to the directive, suggesting that he did not adequately understand what he was instructed to do. The incident report arising from the event illustrated ".. the problems that can arise when there is a lack of understanding between controllers and flight crews" (Air Accidents Investigation Branch, 2008, p. 32).

Although this case highlights how nonunderstandings can permeate readback/hearback exchanges, the extent to which concealed nonunderstandings occur within high-risk environments, and the proportion of these that result in catastrophic errors, remains unclear. This is the case for several reasons. First, many safety investigation authorities, such as the National Transportation Safety Board (NTSB), tend not to examine the contribution of linguistic factors associated with accidents and incidents (Jones, 2003). However, even if investigative organisations did scrutinise accidents for linguistic-related causes, a concealed nonunderstanding describes a particular state of mind where operators *deceptively* fail to

disclose the fact that they have not understood an instruction. Therefore, the documentation of such instances would rely on receivers admitting to deception or it becoming clear through circumstantial evidence. Even in the case of LOT 282, the evidence is circumstantial, and it remains unclear whether the Captain knew that he had failed to understand the instructions (and concealed this fact), or whether he simply misunderstood the instructions (but yet still provided accurate readbacks). Finally, not all cases of concealed nonunderstandings presumably lead to an operational error or incident and therefore, such instances would not necessarily be captured in accident data.

Repair and Recovery from Nonunderstandings

Recovering from, or *repairing* difficulties in communication, including nonunderstandings, can be either *self-initiated* (executed by the communicator of the 'trouble source') or *other-initiated* (executed by the communicator who is the recipient of the trouble source; Schegloff, 2000). In cases where a sender presents a message for a receiver to understand and the receiver indicates difficulties in understanding to the sender, this would constitute an other-initiated repair (as the 'trouble source' was the instruction that could not be understood; here forth referred to as 'receiver-initiated repair'). In cases where a receiver does conceal that the instruction has not been adequately understood, and continues the exchange with an accurate readback, a sender can initiate a repair. Arguably, this would also constitute an other-initiated repair, as the trouble source now becomes the readback that has not been understood by the receiver (here forth referred to as 'sender-initiated repair'; Schegloff, Jefferson, & Sacks, 1977; Schegloff, 2000).

Generally, the communicator who first recognises a problem in communication will initiate the repair (Bazzanella & Damiano, 1999). In the context of a nonunderstanding during a readback/hearback loop, this would normally involve the receiver seeking repetition, clarification or elaboration of the information that was ambiguous or absent (Bazzanella & Damiano, 1999; Rieger, 2003). Receiver-initiated repairs are the recommended remedy for

nonunderstandings during the readback/hearback protocol in naturalistic environments (Barshi & Farris, 2013; International Civil Aviation Organization, 2007; Monan, 1986).

There is evidence to suggest that receiver-initiated repairs of nonunderstanding do take place during communicative exchanges within high-risk environments. In their analysis of 48 hours of pilot-controller communication, Cardosi (1993, p. 6) noted that responses to eight per cent of ATC clearances involved a question, or a request for a different clearance or additional information. Nevertheless, the fact that communication issues still arise in the context of aviation suggests that not all instances of nonunderstanding are clarified (Prinzo & Britton, 1993).

The psychological basis for concealing a nonunderstanding is best considered within the context of social psychology. For example, in an explorative experiment in which participants were asked to give route directions to other participants over the telephone to particular locations within a simulated campus, Skantze (2005) observed a positive relationship between the frequency of disclosed nonunderstandings and the perception of task failure. To extrapolate into highly professional environments, it might be postulated that operators may not request clarification in the case of a nonunderstanding in an attempt to avoid negative evaluations from others. In such circumstances, operators may experience an appraisal of *threat* (to status or self esteem) resulting in a coping response of minimising harm-to-self by concealing evidence of incompetence (Searle & Auton, in press). Therefore, it might be argued that instances where receivers conceal nonunderstandings occur in an attempt to ‘save face’.

In the aviation context, “Garrison (1977) claimed that ‘all instrument pilots’ (p. 7) have experienced the feel of uncertainty about a procedure and of not wishing to admit their perplexity to the air traffic controller” (Murray, 1999, p. 405). Furthermore, Chiang and Pepper (2006) found that nurses commonly cited face-saving concerns as a barrier to reporting medication administration errors. This suggests that face-saving may represent a

primary factor associated with instances where nonunderstandings are not resolved.

Where receiver-initiated repairs of nonunderstanding are unreliable, it may the case that senders develop strategies that are intended to detect instances of nonunderstanding, particularly where the consequences of error are significant (Schegloff et al., 1977). Indeed, Clark and Schaeffer (1987) advocate that senders instigate actions to repair a potential nonunderstanding either by expanding or replacing the original utterance, to ensure common ground. However, what remains unclear is the basis on which a nonunderstanding might be detected. In a readback/hearback context in particular, the readback may have been entirely correct. Therefore, the sender must rely on other features of the communication interaction to detect instances on nonunderstanding.

Modes of Communication

During face-to-face communication, communicative participants utilise multiple sources of information to establish common ground (Massaro & Egan, 1996). In general, information can be derived from the *verbal* content of an utterance, in addition to a variety of *nonverbal* cues (Nygaard & Queen, 2008). Verbal communication refers strictly to the *semantic* (or linguistic) content of an utterance that includes the syllables, words and phrases (Nygaard & Queen, 2008; Soskin, 1953).

Nonverbal communication commonly refers to any type of communication that is independent of the words spoken (Baseheart, 1975; Knapp, Hall, & Horgan, 2014). Nonverbal communication can manifest as both *visual* cues (such as facial expressions, body gestures, posture and eye gaze) and *prosodic* cues (Baseheart, 1975; Knapp, Hall, & Horgan, 2014). Prosody refers to the array of non-linguistic features of speech, including intonation, rhythm, intensity, pitch, timing and differential pausing that enhances the meaning of the linguistic content of an utterance (Bolinger, 1985; Cruttenden, 1986; Crystal, 1969; Monnot, Nixon, Lovallo, & Ross, 2001; Nygaard, Herold, & Namy, 2009).

In comparison to face-to-face communication, voice (or auditory) communication

commonly takes place over the telephone or radio. Such interactions are distinct from face-to-face encounters insofar as they only permit an exchange of semantic and prosodic information, preventing access to all visual information during discourse (Drummond & Hopper, 1991; Kanki & Smith, 2001; Nygaard et al., 2009; Soskin, 1953).

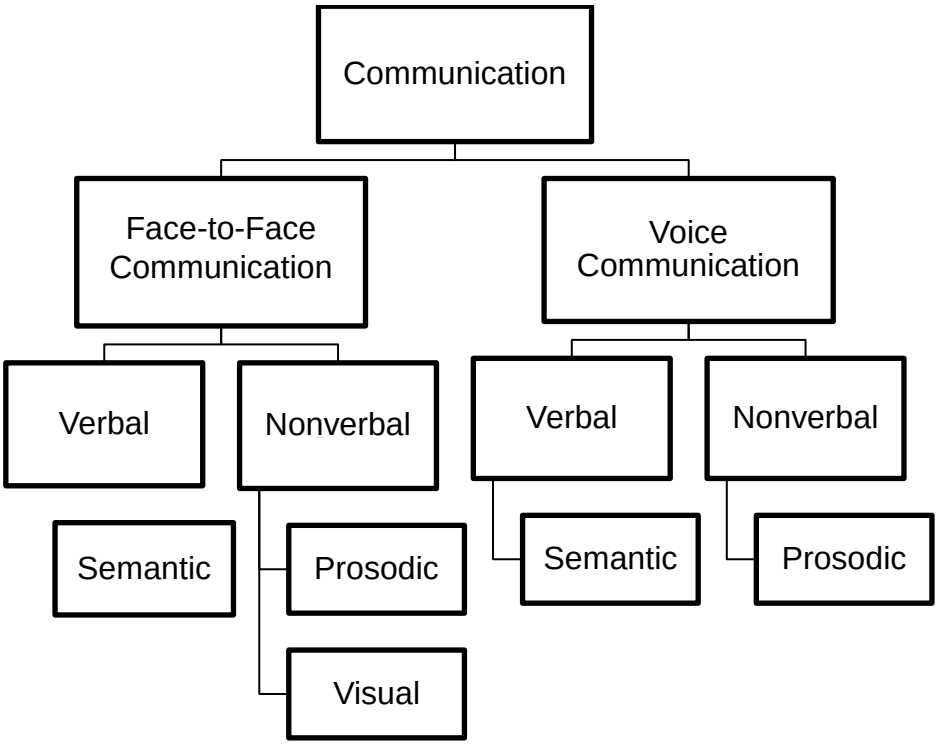


Figure 1. The channels of information that can be utilised during face-to-face and voice communication.

Communication within high-risk environments comprises either face-to-face interaction when communicators are co-located (such as between the pilot and co-pilot within the cockpit of an aircraft) or purely auditory interactions, when communicators are located remotely (such as between a pilot and an ATC; Kanki & Smith, 2001). The context of the communication can affect the efficiency and success of the communication process (Kanki & Smith, 2001; Swerts & Krahmer, 2005). As radio/telephone transactions do not involve the exchange of visual information, there is a great deal more information available during face-to-face exchanges that facilitate the process of communication. Operators working in domains that are heavily reliant on radio communication, therefore, need to successfully adapt to the loss of key visual cues.

Radio Communication

Radio communication is a common means of information exchange in many naturalistic environments, including aviation and electricity transmission control (Australian Transport Safety Bureau, 2007; Garzone et al., 2010). The readback/hearback protocol is used exclusively over the radio or a telephone to deliver and confirm instructions.

As radio communication eliminates access to all visual components of communication, efficiency in this context depends on attending, not only to the semantic content of the speech (*what* is said), but also to the prosodic features of the utterance (*how* something is said; Soskin, 1953). Fegyveresi (1997), for example, describes one particular aviation accident in which an ATC failed to detect a change in pitch or stress in the pilot's voice, which was caused by the pilot's concern with the aircraft's fuel state. As the controller did not recognise the severity of the fuel crisis, the issue was not given a high priority and the aircraft subsequently crashed through fuel exhaustion.

Despite the importance of prosody in verbal communication, it remains unclear whether aspects of prosody are used to infer a lack of understanding, particularly where there is no additional semantic information available. This is especially the case in environments where the readback/hearback protocol is employed, since a readback is effectively the reuse of an already uttered text. Since the semantic content of a readback may not vary, its capacity to signal communication difficulties is limited. In this case, a sender has little choice but to rely on the prosodic cues of the utterance as evidence of a receiver's level of understanding.

Functional Aspects of Prosody

The role of prosody in conveying a speaker's general state, above and beyond the semantic content of the utterance, is well established (Bolinger, 1985; Dromey, Silveira, & Sandor, 2005; Nygaard et al., 2009; Nygaard & Queen, 2008). For example, there is a strong and consistent relationship between prosodic cues and the production and perception of emotion (e.g., Bachorowski & Owren, 2006; Bachorowski, 1999; Banse & Scherer, 1996;

Frick, 1985; Gobl & Chasaide, 2003; Huttar, 1968; Johnson, Emde, Scherer, & Klinnert, 1986; Johnstone & Scherer, 2000; Juslin & Laukka, 2001, 2003; Juslin & Scherer, 2008; Massaro & Egan, 1996; Scherer, 1986, 2003; Schlanger, Schlanger, & Gerstman, 1976; Zuckerman, Amidon, Bishop, & Pomerantz, 1982). Although the unique prosodic profiles that correspond to the presentation of particular emotions have yet to be identified, listeners are relatively accurate in inferring different emotional states, purely on the basis of prosodic cues (Scherer, 1986).

Allied to emotional states, prosody has also been associated with perceptions of task difficulty, especially in high-risk environments (Prinzo & Britton, 1993). Although this research has generally focused on the generation of prosody, rather than the perception of prosody per se, it has been possible to identify specific prosodic cues that vary in an operator's utterance as a function of task complexity (Griffin & Williams, 1987), workload (Brenner, Shipp, Doherty, & Morrisey, 1985; Morrison & Wright, 1989) and stress (Brenner et al., 1985; Griffin & Williams, 1987). For example, in the analysis of recordings from 20 naval aviators, Griffin and Williams (1987) observed an increase in fundamental frequency and the amplitude of an utterance associated with increases in task complexity.

Prosodic Cues of Uncertainty

Investigations of the perception of prosody have largely been restricted to analyses of uncertainty, rather than nonunderstanding (Brennan & Williams, 1995; Krahmer & Swerts, 2005; Smith & Clark, 1993; Swerts & Krahmer, 2005). Uncertainty is a psychological state in which the information received from a sender is considered incomplete, ambiguous, erroneous, and/or imprecise (Woods, 1988). Arguably, it represents the precursor to a nonunderstanding but potentially involves the same, or similar prosodic cues.

The common methodological approach to the generation and assessment of levels of uncertainty has involved the Feeling of Knowing (FOK; Hart, 1965) and the Feeling of Another's Knowing (FOAK) paradigms (Brennan & Williams, 1995). The FOK paradigm is a

three-staged process, the first of which involves participants being asked a series of factual questions, where their verbal responses are tape-recorded. The second stage involves participants rating their feeling that they would recognise the correct answer to each question if it were presented. Their 'feeling of knowing' is lower if they express less certainty of knowing the correct answer. Finally, participants are asked to respond to a recognition test, where they are presented with the same factual questions as those they responded to in Stage 1 and their accuracy is recorded (Hart, 1965; Smith & Clark, 1993).

The FOAK paradigm involves participants (i.e., 'listeners') listening to verbal responses to trivia questions that are purportedly from other participants (i.e., 'speakers'). Listeners are then asked to rate the level of uncertainty perceived in the verbal responses (Brennan & Williams, 1995). The advantage of the FOAK paradigm is the experimental control that can be exercised over the delivery of the responses, particularly in terms of the systematic manipulation of prosodic features.

Using the FOK paradigm, uncertainty in the generation of utterances appears to be especially associated with changes in intonation (Brennan & Williams, 1995; Krahmer & Swerts, 2005; Smith & Clark, 1993; Swerts & Krahmer, 2005), fillers (such as *um* or *uh*; Krahmer & Swerts, 2005; Smith & Clark, 1993; Swerts & Krahmer, 2005) and the duration of inter-turn delays between the offset of the trivia question and the onset of the trivia response (Brennan & Williams, 1995; Krahmer & Swerts, 2005; Smith & Clark, 1993; Swerts & Krahmer, 2005). These prosodic cues also appear to be perceptually relevant to listeners tested within the FOAK paradigm (see Table 1 for a summary of the research findings).

Table 1

Prosodic Cues Associated with the Production and Perception of Uncertainty

Prosodic Cues of Uncertainty					
Production of Uncertainty (FOK Paradigm)			Perception of Uncertainty (FOAK Paradigm)		
Intonation	Filler	Inter-turn Delay	Intonation	Filler	Inter-turn Delay
Speaker uncertainty associated with rising intonation at the end of the response[a]	Speaker uncertainty associated with use of a filler prior to responding[b]	Speaker uncertainty associated with a greater latency to respond[c]	Listener perception of speaker uncertainty associated with rising intonation at the end of the response[d]	Listener perception of speaker uncertainty associated with the use of a filler prior to response[d]	Listener perception of speaker uncertainty associated with a five second latency to respond[d]

Note. [a] Brennan and Williams (1995); Krahmer and Swerts (2005); Smith and Clark (1993); Swerts and Krahmer, (2005)
[b] Krahmer and Swerts (2005); Smith and Clark (1993); Swerts and Krahmer (2005)
[c] Brennan and Williams (1995); Krahmer and Swerts (2005); Smith and Clark (1993); Swerts and Krahmer (2005)
[d] Brennan and Williams (1995).

Prosodic cues of uncertainty defined.

Intonation is generally understood as the pitch patterns of a spoken utterance or the ‘melody’ of speech (Bolinger, 1972, 1985; Denes & Milton-William, 1962; Hunter, 2004; O’Connor & Arnold, 1961). According to Pierrehumbert’s theory of English intonation (Beckman & Pierrehumbert, 1986; Pierrehumbert & Hirschberg, 1990; Pierrehumbert, 1980), intonation contours comprise three classes of pitch events: pitch accents (which mark the prominent syllables), phrase accents (which mark the boundary of intermediate phrases) and boundary tones (which mark the boundary of intonational phrases). Each pitch event is associated with either a high (H) or low (L) tone (Pierrehumbert, 1980). H and L tones refer to relatively high or low pitches on an intonation contour for a particular utterance (Pierrehumbert, 1980). Combinations of H and L tones for the three areas comprise intonation contours (Pierrehumbert & Hirschberg, 1990).

There are many different types of intonation contours in which any single utterance can be spoken, depending on the combination of pitch accents, phrase accents and boundary tones (e.g., low-rise, fall-rise, high-fall, high-rise; Pierrehumbert, 1980). However, final intonation in the production and perception of uncertainty literature was categorised by the researchers/independent transcribers as either ‘rising’ or ‘falling’ purely based on whether the

utterance ended in a relatively high boundary tone (i.e., rising intonation) or relatively low boundary tone (i.e., falling intonation), irrespective of the combination of pitch and phrase accents that preceded it (Brennan & Williams, 1995; Krahmer & Swerts, 2005; Smith & Clark, 1993; Swerts & Krahmer, 2005).[3]

Inter-turn delay refers to the duration of silence between the offset of one speaker's utterance and the onset of the other speaker's response (Roberts, Francis, & Morgan, 2006). An unfilled inter-turn delay refers to a delay that only contains silence, whereas a filled inter-turn delay refers to a delay that contains a filler (a speech interjection, such as *uh* or *um*; Goldman-Eisler, 1961).

Procedural Variations to the Full Readback Requirement

During a routine readback/hearback exchange, a sender is limited in the type of information that can be used as evidence during the grounding process. Specifically, the readback requirement that is implemented in many technical environments stipulates that a receiver is to respond to an instruction with a full readback that repeats "the entire instruction or a fully synonymous equivalent" (Cushing, 1994, p. 40). When this protocol is adhered to, there is minimal semantic variation between a sender's instruction and the receiver's readback, and therefore, a sender is confined to the variation of prosodic cues in a receiver's readback to gauge any potential concealed nonunderstanding during the grounding process.

However, there are many situations within various technical environments where procedural variations to the readback requirement will occur. As Cushing (1994) notes, "such (readback) requirements are often honoured more in the breach that in the observance" (p. 40). For example, in an emergency situation, operators may ignore or neglect to follow standard phraseology and default to the use of plain language (Alderson, 2009). Morrow et al. (1994) also found that operators were less likely to use standard terminology when resolving a miscommunication. Even in routine communication, partial readbacks can often be used in

[3] Although this is a very simplistic method of categorizing intonation contours as it does not take into account preceding pitch and phrase accents, previous research suggests that listeners generally attend to overall pitch shape rather than the phonetic detail of where the intonation contour began (Levis, 1999, 2002).

place of full readbacks. For example, Morrow, Lee, and Rodvold (1993) observed that out of the routine pilot communication they analysed, up to 13% contained partial, rather than full, readbacks. Other investigators have reported similar figures (Cardosi, 1993; Prinzo et al., 2006).

The distinction between a full and partial readback, is that unlike a full readback, the receiver does not repeat all the information in the instruction provided by the sender during a partial readback response (Barshi & Farris, 2013). When receivers respond with a partial, rather than a full readback therefore, senders do not receive any confirmation that the instruction has been accurately heard. As hearing a command accurately, as evident through a full readback, is arguably a precondition to understanding an instruction, the use of partial readbacks consequently fail to provide senders with any evidence to support the notion that common ground (of the instruction) has been achieved (Barshi & Farris, 2013; Cushing, 1994).

Procedural variations as a function of organisation type.

Although procedural variations to the readback requirement occur as a function of situational and contextual factors within an organisation (such as during emergencies), compliance to the readback protocol also tends to vary across organisations. For instance, hospitals across the United States (US) have only recently begun to implement standardised readbacks to improve patient safety, despite their use in aviation for many preceding years (Hanna et al., 2005).

Unlike in the aviation context where the observance of the readback protocol is strongly enforced, the consistent use of standardised readbacks within hospitals is progressing slowly (Prabhakar et al., 2012). In a relatively recent study, Prabhakar and colleagues (2012) explored the perceptions held by medical practitioners in a US hospital regarding the utility of the readback protocol within their healthcare setting. Differences were evident between residents, attending physicians, and nursing staff in the way that the readback protocol was

regarded as an improvement for patient safety. The lack of a common belief in the utility of standardised readbacks to improve patient safety could account for the relatively slow adherence to the readback protocol.

Semantic Cues of Nonunderstanding

As responses to procedural instructions frequently consist of a partial, rather than full, readback, it is important to consider how the semantic cues contained within a response also influence a sender's perceptions of nonunderstanding on the part of the receiver. Although there is the common assumption that the receipt of a full readback from a receiver can be considered evidence of understanding (Cushing, 1994; International Civil Aviation Organization, 2007; Wickens, 2007), it is important to test this assumption empirically as the readback procedure has been criticised as being both superfluous and time-consuming (Schneider et al., 2004).

In addition to investigating whether the semantic evidence contained within a full readback (in comparison to a partial readback) provides a sender with additional evidence of understanding, it is also important to investigate whether the prosodic cues that are potentially useful to operators to gauge a receiver's level of understanding, can also be generalised across non-routine responses to instructions (i.e., during partial readback responses). Although Brennan and Williams' (1995) investigated uncertainty, rather than nonunderstanding associated with communicative exchanges, their data suggests that the utility of the prosodic cues of intonation, filler and inter-turn delay as an indication of a speaker's state of understanding, potentially exists, even when there is no supporting semantic evidence.

Brennan and Williams (1995) informed participants (i.e., listeners) that they would be hearing verbal responses to trivia questions and would be required to estimate how likely it was that the speaker knew the correct answer to the question (i.e., the speaker's level of certainty). They presented listeners with 32 naturalistic responses to trivia questions, where each trivia response produced by a speaker was preceded by a generic question (e.g., "*What is*

the answer to question one?" p. 389), rather than the actual trivia question that elicited each response. Therefore, each listener was generating ratings of uncertainty on the part of the speaker based solely on prosodic cues, rather than acquiring any semantic evidence of certainty (that may have been available if the original trivia question had been included in the communicative exchanges). This suggests that, in the absence of any supporting semantic evidence of the certainty of a speaker, the prosodic cues of intonation, filler and inter-turn delay constitute useful indicators of uncertainty. However, the utility of semantic cues has yet to be tested empirically within the context of the readback/hearback protocol during perceptions of nonunderstandings.

Sender-Initiated Repairs of Perceived Concealed Nonunderstandings

Concealed nonunderstandings represent a real and potential risk that can threaten the effectiveness of the readback/hearback protocol within high-risk environments. Arguably, the purpose of concealed nonunderstandings is to deceive the sender to the effect that the instruction has been understood as intended. In such circumstances, the sender remains as the communicative partner in charge of detecting and resolving the receiver's nonunderstanding to ensure mutual understanding of the instruction, or common ground, is achieved.

When a receiver responds to an instruction with a full readback, a sender can potentially monitor the variation in the prosodic cues to gauge the extent to which the receiver has understood the instruction (as such cues have been found to be useful during perceptions of uncertainty). As procedural variations occur frequently, it is also important to establish whether such prosodic cues can be used during perceptions of nonunderstanding when an operational instruction is followed by a partial readback. Finally, it is of interest to determine whether the semantic content available in a full readback, compared to a partial readback, is used by a sender to indicate that the receiver has understood the instruction.

Introduction: Section 2

Situation Assessment in Operational Environments

Medicine, aviation and electrical rail control are high-risk technical environments in which operators are required to respond under a unique and sometimes demanding set of circumstances. These environments are characterized by ill-structured problems, uncertain and dynamic information, time pressure, high stakes and ill-defined or competing goals (Orasanu & Connolly, 1993, p. 7). Within such constraints, the safe and efficient operation of these environments relies upon operators' capacity to respond in an efficient manner.

Situation assessment is considered a precursory stage to the initiation of a response to a high risk, high-consequence situation (Kaempf, Klein, Thordsen, Marvin, & Wolf, 1996; Wellens, 1993). Situation assessment is a cognitive skill whereby operators attempt to understand the nature of a situation by attending to, and interpreting, the information available (Klein & Calderwood, 1991; Klein, 1989). It is distinct from other, related constructs, such as situation awareness, insofar as it represents the initial stage in the decision making process rather than encompassing the projection of the status of the system state in the future (Endsley, 1995; 1997). A fire fighter, for example, may consider such environmental cues as the colour of the smoke and the force with which it is escaping a building, to draw the preliminary inference that toxic chemicals may be burning (Klein, 2003). Similarly, a medical practitioner may observe a patient exhibiting an elevated temperature, and deduce that an infection is most likely present (Simon, 1992).

Theoretical Models of Situation Assessment

There are a number of different theoretical models that purport to describe the cognitive mechanisms that are expected to underlie the process of situation assessment, including the Recognition-Primed Decision (RPD) model (Klein, 1993, 1997), the Lens Model (Brunswik, 1955) and the Adaptive Control of Thought - Rational (ACT-R*) framework (Anderson, 1982).

The Recognition-Primed Decision model.

The RPD model (Klein, 1993, 1997) describes how operators in highly technical, naturalistic settings rely on their experience to respond effectively to situations in a timely manner. This model underscores the importance of the process of situation assessment to be able to generate a reasonable and timely response (Wellens, 1993).

Cues and pattern recognition.

The RPD model (Klein, 1993, 1997) suggests that the environmental *cues* underpin the process of situation assessment. Conceptually, cues[4] are thought to be associations in memory between feature/s of the environment and an object/event (Ericsson & Kintsch, 1995; Wiggins, 2006, 2012). Cue associations are formed when a feature in the environment (such as rising intonation at the end of a readback response) is repeatedly paired with an event/object (such as a concealed nonunderstanding during a readback response), such that the recognition of a feature in an environment holds meaning for an operator. These cue associations can be auditory, olfactory, tactile or visual in nature (Wiggins, 2006).

As a result of extensive experience within a domain, an operator is expected to have developed a reservoir of patterns of cues that pertain to different events (Coderre, Mandin, Harasym, & Fick, 2003; Croskerry, 2009; Ericsson & Kintsch, 1995; Klein, 1993, 2003). These patterns of cues are presumed to be held in Long-Term Memory (LTM), which is a relatively permanent and unlimited store of information acquired on the basis of past experience (Ericsson & Kintsch, 1995). These recognised patterns of cues are thought to facilitate situation assessment through a relatively rapid process referred to as *pattern recognition* (Kaempf et al., 1996; Kahneman & Klein, 2009; Klein, 1989, 1997, 1999).

Pattern recognition describes the process in which an operator will recognise a new situation as familiar, by 'matching' the pattern of cues observed in the current operational

[4] Within the current thesis, the term *feature* will be used to refer to a piece of information that has no association with an event/object. When a feature becomes coupled with an event or object in memory, it then becomes a *cue*. For example, in Brennan and Williams' (1995) investigation, intonation, inter-turn delay and filler were prosodic features evident in a speaker's response. When it was evident that they were associated with perceptions of uncertainty ('event'), they became prosodic cues of uncertainty as it was established they held meaning for operators. It is important to note that in Section 1 of the Introduction, the term cue was used exclusively, even in circumstances where feature was a more accurate description. This was intentional as the theoretical distinction between 'feature' and 'cue' was considered more appropriate to discuss in Introduction: Section 2.

environment it to a pattern encountered in the past that is stored as a template in LTM (Coderre et al., 2003; Kaempf et al., 1996; Klein, Calderwood, & Clinton-Cirocco, 1986; Klein, 2003; Schmidt & Boshuizen, 1993; Simon, 1992; Wickens & Flach, 1988; Wickens & Hollands, 2000). Pattern matching allows an operator to rapidly classify a situation as 'typical' (if an exact match in LTM exists) or 'novel' (if an exact match in LTM is unavailable; Lipshitz, 1993). Recognition of a typical situation based on a pattern of cues is expected to activate an 'action script' in memory which, amongst other things, specifies typical ways to respond to the situation (Klein, 1989, 2003; Lipshitz, 1993). The RPD model, therefore, describes how experienced operators can make effective decisions through a *recognition-primed* process of situation assessment (Klein, 2003, 2008).

Within some situations, the presence of a single, highly diagnostic cue (rather than a *pattern* of cues) can be used reliably by operators to assess the system state (Schriver, Morrow, Wickens, & Talleur, 2008). For example, Ebbesen and Konečni (1975) examined the types of information (cues) that influenced judges' decisions in bail hearings, including the gender and age of the defendant, the severity of the crime and the strength of community ties, and noted that judges based their decisions almost exclusively on the district attorneys' recommendations. In turn, the district and defence attorneys' bail recommendations were also based on one cue: the severity of the crime.

In many instances, however, the presence of a single cue in the environment will not be sufficient to indicate the system state, since it is associated with several other object/events in LTM (Klein, 1989; Schriver et al., 2008). As a result, operators often rely on a pattern of cues to generate recognition and a subsequent assessment of a situation (Klein, 1993; Wickens & Hollands, 2000). In their investigation into the utility of prosodic cues to signal uncertainty, for example, Brennan and Williams (1995) found that the presence of one cue (rising intonation) at the end of speaker responses, led listeners to detect a greater level of uncertainty on the part of speakers, compared to the absence of a cue (i.e., falling intonation). However,

listeners perceived an even *greater* level of uncertainty on the part of speakers when two prosodic cues were present in speaker responses (i.e., rising intonation *and* a longer inter-turn delay), and even more so, when three prosodic cues were evident in speaker responses (i.e., rising intonation, longer inter-turn delay *and* a filler[5]). This suggests that while listeners can use one prosodic cue to assess uncertainty on the part of speakers, the use of two or more cues led to greater confidence of speaker uncertainty.

Pattern recognition and expert performance.

According to Kahneman (2003), an operator may engage in one of two processes when encountering a new situation that requires assessment. *System 1* reasoning describes judgements that are rapid, automatic and intuitive mental-shortcuts, such as pattern recognition (Coderre et al., 2003; Croskerry, 2009; Evans, 2008; Kahneman & Frederick, 2005; Kahneman & Klein, 2009; Kahneman, 2003). System 1 processes are 'fast and frugal' and rely heavily on the experience of operators to recognise as familiar, patterns of information in the environment, and intervene in a manner that has proven successful in similar situations (Gigerenzer & Todd, 1999; Klein, 1989, 1999). The engagement of System 1 processes facilitates a rapid assessment of a situation and ensures that cognitive resources are available for other necessary tasks vital to efficient performance in uncertain and dynamic environments. Those operators who do not possess established cue associations in memory to appropriately interpret environmental information, are unable to engage in pattern recognition (Coderre et al., 2003).

As an alternative, operators can assess a situation through the interpretation of environmental stimuli using *System 2* processes (Kahneman, 2003). These processes use deductive reasoning to assess a situation and are relatively slower, systematic and effortful, requiring a greater expenditure of time and cognitive resources (Croskerry, 2009; Kahneman & Klein, 2009; Stolper et al., 2011). Due to the time constraints inherent in high-risk

[5] Although the mean ratings of uncertainty differed depending on the presence of one, two or three prosodic cues contained in a speaker's utterance, Brennan and Williams (1995) did not report whether these differences were *significantly* different from each other.

environments, experienced operators depend heavily System 1 processes during situation assessment (Coderre et al., 2003; Kaempf et al., 1996; Klein, Calderwood, & Clinton-Cirocco, 1988; Klein & Klinger, 1991; Orasanu & Connolly, 1993). The use of System 2 reasoning is expected to be used by operators as a secondary process that is activated when a clear 'match' is unavailable during the initial pattern recognition process (Klein & Klinger, 1991; Klein, 1989; Young, Smith, Guerlain, & Nolley, 2007). In the absence of a match, operators must engage System 2 processes during situation assessment, as their lack of experience renders them unable to access a repertoire of patterns in memory that might facilitate intuitive judgements (Croskerry, 2009; Lipshitz, Klein, Orasanu, & Salas, 2001).

Origins of the Recognition-Prime Decision model.

The RPD model was developed initially by Klein's (1989) descriptive inquiry into experienced fire commanders' decisions under time pressure and uncertainty. Through observations and interviews with these operators, Klein (1993) noted that the majority of commanders were not generating and evaluating multiple options when assessing a situation. Due to the time constraints inherent the environment, the fire commanders would, instead, rely on their experience to classify the situation as typical and 'recognise' a course of action that had proven successful in past situations (Kaempf & Klein, 1994; Klein, 1993).

The RPD model (Klein & Klinger, 1991; Klein, 1993; Figure 2) describes how recognition-primed decision-making can occur in both simple and complex forms, which differ in the extent to which System 1 and System 2 processes are engaged (Kaempf et al., 1996). Both forms of the model propose that operational environments contain different types of information, including plausible goals, cues to monitor, expectancies about the situation, and typical reactions.

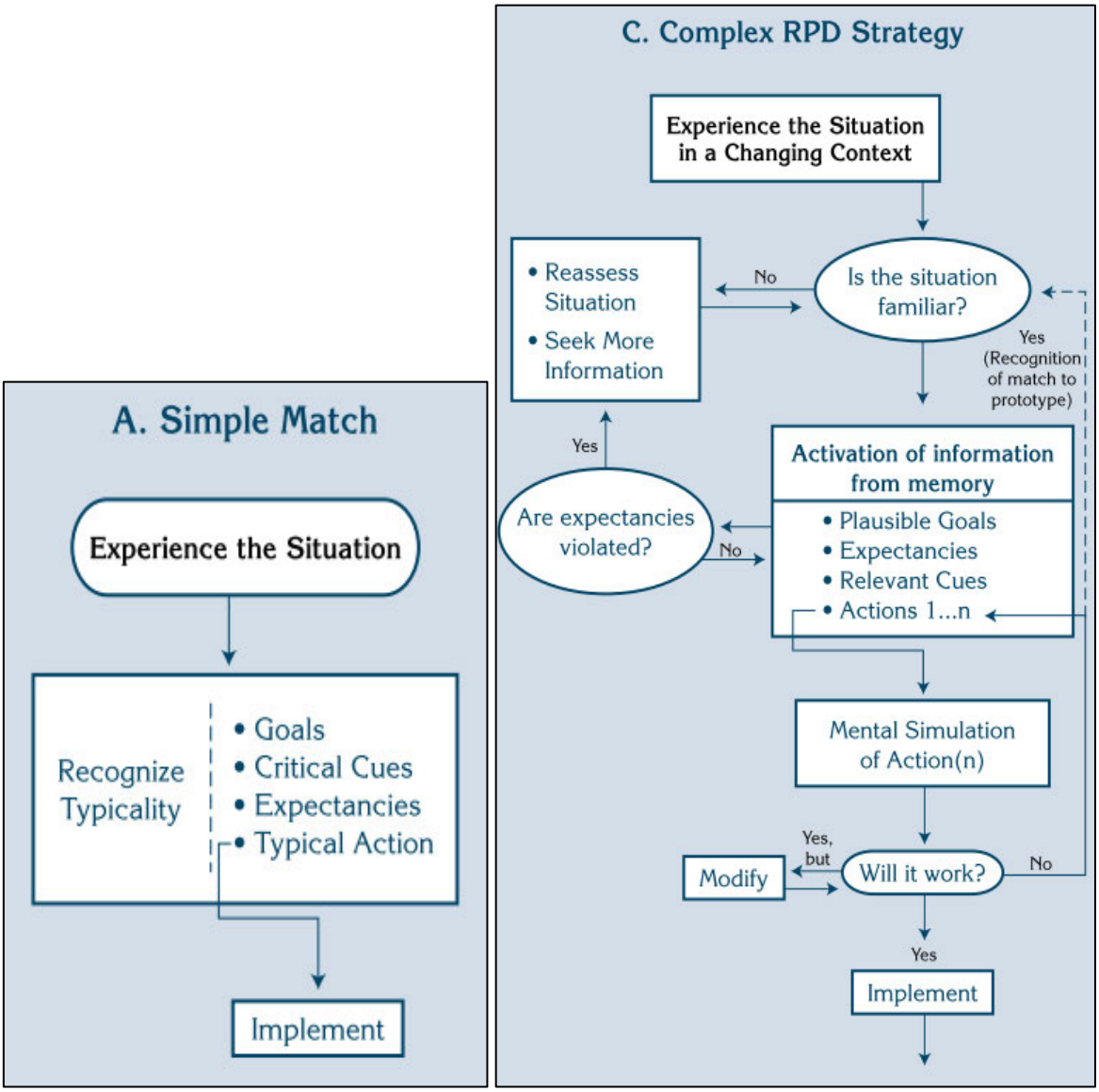


Figure 2. An abbreviated version of the RPD model illustrating both simple (left) and complex (right) recognition-primed decision strategies (Klein & Klinger, 1991; Klein, 1993).

In the simplest version of the RPD model (see 'Simple Match'; Figure 2), operators will recognise a situation as typical through pattern recognition (Kaempf et al., 1996; Klein, 1997). For experience operators, this recognition of a pattern of cues will activate an action script, which will include routines for responding (Klein, 2003). Within the domain of electricity transmission control, for example, an experienced system operator may make an urgent call to a field operator and direct him to execute a difficult sequence of steps to curtail a critical situation unfolding in the field. From past experiences with critical incidences within this domain, the system operator may recognise that a field operator who has relative inexperience

(cue) and unfamiliarity with organisational procedures (cue), hesitates before responding with a readback (cue), and responds with a partial (cue) and inaccurate (cue) readback of the operational steps, is usually incapable of responding in an appropriate and efficient manner to a rapidly unfolding situation. As many of these cues are present in the confronting situation, the field operator will recognise the situation as 'typical', and this recognition will then activate an action script that will guide the operator to respond in a way that was successful in the past (in this case, it might involve calling up another, more experienced operator to deal with the problem). Consistent with System 1 operations, a Simple Match strategy, through the recognition of patterns of cues, is largely a non-conscious process that does not require any extensive deliberation of the observed cues nor courses of action (Klein, 1997).

In the more complex version of the RPD (see 'Complex RPD Strategy'; Figure 2), the process of pattern recognition will not necessary produce a clear match for the current situation in LTM, and therefore operators are not primed to an obvious course of action. During these circumstances, operators will engage in System 2 processes and mentally simulate a course of action to determine its potential success within a particular situation. The course of action may be considered appropriate (and consequently implemented), it may expose flaws (requiring modification), or it may be considered inadequate and rejected in favour of the next best solution. The complex RPD strategy highlights how even in the more ambiguous situations, operators do not have to apply heavy analytical strategies to generate an appropriate response. Rather, they engage a simpler strategy referred to as *satisficing*, which involves selecting the first workable alternative, rather than considering all of the alternatives available to select the optimal response (Klein, 1999; Simon, 1978).

In summary, the RPD model describes how decisions made in naturalistic settings can involve purely System 1 processes when a situation is encountered, recognised as typical through pattern matching, and the first course of action is implemented (Klein, 2008). In more complex situations where a clear match to the situation is unavailable, a combination of

System 1 processes and the more analytical, conscious and deliberate processes of System 2 are needed to arrive at a response (Klein, 2008).

The RPD model has been applied to a variety of different contexts, including military control, commercial aviation, offshore drilling and fire ground commanders, where recognition-primed judgments constitute the most common situation assessment strategy within these environments (Lipshitz et al., 2001). Kaempf and colleagues (1996), in particular, investigated how experienced naval officers made decisions within complex, uncertain and time-pressured conditions, and concluded that: (1) Officers relied predominantly on the process of pattern matching through cue utilisation to classify a situation as typical, and that; (2) The recognition of the situation as typical, enabled the officers to recognise an appropriate action based on factual or procedural knowledge, rather than comparing multiple options.

The RPD model, therefore, describes how operators in naturalistic settings use their experience in the form of a repertoire of patterns in memory to recognise their situation as familiar, thereby enabling rapid responses (System 1 processing). In more complex situations, the RPD model describes how operators use a combination of recognition-based processes as well as more deliberative and analytic strategies to evaluate the validity of response options in a serial manner (System 2 processing; Klein, 2008).

The Lens Model.

Using mechanisms similar to those that underscore RPD, Brunswik (1955) proposed a ‘lens’ analogy to suggest that the process of cognitive judgement, such as situation assessment, takes place through a process of cue utilisation. According to Brunswik’s (1955) Lens Model (Figure 3), cue utilisation is a process in which an operator will utilise many different environmental cues to make a cognitive judgment about the true state of a situation.

On the left hand side of the lens (Figure 3), an environment will contain several cues that are differentially associated with an outcome/situation (or *true state*; Newell, Lagnado, & Shanks, 2007). The precise relationship between individual cues and the true state of a system,

however, is not always evident. The right hand side of the lens depicts how an operator, or *judge*, will attend to particular cues present in the environment to form a judgment.

As depicted in Figure 3, the relative importance (weights) of the cues may vary between those predicting the true state and those used to judge the state, suggesting that a judge will not always utilise the most predictive cues when forming a judgement (Newell et al., 2007). The manner by which a judge will integrate and weigh the importance of cues to form a judgement is dependent on the associations that have been established in memory through experience between features and events (Balzer, Doherty, & O'Connor, 1989; Bisantz & Pritchett, 2003). An experienced Air Traffic Controller, for example, may take into account a pilot's non-native English speaking background, use of non-standard phraseology and lack of adherence to the full readback requirement during a readback response to judge the extent to which a pilot has understood an instruction. Although these cues may be the most predictive indicators of a pilot's state of nonunderstanding as reflected in empirical data, a less experienced pilot may rely on other, less predictive cues, such as the pilot's gender or tempo of the readback, to ascertain a state of nonunderstanding due to a relative lack of exposure to such situations. The overarching line connecting the criterion and the judgement represents the accuracy of the judge in perceiving the situation/outcome (Newell et al., 2007).

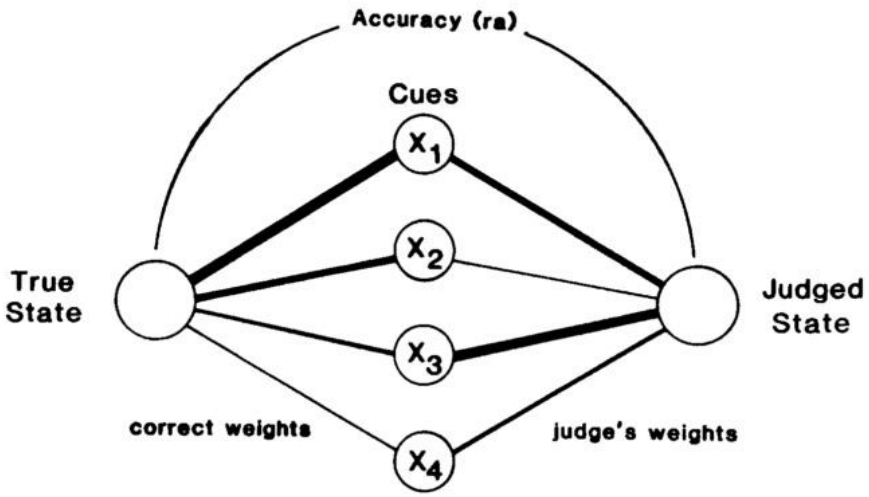


Figure 3. A schematic illustration of Brunswik's Lens Model that conceptualises the process of intuitive judgements, such as situation assessment (Wigton, Hoellerich, & Patil, 1986).

Using the Lens Model as a basic framework, a number of researchers have investigated how operators use probabilistic cues from a range of uncertain environments to form judgments (Karelaia & Hogarth, 2008; Wigton, 1988). There has been general support for the principles that underlie the Lens Model and, in particular, the proposition that operators extract and utilise cues during situation assessment as a function of the associations formed in memory through experience between the cues and the true state (Balzer et al., 1989; Bisantz & Pritchett, 2003; Karelaia & Hogarth, 2008; Wigton, Patil, & Hoellerich, 1986). The Lens Model (Brunswik, 1955) therefore, describes situation assessment as a process whereby operators rely on a range of idiosyncratic cues in the environment to derive meaning from an uncertain situation.

Adaptive Control of Thought-Rational model.

Using heuristics as the basic mechanism of association, Anderson's (1982) Adaptive Control of Thought - Rational (ACT-R*) framework also refers to experience as a key component in formulating rapid and reasonable assessments of uncertain situations. Although primarily a framework to describe cognitive skill acquisition, rather than the process of situation assessment per se, the ACT-R* model proposes the involvement of two stages in the development of a cognitive skill: the *declarative* stage and the *procedural* stage. The declarative stage involves the acquisition of factual knowledge about the skill domain. During the *procedural* stage, this knowledge is transformed into procedural 'rules' or productions in memory that are necessary to undertake skilled performance (Anderson, 1982).

Productions are 'IF-THEN' rules or condition-action statements that are essentially knowledge representations that guide behaviour within specific situations (Anderson, 1987). Production rules specify that 'IF' a particular pattern of information is encountered or satisfied, 'THEN' a particular response can be executed (Anderson, 1993). When required to transform a singular word to a plural, a particular production may be activated in memory: 'IF' the goal is to generate the plural of a noun, 'THEN' say "noun + s" (Anderson, 1982, p. 373).

The ACT*R model purports that operators develop and strengthen these productions through many successful pairings of certain patterns of information (cues) with appropriate responses (actions; Anderson, 1993). From past experiences with detecting concealed nonunderstandings on the basis of a filler and rising intonation in a readback response, a radio operator may begin to clarify critical elements of the instruction to ensure that the receiver has understood the instruction. If this rule is successful (i.e., receivers may concede that they did not actually understand the instruction during those instances), this will form a production rule in the form of a simple 'IF-THEN' statement (such as, 'IF' the receiver uses a filler and rising intonation at the end of the readback response, it may indicate a nonunderstanding, 'THEN' clarify instruction for receiver). These conditional statements are expected to eventually become a heuristic specifying how to respond in such situations (Anderson, 1993).

The ACT*R model is akin to the predictions of the RPD model and Lens Model, such that the recognition of a pattern of cues in the environment and the consequent assessment of a situation falls within the 'IF' condition, and the response selection falls within the 'THEN' condition (Greitzer, Podmore, Robinson, & Ey, 2010; Wiggins, 2006). Therefore, the ACT*R model describes how the recognition of an action is primed once the salient cue or pattern of cues is recognised in the environment, without the need to engage in conscious deliberation.

Consistent with the predictions of the ACT*R model, inexperienced and experience operators have distinctly different, nuanced productions in memory, which is evident through their differential use of heuristics in problem solving. Inexperienced operators generally apply weak heuristics that are founded on domain-general knowledge (Chi, Glaser, & Rees, 1982). In comparison, experienced operators are able to use their superior domain-specific knowledge base to apply more appropriate heuristics to achieve efficient performance (Chi et al., 1982). Lau and Redlawsk (2001), for example, found that, while all voters utilised cognitive heuristics during a mock presidential election campaign, politically sophisticated voters were more likely to use the Endorsement and Ideological Schema heuristics that

required an organised and coherent understanding of politics to be used effectively (Sniderman, Brody, & Tetlock, 1991). Therefore, the ACT*R model predicts that experienced operators are able to rapidly assess and respond to problems through the use of highly developed and refined productions in memory.

Empirical Support for Recognition-Primed Situation Assessment

Although distinct, the RPD model (Klein, 1993, 1997), the Lens Model (Brunswik, 1955) and the ACT*R model (Anderson, 1982) all posit that pattern recognition, through the utilisation of associations between features and events, is the cognitive mechanism that underlies situation assessment. Empirical evidence to support this recognition-driven approach to situation assessment has been explored extensively using the expert-novice paradigm. Using a cross-sectional methodology, this paradigm compares the performance of experts and novices to identify the cognitive processes that underlie superior situation assessment (Abernethy, 1990; Chase & Simon, 1973; Coderre et al., 2003; de Groot, 1965; Eccles, Walsh, & Ingledew, 2002; Gray, 2004; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Kee et al., 2003; Müller, Abernethy, & Farrow, 2006; Myles-Worsley, Johnston, & Simons, 1988; Northcraft & Neale, 1987; Simnett, 1996; Stokes, Kemper, & Kite, 1997; Stokes, Kemper, & Marsh, 1992; Wiggins & Henley, 1997; Wiggins & O'Hare, 1995).

One of the most robust findings arising from this paradigm is how experienced operators can be distinguished from their less skilled counterparts through their more refined domain-specific knowledge representations within LTM that facilitate the process of situation assessment (Calderwood, Klein, & Crandall, 1988; Coderre et al., 2003; de Groot, 1965; Ettenson, Shanteau, & Krogstad, 1987; Jarodzka et al., 2010; Müller et al., 2006; North, Williams, Hodges, Ward, & Ericsson, 2009; Slovic, Rorer, & Hoffman, 1971; Wiggins, Azar, Hawken, Loveday, & Newman, 2014; Williams, Ward, Knowles, & Smeeton, 2002; Williams, Hodges, North, & Barton, 2006; Young et al., 2007). This reservoir of recognised patterns is presumably a product of extensive experience or deliberate practice within a domain

(typically a decade; Ericsson, Krampe, & Tesch-Römer, 1993). Less experienced operators, by comparison, do not possess such refined associations in memory and are forced to integrate environmental features into working memory to produce a solution (consistent with a System 2 solution; Stokes, Kemper, & Marsh, 1992; Stokes et al., 1997).

The application of cues during situation assessment is reasonably well-established within a variety of domains, including aviation (Schriver et al., 2008; Stokes et al., 1992; Wiggins & O'Hare, 1995, 2003a,b), chess (Chase & Simon, 1973; de Groot, 1965), sport (Abernethy, 1990; Gray, 2004; North et al., 2009) and medicine (Hoffman, Aitken, & Duffield, 2009; Loveday, Wiggins, Searle, Festa, & Schell, 2013; Myles-Worsley et al., 1988; Wigton, Hoellerich, et al., 1986). In seminal research within the area of skilled recognition, Chase and Simon (1973) showed how chess masters were more capable of reconstructing chess pieces on a clear chessboard from memory, compared to less experienced players. This effect did not hold when players were asked to reconstruct a board containing randomly placed pieces. The chess masters' superior perceptual skill in reconstructing meaningful (but not meaningless) configurations was based on their capacity to access a repertoire of recognisable patterns refined by their extensive experience.

Skilled recognition has also been demonstrated in different sports, and particularly, in those sports that require responses that are both rapid and accurate. For example, skilled tennis players exhibit superior anticipatory performance compared to less skilled players (Williams et al., 2002). Experts' anticipatory skill was based on the extraction of more refined and subtle cues, such as the degree of rotation of the torso or the relative position of the shoulders to hips, to anticipate an opponent's swing. Less skilled players were more likely to attend to more obvious cues, such as the racquet and ball regions.

Within the context of software engineering, the utilisation of cues during situation assessment has also been associated with broad indicators of workplace expertise. In their investigation, Loveday, Wiggins, and Searle (2013) used an online situational judgement test,

EXPERTise 1.0 (Wiggins, Harris, Loveday, & O'Hare, 2010) to distinguish between two 'typologies' of practitioners based on their relative level of visual cue utilisation during domain-relevant activities. Those operators who possessed relatively greater levels of visual cue utilisation demonstrated superior error management (assessed over a three month period), and were also more likely to be nominated as 'expert developers' by their peers. Overall, empirical support for the significance of cues in facilitating accurate situation assessment is evident across a variety of domains.

Limitation of the Construct of Cue Utilisation Underlying Expert Performance

The application of the expert-novice paradigm has provided reasonably consistent support for the proposition that experienced operators can be reliably discriminated from their less experienced counterparts by their greater ability to recognise and utilise task-relevant cues during situation assessment (e.g., Calderwood et al., 1988; Coderre et al., 2003; de Groot, 1965; Ettenson et al., 1987; Hoffman et al., 2009; Jarodzka et al., 2010; Müller et al., 2006; North et al., 2009; Shim, Carlton, Chow, & Chae, 2005; Williams et al., 2002). However, a major limitation of this literature is that much of the work in the assessments of cue utilisation as the foundation for situation assessment has involved the application of *visual* cues, to the exclusion of cues from other modalities. In some cases, this focus on visual cues has involved controlling for, and thereby excluding other cues, including auditory cues, despite the fact that many of the environments within which assessments have been undertaken incorporate visual information, and information from other modalities. For example, Scialfa et al., (2011) compared the ability of novice and experienced drivers to assess hazards from filmed traffic scenes, and confined participants to judgements based on visual cues in the absence of the auditory, tactile and olfactory cues that would normally be present.

The focus on visual cues to exclusion of other modalities is significant since, in many high-risk environments, operators are expected to integrate cues from more than one modality when assessing a situation. Skilled driving for example, requires drivers to attend to both an

array of visual cues (such as recognising brake lights or monitoring a speedometer), and cues derived from other modalities (such as the sound of a car horn or the proximity warning). By excluding a modality, the potential for interactions between modalities is lost.

If superior performance during situation assessment is determined by cue utilisation (irrespective of modality), rather than simply as a by-product of visual processing, then the same effects that occur in response to visual cues should be apparent for assessments made on the basis of auditory, olfactory or tactile cues. This would provide greater support for the process of cue utilisation underpinning expert performance during situation assessment and help inform training within technical environments.

Introduction: Section 3

Research Aims and Design

In the context of multi-crew operations, clear and efficient communication is one of the most important processes associated with maintaining the safety and security of technical environments. One of the unique aspects of communication within such environments is the frequent reliance on radio communication as a means of delivering and confirming information between operators. Despite the implementation of the readback/hearback protocol, nonunderstandings still have the potential to compromise the reliability of these communicative exchanges.

An incident where a receiver conceals the fact that an instruction has not been understood, represents an opportunity for a sender to recognise and repair the concealed nonunderstanding to ensure common ground and accurate communication. Within time-pressured and complex environments, assessments of nonunderstanding are expected to be facilitated through the application of cues. Due to the context in which these communicative exchanges take place (i.e., over the radio), only cues of an auditory (prosodic and semantic) nature can be attended to when making assessments of a speaker's state of understanding.

During perceptions of uncertainty, listeners attend to the prosodic cues of intonation, filler and inter-turn delay (Brennan & Williams, 1995). Specifically, listeners detect a greater level of uncertainty on the part of speakers when responses end in rising intonation, are preceded by a filler, and are produced after a relatively long delay (Brennan & Williams, 1995). As the receipt of uncertain information (Woods, 1988) arguably precipitates the occurrence of nonunderstandings, it is likely that those prosodic cues associated with perceptions of uncertainty are also associated with perceptions of nonunderstanding.

In addition to these prosodic cues, it could be argued that the frequent procedural deviations in the readback procedure observed within many technical environments, also provides senders (i.e., listeners) with the opportunity to monitor variations in semantic cues of the readback response. Specifically, a full readback response arguably provides senders with

greater evidence that an instruction has been understood, in comparison to a partial readback response. The utility of semantic cues during perceptions of nonunderstanding however, has yet to be investigated.

Within high-risk technical settings, there has been consistent support for the notion that experienced operators can be reliably discriminated from their less experienced counterparts through their relatively greater ability to recognise and utilise task-relevant cues when assessing complex and ambiguous situations (e.g., Calderwood et al., 1988; Coderre et al., 2003; de Groot, 1965; Ettenson et al., 1987; Hoffman et al., 2009; Jarodzka et al., 2010; Loveday et al., 2013; Müller et al., 2006; North et al., 2009; Shim et al., 2005; Williams et al., 2002). It could be argued that those operators who can be distinguished by a greater capacity of visual cue utilisation within operational contexts, would also be more capable in responding to the prosodic (intonation, inter-turn delay and filler) and semantic (partial and full readbacks) cues that are potential indicators of nonunderstandings. However, the specific prosodic and semantic features that are utilised by operators to assess instances of nonunderstandings, and whether skilled operators have a greater capacity to utilise relevant auditory cues during such assessments, have yet to be established empirically.

The overall aim of this body of research was to determine whether operators within technical domains, use particular prosodic and semantic features contained within an auditory readback response to assess the extent to which a receiver had understood an instruction. An allied aim was to establish whether the use of such prosodic and semantic cues during assessments of nonunderstandings were associated with operators' relative levels of visual cue utilisation.

Specific Research Questions

This thesis presents four papers that examine the utility of particular prosodic and semantic features during perceptions of nonunderstanding and whether the utilisation of these cues is associated with visual cue utilisation. Together, the papers address a series of research questions that are summarised in Figure 4.

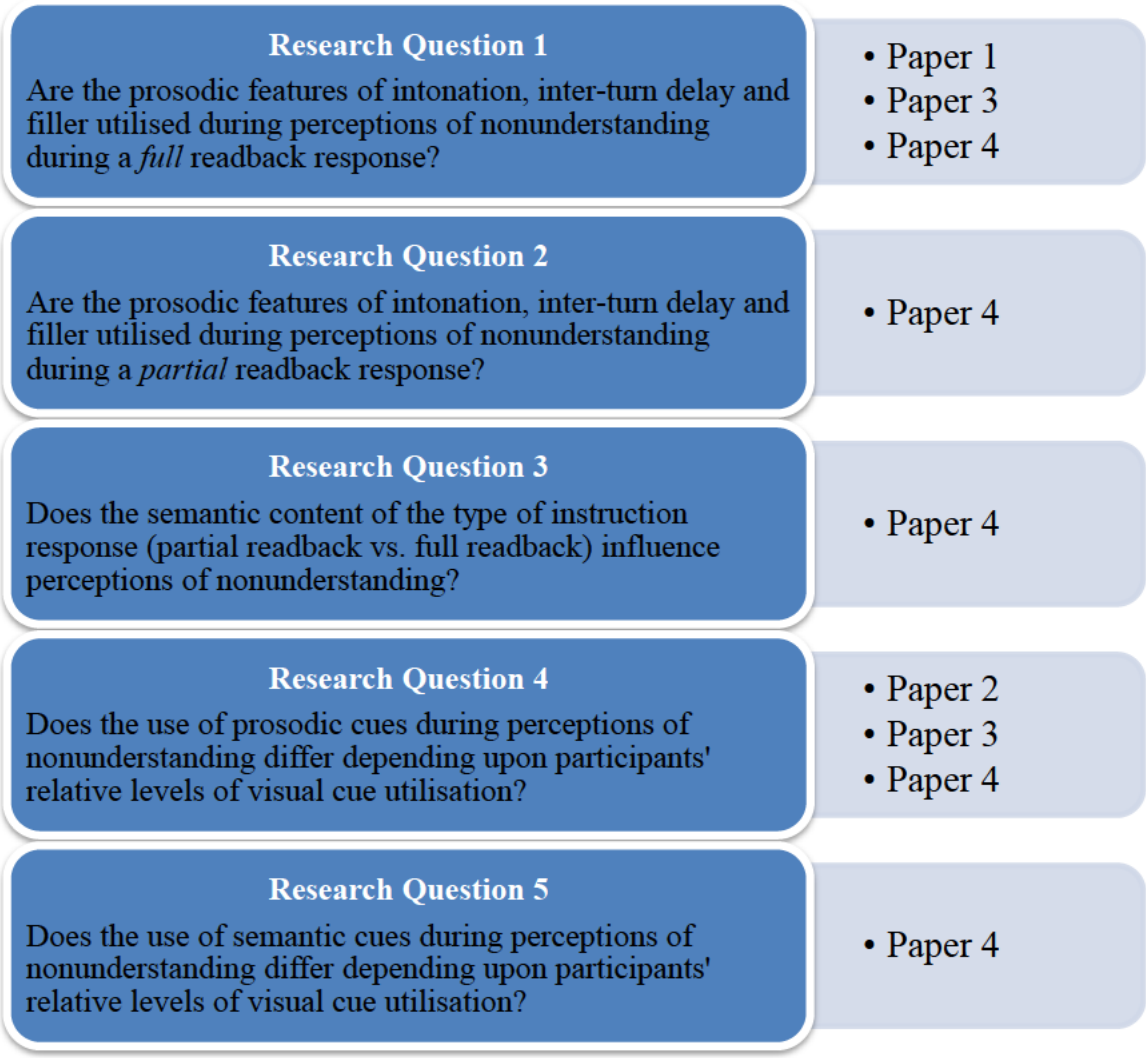


Figure 4. Diagrammatic representation of research questions and associated papers.

Within the context of electricity transmission control, Paper 1 investigated whether the prosodic cues that have been found useful during perceptions of uncertainty during trivia games (Brennan & Williams, 1995), are also utilised by listeners during perceptions of nonunderstanding on the part of speakers during full readback responses during radio communication. Using the same context as Paper 1, the study reported in Paper 2 assessed

whether the utilisation of prosodic cues, including intonation, inter-turn delay and filler, was associated with visual cue utilisation. Conducted within the context of electrical rail control, Paper 3 examined whether intonation, inter-turn delay and filler also had utility during assessments of nonunderstanding in a domain distinct from that used in Papers 1 and 2. Using a more ecologically valid measurement of prosodic cue utilisation, Paper 3 also replicated Paper 2 to examine whether the use of these prosodic cues was associated with visual cue utilisation. Finally, the study described in Paper 4 investigated the utility of intonation, inter-turn delay and filler as prosodic cues during perceptions of nonunderstanding in both *full* and *partial* readback responses (semantic cues) within the domain of hydroelectric power generation. Paper 4 also investigated whether the diagnostic value of both the prosodic and semantic cues was associated with operators' visual cue utilisation performance.

Measuring Cue Utilisation

Prosodic cue utilisation.

The main objective of this thesis was to address a research question that has practical implications for many high-risk, high-consequence environments: Are there specific auditory (prosodic and semantic) cues that can be used by a sender (i.e., the listener of a readback response) to gauge the extent to which a receiver (i.e., the speaker of a readback response) has understood an instruction within the context of a readback/hearback communicative exchange?

To investigate this research question, it was necessary to focus on the prosodic and semantic cues that are used by listeners to judge nonunderstandings on the part of the speaker, rather than those cues produced by speakers when they are in a state of nonunderstanding. According to theoretical models of situation assessment, including the RPD model (Klein, 1993, 1997), the Lens Model (Brunswik, 1955), and the ACT*R model (Anderson, 1982), such auditory cues will only be rated by listeners as indicative of nonunderstandings if feature-event cue associations reside in memory between auditory features (e.g., intonation)

and a miscommunication event (e.g., nonunderstanding; Wiggins, 2006).

Nevertheless, the research outcomes from Brennan and Williams (1995) and Keysar and Henly (2002) suggest that there is a strong relationship between actual levels of uncertainty and understanding as reported by a speaker, and perceptions of uncertainty and understanding on the part of the listener. Providing support for theoretical models of situation assessment, this suggests that, through experience, operators have developed skills necessary to identify key features of utterances that enable the accurate interpretation of cognitive function (such as a speaker's level of uncertainty or nonunderstanding). Therefore, all four papers focus on auditory cues associated with *perceptions* (or assessments) of nonunderstanding (i.e., the auditory cues that influence a listener's perception of a speaker's understanding of an instruction), rather than the auditory cues associated with the *production* of nonunderstanding (i.e., the auditory cues produced by speakers during a state of nonunderstanding).

The studies described in all four papers utilised the same framework for measuring perceptions of nonunderstanding within a readback/hearback communicative context. Each cohort of participants was presented with a series of scripted audio-only communicative exchanges emulating the readback/hearback protocol. Each exchange consisted of two elements: Firstly, participants heard a 'sender' deliver an instruction to a 'receiver'. Participants then heard the receiver repeat the instruction back to the sender (i.e., deliver the readback). The readback portion of the exchange was systematically manipulated to vary the prosodic and semantic features that were expected to influence perceptions of nonunderstanding.

After listening to each communicative exchange, participants were asked to rate the extent to which they perceived that the receiver had understood the instruction, based on the way that the instruction had been read back. Ratings were recorded on a 7-point Likert scale from 1 (*did not understand at all*) to 7 (*understood completely*) to maintain consistency with Brennan and Williams' (1995) measurement of perceptions of uncertainty. These ratings,

therefore, constituted participants' perceptions of nonunderstanding on the part of the receiver. In summation, the current thesis conceptualised nonunderstandings as a gradual phenomenon, which can vary from a partial to a complete lack of understanding (Hirst et al., 1994).

Visual cue utilisation.

It was necessary to measure the visual cue utilisation performance of practitioners to assess the association with the use of auditory cues during perceptions of nonunderstanding. The studies reported in the current thesis measured operators' overall capacity for visual cue utilisation using the EXPERT intensive skills evaluation (*EXPERTise 1.0*) software package (Wiggins et al., 2010). EXPERTise 1.0 is an online Situational Judgement Test (SJT) that can be customised to assess operators' utilisation of visual cues during task-related activities in a variety of domains. It has differentiated high from low performers on the use of task-related visual cues within the domains of power control (Loveday, Wiggins, Harris, O'Hare, & Smith, 2013), pediatric diagnosis (Loveday, Wiggins, Searle, et al., 2013), aviation (Wiggins et al., 2014) and software engineering (Loveday, Wiggins, & Searle, 2013). The test-retest reliability of EXPERTise 1.0 has been demonstrated as satisfactory (Loveday, Wiggins, Festa, Schell, & Twigg, 2013).

Domains of Research Interest

To reliably assess how both auditory and visual cue associations facilitate situation assessment in complex situations, it was imperative to select participants within real-world technical environments where such cues have predictive value. It was also necessary to select domains that commonly use the readback/hearback protocol to deliver and confirm instructions over the radio/telephone. Recruiting from these domains ensures that operators are given sufficient exposure and opportunities to acquire the semantic and prosodic cues that manifest verbally during nonunderstandings (Kahneman & Klein, 2009). Finally, it was important to select domains where the application of visual cues during situation assessment occurs in a manner that is amenable to experimental replication and assessment.

To satisfy these criteria, the domains of electricity transmission control (Papers 1 and 2), electrical rail control (Paper 3), and hydroelectric power generation (Paper 4) were selected. Within each domain, a specific cohort of operators was identified[6] whose job function: (a) required adherence to the readback/hearback procedure during instructional exchanges over the radio/telephone, and; (b) primarily involved monitoring and maintaining the integrity of distribution networks through the use of supervisory screens.

These three domains represent many professional and technical environments where operators are required to attend to task-relevant information from visual interfaces to aid situation assessment (Hoc, Amalberti, & Boreham, 1995). Within the expert/novice framework, there is sufficient evidence to suggest that experienced operators can be distinguished from their less experienced counterparts through their superior ability to extract and utilise cues from visual interfaces (e.g., Elkerton & Williges, 1985; Speier, 2006; Sutter & Ziefle, 2005). For example, Elkerton and Willigies (1985) observed differences between expert and novice computer users with respect to their retrieval performance in a complex information-retrieval computer task. Therefore, the use of domains where situation assessment involves a process of extracting task-relevant cues from standardised visual displays is appropriate for investigating visual cue utilisation, as they enable a higher degree of experimental control while still maintaining ecological validity.

Summary of Thesis Contributions

The aim of this thesis is to identify particular prosodic and semantic cues that can aid the recognition of nonunderstandings within readback/hearback exchanges that take place within operational environments. This thesis is also expected to contribute to the literature supporting the role of visual cue utilisation in the process of situation assessment. Specifically, the current thesis is expected to highlight any relationships that may exist between visual and auditory cue utilisation during situation assessment within a variety of operational

[6] These specific cohorts of operators included *network operators* within electricity transmission control; *electrical control operators* within electrical rail control and *hydroelectric power generation controllers* within hydroelectric power generation.

environments. The research questions proposed by this thesis are addressed by detailed experimental studies, where the research outcomes are expected to better inform the direction of naturalistic studies that might extend the literature on nonunderstandings in the future. Operator performance is examined across a variety of distinct operational and organisational cultures to assess consistency of cue utilisation within the naturalistic environment.

Paper 1: Prosodic Cues and Perceptions of Nonunderstanding

Paper 1 Aim

The aim of Paper 1 was to determine whether the prosodic cues found to be important during perceptions of uncertainty (namely, intonation, inter-turn delay and filler; Brennan and Williams, 1995), are also used by operators in high-risk environments to gauge the extent to which a speaker has understood an instruction during a readback response over the radio. To ensure a high degree of ecological validity, it was important to recruit participants from high-risk, high-consequence environments.

Paper 1 was conducted within the context of high-voltage electricity transmission control, which involves the bulk transfer of electrical energy from generators to areas of demand. Electricity transmission network operators, who are responsible for monitoring and diagnosing system events from remote data screens, were recruited from two major power control companies in both Australia and New Zealand. Network operators were selected, as their role requires them to follow the readback/hearback protocol during operational exchanges over the telephone. As network operators are required to monitor visual interfaces for indicators of faults and consequently, initiate proceedings to resolve such issues, these operators are frequently responsible for delivering instructions to operators out in the field (i.e., 'field operators'). Arguably then, network operators have had ample opportunity to develop cognitive associations between the prosodic features of intonation, inter-turn delay and filler, and issues of nonunderstanding, if such associations do exist (Kahneman & Klein, 2009). In other words, network operators are frequently positioned in the 'sender' role during readback/hearback communicative exchanges, with the responsibility for detecting nonunderstandings on the part of the 'receiver' (i.e., field operator)

Paper 1 Publication History

Paper 1 was published in the Journal of Communication on the 11th May 2013. The Journal of Communication has an impact factor of 2.076 and ranks 3rd of 74 journals in the

Communication category for the 2013 Journal Citation Report. The author of the present dissertation wrote approximately 85% of Paper 1. See Appendix A for the original submission (N.B. the original submission has been subsequently removed from the Appendices for copyright purposes).

A modified version of Paper 1 was also published in the Proceedings of the *56th Annual Meeting of the Human Factors and Ergonomics Society (HFES), 2012*. This version included a modified design of that presented in Paper 1, and only included the Australian sample of network operators. The original submission of the HFES proceedings paper is provided in Appendix B of the thesis (N.B. the original submission has been subsequently removed from the Appendices for copyright purposes). Both these publications use American, rather than Australian, English.

Finally, as Paper 1 is largely grounded in communication theory, it does not include a theoretical discussion of the role of cue utilisation in the assessment of nonunderstandings. As such, the term ‘prosodic cue’ is exclusively used, rather than the more nuanced distinction between prosodic features and prosodic cues.

Prosodic Cues Used During Perceptions of Nonunderstandings in Radio Communication

Jaime C. Auton[1], Mark W. Wiggins[1], Ben J. Searle[1], Thomas Loveday[1], & Nan Xu Rattanasone[2]

1 Department of Psychology, Macquarie University, Sydney, Australia

2 Department of Linguistics, Macquarie University, Sydney, Australia

Author Note

This research was supported in part by grants from the 'Australian Research Council' and TransGrid Pty Ltd under the former's Linkage Program (Grant Number LP0884006).

Correspondence concerning this article should be addressed to Jaime Auton, Department of Psychology, Macquarie University, New South Wales, Australia, 2109.

E-mail: jaime.auton@mq.edu.au

Abstract

The readback/hearback protocol is a radio procedure intended to reduce communication errors in technical industries. It consists of the delivery of an instruction to a receiver, and the readback of that instruction by the receiver to confirm to the sender that it has been heard accurately. It does not, however, ensure that the receiver has understood the instruction. Using two samples within electricity transmission control, the present research explored whether the prosodic cues that listeners use to interpret uncertainty, are also used to judge whether a receiver is perceived to have understood an instruction. Rising intonation and filled inter-turn delays were identified as prosodic cues that operators use to identify nonunderstandings. Theoretical and practical implications of the findings are considered.

Keywords: prosodic cues, perceptions of nonunderstanding, spoken communication

Prosodic Cues Used During Perceptions of Nonunderstandings in Radio Communication

Communication is fundamental to the operation of many technical environments, including air traffic control, rail control, and high-voltage electricity transmission control (Australian Transport Safety Bureau, 2007; Fegyveresi, 1997). These technical environments are characterized by ill structured problems, uncertain and dynamic conditions, time stress and the potential for significant errors (Orasanu & Connolly, 1993).

The principal goal of any communicative exchange is to ensure mutual understanding between all participants (Cohen & Levesque, 1994; Dascal & Berenstein, 1987; Weigand, 1999). To achieve this goal, communicative participants usually follow the collaborative model, which places joint responsibility on both participants to arrive at a mutual understanding (Clark, 1994; Clark & Brennan, 1991; Clark & Schaefer, 1987, 1989; Clark & Wilke-Gibbs, 1986). This model highlights two distinct stages of a communicative exchange: the presentation of content and the grounding of that content (Clark & Schaefer, 1987).

During the first stage of a communicative exchange, the sender will present information for the receiver to interpret and derive understanding. During the second stage, the sender and receiver collaborate to establish the mutual belief that the receiver has understood correctly the sender's intended message (Clark & Schaefer, 1987). Grounding is accomplished when the receiver provides evidence that he/she has understood the sender's proposal and the sender accepts the validity of such evidence (Clark & Schaefer, 1987). Despite this goal of mutual understanding, problems in communication occur frequently during discourse (Hirst, McRoy, Heeman, Edmonds, & Horton, 1994; Kim, Park, Han, & Kim, 2010). In fact, problematic communication has been identified as one of the leading causes of errors in many technical industries (Kim, et al., 2010).

To minimize the risk of errors in communication, a number of industries have implemented preventative strategies, one of which is the readback/hearback protocol (Australian Transport Safety Bureau, 2007). This protocol is a radio or telephone procedure

that requires the receiver of an instruction to repeat, or *read back* the instruction to the sender (Streitenberger, Breen-Reid, & Harris, 2006). It consists of four distinct stages: First, a sender delivers a verbal instruction to a receiver; second, the receiver actively listens to the instruction; third, the receiver repeats the instruction verbatim to the sender (known as the *readback*); and finally, the sender actively listens for an accurate readback from the receiver (known as the *hearback*; Prinzo & Britton, 1993).

The primary objective of the readback/hearback protocol is for the sender to verify that the receiver has accurately heard the instruction (Prinzo & Britton, 1993; Schneider, Healy, & Barshi, 2004; Streitenberger et al., 2006). Nevertheless, this protocol does not necessarily establish whether the receiver has understood the instruction, since the ability to repeat a string of words does not necessarily ensure an understanding of the meaning of those words (Jones, 2003). As a result, the delivery of an accurate readback does not necessarily provide the sender with sufficient evidence during the grounding process to suggest that mutual understanding has taken place. The broad objective of this study was to establish those cues, in addition to an accurate readback that are used by senders to infer a lack of understanding on the part of the receiver.

Occasions where a receiver fails to understand the meaning of an instruction have the potential to result in errors. This type of error in communication is referred to as a *nonunderstanding* (Skantze, 2005). A nonunderstanding can occur when the receiver has either failed to obtain any meaningful interpretation of the instruction or cannot choose between two or more interpretations of the instruction (Hirst et al., 1994; Skantze, 2005).

An important component of a nonunderstanding is that receivers will immediately recognize that they have failed to understand an instruction (Hirst et al., 1994). This is distinct from the more common *misunderstanding*, in which receivers are under the false assumption that they have correctly interpreted the meaning of an instruction (Hirst et al., 1994; Skantze, 2005). Therefore, in the case of a misunderstanding, neither communicative partner is

immediately aware that an error has occurred.

The conscious nature of nonunderstandings is such that receivers can either seek clarification from the sender to establish mutual understanding or they can deceptively conceal their state of nonunderstanding and continue with the exchange (Weigand, 1999). Clearly, those instances where a receiver conceals a nonunderstanding pose the greatest risk for error as the sender remains unaware that a failure has occurred. Despite the implementation of readback/hearback communication strategies, nonunderstandings remain a common cause of communication problems in technical environments (Barshi & Farris, 2013; Howard, 2008; Morrow, Rodvold, & Lee, 1994).

The collaborative model suggests that, for grounding to take place, the sender will rely on evidence provided by the receiver to ascertain whether or not the intended message has been understood (Clark & Brennan, 1991). Since the readback/hearback protocol takes place predominantly over the radio or telephone, communication is strictly an exchange of spoken information (Fegyveresi, 1997). Therefore, a sender cannot monitor visual cues (such as head nods or raised eyebrows) as indicators of a receiver's level of understanding (Drummond & Hopper, 1991; Fegyveresi, 1997). Rather, senders must rely on the semantic content of the speech (the syllables, words or phrases of an utterance) and the prosodic cues of the voice to detect a nonunderstanding (Nygaard, Herold, & Namy, 2009; Nygaard & Queen, 2008).

Broadly defined, prosody refers to the array of paralinguistic cues that do not establish *what* speakers say, but rather *how* they say it (Bolinger, 1985; Cruttenden, 1986). Prosody encompasses a variety of speech properties including intonation, rhythm, intensity, pitch, timing and fillers (interjections such as *um* or *uh*; Hunter, 2004; Nygaard et al., 2009).

Although radio communication comprises both semantic (i.e., *what* is said) and prosodic (i.e., *how* it is said) information, a readback is intended to be an exact restatement of an already uttered text. When this protocol is adhered to, the semantic content of a readback will not vary and thus, cannot provide the sender with evidence of a receiver's level of

understanding. Therefore, a sender can only rely on the prosodic cues of an utterance as evidence of a receiver's level of understanding during the grounding process.

While there is a significant body of research relating to the prosodic and other nonverbal cues associated with the detection of various types of deceptive communication (see DePaulo et al., 2003 for an overview), there is relatively limited evidence available concerning the specific prosodic cues that are perceptually relevant to listeners' detection of concealed nonunderstandings (a type of deceptive communication). It could be argued that nonunderstandings arise where there is perceived uncertainty as to the nature of the information since uncertain information is considered incomplete, ambiguous, erroneous and/or imprecise (Woods, 1988). This highlights the potential relationship that exists between the constructs of uncertainty and nonunderstandings within a communicative exchange.

While there has been a great deal of interest in uncertainty as a construct within theories of communication, much of this research has focused on its impact in the context of romantic, intercultural, organizational and health communication (Berger, 2011). However, a small body of research has investigated the prosodic cues that listeners draw upon to detect perceived uncertainty amongst players of trivia question games. Specifically, Brennan and Williams (1995) required listeners to listen to audio-only responses to trivia questions, and rate the extent to which players 'sounded certain' about the accuracy of their responses. Brennan and Williams found that rising intonation, increased duration of inter-turn delays and the presence of a filler were salient prosodic cues that were detected by listeners in the responses of players where they were experiencing a level of uncertainty.

Intonation refers to the variations in pitch during the production of an utterance (Hunter, 2004). Brennan and Williams (1995) found that listeners perceived players' responses as more uncertain when they articulated their response with a rising intonation (i.e., response rising in pitch), rather than a falling intonation (i.e., response declining in pitch). Further, Brennan and Williams identified inter-turn delays (both filled and unfilled) as prosodic cues

that were identified by listeners as reflecting players' uncertainty.

An inter-turn delay refers to the duration of silence between the offset of one speaker's utterance and the onset of the other speaker's response (Roberts, Francis, & Morgan, 2006). Inter-turn delays containing fillers such as *um* and *uh* are labelled filled delays, whereas delays containing silence are known as unfilled delays (Goldman-Eisler, 1961). Brennan and Williams (1995) found that listeners rated players as more certain of the accuracy of their responses when the utterance was preceded by a shorter (one second), rather than a longer (five second) unfilled inter-turn delay. By contrast, when a filler occupied an inter-turn delay, listeners perceived a greater level of uncertainty amongst players (Brennan & Williams, 1995). This suggests that fillers may augment other prosodic features, such as unfilled inter-turn delays, and thereby enhance the listener's capacity to detect uncertainty in the communication of information.

Although Brennan and Williams (1995) offer a significant contribution, they did not consider the related stage of the communicative process whereby the uncertainty experienced by the receiver is succeeded by a nonunderstanding. Moreover, they did not consider the relatively unique situation in which the receiver of an instruction is required to read the instruction back to the sender verbatim in the absence of visual cues. This is an unusual situation since the opportunities for a sender to detect nonunderstandings are constrained to prosodic cues where the receiver is familiar with the terminology and can read back the instruction accurately, but may deliberately conceal a nonunderstanding.

On the basis of the relationship between uncertainty and nonunderstanding, the aim of the present research was to determine whether the prosodic cues that are used by listeners during perceptions of uncertainty are also used by listeners when perceiving the extent to which a receiver has understood an instruction during a simulated readback/hearback exchange. Specifically, it was hypothesized that readbacks ending in rising intonation would attract lower ratings of perceived understanding compared to readbacks ending in falling

intonation (H1). It was also predicted that unfilled inter-turn delays of longer duration would attract lower ratings of perceived understanding compared to unfilled inter-turn delays of shorter durations (H2). Finally, it was hypothesized that readbacks that were preceded by a filled inter-turn delay would attract lower ratings of perceived understanding compared to readbacks that were preceded by an unfilled inter-turn delay of the same duration (H3).

The Current Study

The primary aim of the present study was to identify, within the context of a technical domain, those prosodic cues that enable listeners of a readback response to discern different levels of understanding on the part of the receiver. To this end, it was important that the participants represented actual operators who frequently engaged in readback/hearback communicative exchanges. Therefore, the context for the present study was high-voltage electricity transmission control since it is an environment within which there is a requirement for readbacks verbatim, where communication occurs in the absence of visual cues, and where the consequences of an error are significant.

Electricity transmission control involves the safe and efficient transportation of high voltage electricity from electricity generators to areas of demand. Due to the small sample of network operators within each high-voltage electricity transmission network, the network operators who acted as participants were recruited from both an Australian and a New Zealand electricity transmission control network. Electricity transmission network operators are responsible for monitoring and diagnosing system events from remote data screens.

The participants were presented with a series of scripted, audio-only communicative exchanges emulating the readback/hearback protocol. Each exchange consisted of two elements: Firstly, a sender delivered an instruction to a receiver, and secondly, the receiver repeated the instruction to the sender (i.e., the readback). The readbacks were systematically manipulated based on the three prosodic cues that had previously been associated with interpretations of uncertainty (namely, intonation, unfilled inter-turn delays and filled inter-

turn delays). After listening to each exchange, participants were asked to rate the extent to which they thought that the receiver had understood the instruction based on the way that the instruction had been read back.

The assessment of perceived nonunderstanding was important in this case, since it is the sender (i.e., the recipient of the readback) who is charged with ensuring the accuracy of the readback and ultimately, the safety and security of the system. Moreover, both Brennan and Williams (1995) and Keysar and Henly (2002) have established that there are strong associations between perceptions of uncertainty and nonunderstanding on the part of a third-party listener, and the actual levels of uncertainty and nonunderstanding on the part of the receiver. This suggests that, through experience, operators have developed the skills necessary to identify key features of utterances that enable accurate interpretations of cognitive function.

Method

Participants

A total of 52 electricity transmission network operators from both an Australian and New Zealand electricity transmission control network participated in this study. Thirteen participants were initially excluded; eight due to technical difficulties, four who were not native speakers of English and one who did not complete the task. The remaining 39 electricity transmission network operators were native speakers of English who were predominantly male (97.44%) and ranged in age from 26 to 62 years (M = 43.79, SD = 10.45). The participants had accumulated between one and 40 years of experience within high-voltage electricity transmission control, with a mean of 20.00 years (SD = 12.74).

Design

The study comprised a 2 (Communicative Exchange 1 vs. Communicative Exchange 2) x 2 (Intonation: Rising vs. Falling) x 6 (Inter-turn Delay: -300 ms [unfilled], +300 ms [unfilled], +600 ms [unfilled], +600 ms [filled], +1200 ms [unfilled], +1200 ms [filled]) x 2 (Country: Australia vs. New Zealand) mixed design, with communicative exchange,

intonation, and inter-turn delay as the within-groups variables, country membership as the between-groups variable and years of experience within high-voltage electricity transmission control as a covariate, resulting in 24 distinct conditions.

The variable Communicative Exchange referred to the type of semantic content contained within the communicative exchange. Each exchange consisted of a sequence of two voices. Participants first heard a sender delivering an operational instruction to a receiver. They then heard the receiver repeating or reading back the instruction to the sender. To control for the potential effects of the semantic content of readbacks on ratings of perceived understanding, two semantically distinct communicative exchanges were included (i.e., Communicative Exchange 1 and Communicative Exchange 2). The two exchanges included instructions that specified different operational steps for the receiver to execute at different locations (see the Appendix for a de-identified version of the two communicative exchanges used). These exchanges were derived from the electricity company in Australia, and these differed from those commonly used in the New Zealand network insofar as they were considerably shorter and embodied different content (i.e., they referred to different locations and used different jargon). To control for any effects associated with the utilization of company-specific exchanges, the country from which participants were employed was included as a between subjects variable (i.e., Australia vs. New Zealand).

The variable Intonation referred to whether the readback portion of the exchange ended with either rising or falling intonation. Inter-turn Delay specified the duration between the offset of the sender delivering the instruction and the onset of the receiver providing the readback (where a positive number indicated a delay between speakers and a negative number indicated an overlap of speech between the end of the instruction and the beginning of the readback) and whether or not the delay was unfilled (i.e., contained only silence) or filled (i.e., contained the filler *um* embedded within the silence).

Finally, it could be argued that the skill in detecting the extent to which a receiver has

understood an instruction is a function of the number of years of experience within the domain. To account for this explanation, years of experience within electricity transmission control was included as a covariate.

The participants listened to the 24 exchanges in randomized order. After listening to each exchange, they were asked to rate the extent to which they perceived that the receiver had understood the meaning of the instruction. The participants subsequently rated the receiver's level of understanding on a 7-point scale from 1 (*did not understand at all*) to 7 (*understood completely*). This type of scale maintained consistency with Brennan and Williams' (1995) to measure perceptions of uncertainty. These ratings represented the dependent variable.

Stimuli Development

Communicative exchanges.

Auditory depiction of the communicative exchanges.

Two Australian English male speakers performed the roles of the sender and the receiver. The actors maintained the same system sender-receiver roles for both exchanges to control for any variation in responses that might result from the voice quality of the receiver. Actors were recorded reading their portion of the exchange (either the instruction or the readback) using a digital audio editor program (Audacity™) on a MacBook™ laptop. The mean duration of the first exchange was 33.47 seconds (*SD* = 0.57), while the mean duration of the second exchange was 34.09 seconds (*SD* = 0.57).

Semantic content of the communicative exchanges.

In consultation with two subject matter experts (selected on the basis of peer review and industry position) and actual documentation from the Australian electricity transmission company, two communicative exchanges were devised. Each exchange included both an instruction and a readback of that instruction. The instruction represented a typical operational instruction that receivers within the organization were likely to receive. Each readback of the

instruction was consistent with the protocol as to how instructions were expected to be read back within the organization, beginning with an acknowledgement of receipt of the instruction (*Yes I understand)* followed by the repetition of the instruction verbatim.

Prosodic manipulation of readbacks.

Each readback was manipulated to vary systematically the intonation, the length of the inter-turn delay between the offset of the sender's instruction and the onset of the receiver's readback, and the presence or absence of a filler during the inter-turn delay.

Intonation.

Intonation refers to the changes in pitch during the production of an utterance (Hunter, 2004). Intonation contours are described as sequences of high and low tones in pitch (Pierrehumbert, 1980). In general, when an utterance rises in pitch from onset to offset, it is referred to as rising intonation (Guy & Vonwiller, 1984; McGregor & Palethorpe, 2008). In contrast, utterances declining in pitch from onset to offset are referred to as falling intonation (Pierrehumbert, 1980).

To generate readbacks concluding with the appropriate intonation contour, the actor who performed the role of the receiver was instructed to utter the readback using a rising or falling intonation towards the *end* of the instruction (i.e., the last one or two words). To ensure that listeners perceived the intonation manipulations accordingly, a manipulation check was conducted.

Manipulation check for intonation contours.

Twenty-three undergraduate students from Macquarie University in Australia participated in the manipulation check. The sample was predominantly female (86.96%), ranging in age from 18 to 46 years (M = 21.00, SD = 6.27). All participants were native speakers of English. The participants were required to listen to the readback portion of both the rising and falling intonation versions of the two communicative exchanges. They were then asked to decide whether the speaker used a rising or falling intonation at the end of his

utterance. It is of note that the rating of intonation was validated by a predominantly female cohort (86.96%), where the vast majority of the electricity transmission network operators were male (97.44%). Although there have been gender differences reported in the perception of nonverbal cues (e.g., Hall, 1984), Edelsky (1979) reported no gender differences in the categorization of intonation contours during a manipulation check similar to the one used in the present study.

Overall, participants listened to four repetitions of each readback, listening and responding to a total of 16 readbacks. The results indicated that participants were able to identify rising intonation with 93.48% accuracy for the first communicative exchange and 94.57% for the second communicative exchange. Participants identified falling intonation with 97.83% and 90.22% accuracy for the first and second communicative exchanges respectively. Overall, participants distinguished rising from falling intonation across both communicative exchanges with an accuracy rate of 94.03%.

Because the manipulation of intonation (rising and falling) at the end of the readbacks was naturalistic and not electronically manipulated to preserve ecological validity, all of the other prosodic cues between the rising and falling versions of each readback could not be artificially controlled. This meant that other prosodic cues such as tempo, average pitch level and dynamics were not matched for the rising and falling intonation versions of the readback from Communicative Exchange 1 (and similarly, for Communicative Exchange 2). To minimize the risk that any significant effects found between falling and rising intonation were a product of some other difference between the two versions of the same readback, two semantically distinct readbacks were presented to participants. It should also be noted that this approach is consistent with the approach taken by Brennan and Williams (1995), who only matched the audio responses to trivia questions on intonation contours without controlling for other prosodic cues.

Unfilled inter-turn delays.

Brennan and Williams (1995) compared the effects of one second and five second unfilled inter-turn delays on perceptions of uncertainty. However, consultation with a subject matter expert, in conjunction with actual recordings of communicative exchanges derived from the Australian electricity transmission control company, suggested that a delay of five seconds between speakers would be unrealistic and thereby threaten the ecological validity of the study. As a result, the length of the inter-turn delays between the instructions and readbacks in the current study were based on previous research that has demonstrated that even within a relatively constrained range of delays (between -300 ms and +1200 ms), inter-turn delays were longer during problematic communication (including a lack of willingness to comply with a request, or issues of misunderstanding; Roberts, et al., 2006; Shimojima, Katagiri, Koiso, & Swerts, 2002).

On the basis of this research, four lengths of inter-turn delays were used to investigate the relationship between delay and perceptions of nonunderstanding: -300 ms (i.e., a 300 ms portion of overlapping speech between the end of the instruction and the beginning of the readback), +300 ms, +600 ms and +1200 ms. Silences of these durations were inserted to lengthen the delay between the offset of the sender giving the instruction and the onset of the receiver repeating the instruction. Considering these durations were relatively short, a manipulation check was run to ensure listeners would perceptually distinguish such durations.

Manipulation check for inter-turn delay durations.

During the same manipulation check that sought to validate the intonation contours, participants were also asked to attend to the inter-turn delays within the communicative exchanges to ensure that the inter-turn delay durations were not all perceived as the same length. To reiterate, 23 undergraduate students from Macquarie University in Australia took part in the manipulation check. All participants were native speakers of English, predominantly female (86.92%), ranging in age from 18 to 46 years (M = 21.00, SD = 6.27).

The participants were required to listen to 16 of the exchanges presented to the electricity transmission network operators (excluding the filled inter-turn delay exchanges) and were asked to decide whether the delay between the sender delivering the instruction and the receiver repeating the instruction was either ‘short’ or ‘long’. The participants identified -300 ms as short 100% of the time, +300 ms as short 96.74% of the time, +600 ms as short 51.09% of the time and +1200 ms as long 96.74% of the time. This pattern of results suggested that, although the range of inter-turn delays was restricted in comparison to Brennan and Williams (1995), listeners were still able to perceptually decipher most of the durations, as they did not categorize them all as either ‘short’ or ‘long’.

Filled inter-turn delays.

Unfilled inter-turn delays with a duration of either +600 ms and +1200 ms were also duplicated to include a filler *um* before the commencement of the readback. To construct the filled inter-turn delays, the actor performing the role of the receiver was asked to produce an *um* before repeating each instruction. This filler was then digitally inserted within an inter-turn delay of +600 ms (100 ms silence, *um* [400 ms], 100 ms silence) and +1200 ms (700 ms silence, *um* [400 ms], 100 ms silence). The type of filler was not counterbalanced across conditions with another filler such as *uh* as this would have required participants to listen to too many communicative exchanges. The use of the filler *um* was selected in preference for *uh* since Clark and Fox Tree (2002) found that speakers use *um* to signal a major delay and *uh* to signal a minor delay. Therefore, it was expected that the use of *um*, rather than *uh* would ensure that the filler would be more salient to participants.

It was not feasible to include a filled and unfilled inter-turn delay for the remaining durations (-300 ms and +300 ms), as the filler itself took approximately 400 ms to produce. Although inter-turn delay represents two distinct independent variables (both filled and unfilled), they were treated as one independent variable in the design of this study as filled delays are embedded within only some levels of unfilled delays (+600 ms and +1200 ms only).

Procedure

Participants took part in the study either during a series of work-related training sessions (Australia) or during scheduled work breaks (New Zealand). All participants were briefed on the purpose of the study before electing to participate. They completed a brief, demographic questionnaire before commencing a series of performance-based tests via computer. The task described in this study (the Auditory Readback task) was the first of five tasks to be completed by participants and took approximately 20 minutes to complete. The battery of tests was accessed by participants using a laptop computer and mouse. Participants worked in a sound attenuating room and wore over-ear noise-cancelling Panasonic RP-HC150 headphones throughout the Auditory Readback task. The Panasonic headphones had a frequency response range of 10Hz - 27kHz and sensitivity of 90 dB.

To prompt participants to attend to the prosodic cues embedded in the readbacks, rather than the semantic content when making judgements of the receiver's level of understanding, participants were told that the receiver would read back the instruction correctly (i.e., they would read back the correct operational steps and locations). Therefore, participants were instructed to listen to *how* the instruction was repeated, rather than *what* was repeated. After completing the battery of tests, they responded to a second questionnaire concerning their perceptions of the tasks that they had just completed, and self-assessments of fatigue, stress, workload and operational performance.

Results

The data were analyzed using a 2 (Communicative Exchange 1 vs. Communicative Exchange 2) x 2 (Intonation: Rising vs. Falling) x 6 (Inter-turn Delay: -300 ms [unfilled], +300 ms [unfilled], +600 ms [unfilled], +600 ms [filled], +1200 ms [unfilled], +1200 ms [filled]) x 2 (Country: Australia vs. New Zealand) mixed design analysis of covariance, with communicative exchange, intonation and inter-turn delay as the within subjects variables, country membership as the between subjects variable and years of experience within high-

voltage electricity transmission control as a covariate. Mean ratings of the perceived level of understanding of the receiver were the dependent variable.

The covariate was not significantly related to ratings of perceived understanding, $F(1, 35) = 0.74$, $p = .396$, $\eta_p^2 = .02$. This suggests that years of experience within the electricity transmission industry did not influence how participants rated levels of perceived understanding associated with the readbacks. Neither was there a significant effect for country membership, indicating that ratings from participants in Australia and New Zealand were similar, $F(1, 35) = 0.01$, $p = .942$, $\eta_p^2 = .00$.

By including the type of communicative exchange as an independent variable, the aim was to determine whether the utility of the prosodic cues to signal levels of understanding would depend upon the semantic content of the communicative exchange. When controlling for years of experience within high-voltage electricity transmission control, no main effect was evident for communicative exchange, $F(1, 35) = 0.01$, $p = .940$, $\eta_p^2 = .00$. Participants were responding relatively consistently across the two different communicative exchanges suggesting that the distinct semantic content of the two communicative exchanges did not influence ratings of perceived understanding.

Consistent with Hypothesis 1, the results revealed a statistically significant main effect for intonation when controlling for years of experience within high-voltage electricity transmission control, $F(1, 35) = 6.12$, $p = .018$, $\eta_p^2 = .15$. Participants associated a rising intonation with a relatively lower mean rating of perceived understanding ($M = 5.23$, $SE = 0.18$) in comparison to falling intonation ($M = 6.01$, $SE = 0.12$). This suggested that, as a prosodic cue, intonation may have the capacity to signal a receiver's level of perceived understanding during radio (auditory only) communication.

Mauchly's Test indicated that the assumption of sphericity had been violated for inter-turn delay, $\chi^2(14) = 29.85$, $p = .008$. Therefore, the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\varepsilon = .75$). In addition to intonation, a

significant main effect was also evident for inter-turn delay when controlling for years of experience within high-voltage electricity transmission control, $F(3.76, 131.52) = 5.66$, $p < .001$, $\eta_p^2 = .14$. Post-hoc pairwise comparisons using a Bonferroni adjustment indicated that no significant pairwise differences were evident for unfilled inter-turn delays of durations of -300 ms, +300 ms, +600 ms and +1200 ms, ($p > .05$), providing no support for Hypothesis 2. These results suggest that inter-turn delays of increasing duration do not necessarily signal to the listener of a readback exchange, the extent to which a receiver is perceived to have understood an instruction.

Consistent with Hypothesis 3, post-hoc pairwise comparisons indicated that there was a significant difference between the +600 ms (unfilled) and +600 ms (filled) *and* between the +1200 ms (unfilled) and +1200 ms (filled) inter-turn delays. For the inter-turn delay of +600 ms, a lower mean rating of perceived understanding was associated with the filled delay (M = 5.47, SE = 0.15) in comparison to the unfilled delay (M = 5.90, SE = 0.13), p = .001. A similar pattern was evident for the +1200 ms inter-turn delay, where the filled delay was associated with a lower mean rating of perceived understanding (M = 5.27, SE = 0.16) compared to the unfilled delay (M = 5.75, SE = 0.15), p = .004. Overall, these results suggest that filled inter-turn delays may signal a receiver's level of understanding during a readback. Specifically, participants reported a lower mean rating of understanding for readbacks that were preceded by *um* compared to those readbacks preceded by unfilled delays of the same duration.

Discussion

The overall aim of this study was to determine whether the prosodic cues that have been shown to signal uncertainty to listeners (namely intonation, unfilled inter-turn delays and filled inter-turn delays) were also useful to electricity transmission network operators when judging the extent to which a receiver had understood an instruction during a simulated readback/hearback communicative exchange. Consistent with Hypothesis 1, listeners

perceived that receivers had a lower level of understanding of the instruction when receivers read back the instruction with a rising intonation. Similarly, support was evident for Hypothesis 3, in which listeners perceived readbacks preceded by filled inter-turn delays as signalling a relatively lower level of understanding, compared to unfilled inter-turn delays of the same duration. However, the results failed to provide support for Hypothesis 2. Although unfilled inter-turn delays of longer durations have been associated with perceptions of greater uncertainty (Brennan & Williams, 1995), there was no association between perceptions of nonunderstanding and increases in the duration of inter-turn delays.

The lack of support for Hypothesis 2 could be attributed to one of two explanations. Firstly, the lengths of inter-turn delays used in the present study may not have been sufficient to signal a difficulty in communication within high-voltage electricity transmission control. Although previous research has suggested that delays of less than one second (such as those utilized in this study) are sufficient to signal a problem in communication (e.g., Shimojima et al., 2002), Jefferson (1989) suggests that a delay of approximately one second is needed to signal problematic communication. Therefore, future studies could investigate whether inter-turn delays of one second or longer signal issues of understanding within technical environments.

A second explanation for the lack of support for Hypothesis 2 is that inter-turn delay, irrespective of duration, is simply not a useful cue for a sender to gauge a receiver's level of understanding during a readback/hearback exchange. A substantial delay during an instructional exchange within a technical environment may be attributed to a variety of operational causes that are not necessarily a function of the cognitive state of the receiver. Therefore, it is possible that operators within electricity transmission control may have established that the inter-turn delay is a relatively unreliable cue when judging a receiver's level of understanding.

Although the prosodic cues of intonation, filled and unfilled inter-turn delays appear in

previous research to signal uncertainty to listeners (Brennan & Williams, 1995), the current study suggested that only intonation and filled inter-turn delays were perceptually relevant to electricity transmission network operators when judging levels of perceived understanding during radio communication. This is significant insofar as it suggests that it may be possible to detect nonunderstandings on the basis of prosodic cues, and suggests that nonunderstanding differs from uncertainty as an underlying construct.

Limitations and Future Research

Despite the positive outcomes of the present study, there were three issues that need to be considered in future research. Firstly, perceptions of understanding were made by a third party listener to a readback/hearback exchange, rather than an interactant who was part of the exchange. The role of a sender in a readback exchange typically includes the role of speaking as well as listening to a readback. The added process of speaking may impose greater cognitive demands than a third-party listener who is simply required to listen to a readback. This additional cognitive load may impact a sender's ability to attend to the prosodic cues, thereby reducing the likelihood that a nonunderstanding would be detected. Support for this proposition can be drawn from Keysar and Henly (2002) who found that, compared to a third party listener of an exchange, the cognitive demands inherent in speaking leads speakers to overestimate their communication effectiveness. Although this is an area for future research, the outcomes of the current study nevertheless demonstrate that, in the absence of the cognitive load imposed by speaking, intonation and filled inter-turn delays are important prosodic cues used by electricity transmission network operators to gauge a receiver's level of understanding.

A second limitation relates to the presentation of the 24 communicative exchanges in the same sequence. Although the order of exchanges was randomized within that sequence, the capabilities of the online system used during the study did not permit the exchanges to be presented to participants in a completely counterbalanced sequence. As a result, there was the

possibility that as participants became more familiar with the exchanges, there was a learning effect that may have resulted in differences in interpretations. However, all participants were informed prior to the commencement of the study, that the readbacks would vary in the form in which they were communicated (i.e., their prosodic cues would vary). The purpose of this instruction was to reduce the possibility of a demand characteristic, such that the detection of different prosodic cues would not alter ratings of nonunderstanding unless the cues were, in fact, meaningful.

Finally, in considering the utility of prosodic cues during perceptions of nonunderstanding, the current study targeted the three prosodic cues that Brennan and Williams (1995) investigated in their examination of perceptions of uncertainty. Although the findings from the current study point to both intonation and filled inter-turn delays as cues used in perceptions of nonunderstanding, other prosodic cues might have an equivalent or greater capacity to signal nonunderstandings. Therefore, there are opportunities for future research to consider the relative value of other features of nonverbal communication, including those common to deceptive communication such as rate of speaking and vocal tension (DePaulo et al., 2003).

At an applied level, the outcomes of the present study highlight the cues that need to be acquired by less experienced and/or non-native speakers to facilitate the grounding process during readback/hearback exchanges in operational environments. The acquisition of these cues may reduce the cognitive demands associated with the recognition of nonunderstandings on the part of the sender, thereby improving performance. Previous research has demonstrated the success of such cue-based learning approaches to improve cue detection in areas such as aviation (Wiggins & O'Hare, 2003), mining (Blignaut, 1979) and sport (Abernethy, Wood, & Parks, 1999; O'Hare, Wiggins, Williams, & Wong, 1998; Smeeton, Williams, Hodges, & Ward, 2005; Williams, Ward, Knowles, & Smeeton, 2002).

Although the adoption of a cue-based learning approach has the potential to assist the

identification of nonunderstandings in operational contexts, the prosodic cues identified in the present study have utility only to signal issues with understanding where the semantic content of an utterance is not sufficient. Therefore, the results of the current study are generalizable only to environments that utilize the readback/hearback protocol or another form of standardized phraseology. Air traffic control is one environment where the repetitive and procedural nature of communication may be vulnerable to errors (Cushing, 1994; Krieger, 2005), and thus, providing operators with another defense mechanism may help to reduce the frequency and/or severity of communication errors.

Using samples of electricity transmission network operators from Australia and New Zealand, the current research identified intonation and filled inter-turn delays as prosodic cues that might be used by electricity transmission network operators during perceptions of nonunderstandings in readback/hearback communicative exchanges. The outcomes have implications for the training of radio operators to detect nonunderstandings in readback/hearback exchanges in a variety of industrial settings, thereby reducing the potential for communication error.

Paper 1 References

Abernethy, B., Wood, J. M., & Parks, S. (1999). Can the anticipatory skills of experts be learned by novices? *Research Quarterly for Exercise and Sport, 70*, 313-318. doi:10.1080/02701367.1999.10608050

Audacity (1.2.5) [Free Audio Editor and Recorder Software]. Retrieved from http://audacity.sourceforge.net/

Australian Transport Safety Bureau. (2007). *Radiotelephony readback compliance and its relationship to surface movement control frequency congestion* (Investigation Report No. B2006/0053). Retrieved from http://www.atsb.gov.au/media/29962/b20060053.pdf

Barshi, I., & Farris, C. (2013). *Misunderstandings in ATC communication: Lanugage, cognition, and experimental methodology.* Burlington, USA: Ashgate Publishing Group.

Berger, C. R. (2011). From explanation to application. *Journal of Applied Communication Research, 39,* 214-222. doi:10.1080/00909882.2011.556141

Blignaut, C. J. H. (1979). The perception of hazard: II. The contribution of signal detection to hazard perception. *Ergonomics, 22,* 1177-1183. doi:10.1080/00140137908924692

Bolinger, D. (1985). *Intonation and its parts.* London: Edward Arnold.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34,* 383-398. doi:10.1006/jmla.1995.1017

Clark, H. H. (1994). Managing problems in speaking. *Speech Communication, 15,* 243-250. doi:10.1016/0167-6393(94)90075-2

Clark, H. H., & Brennan, S. A. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: APA Books.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um spontaneous speaking. *Cognition, 84,* 73-111. doi:10.1016/S0010-0277(02)00017-3

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes, 2,* 19-41. doi:10.1080/01690968708406350

Clark, H. H., & Schaefer, E. (1989). Contributing to discourse. *Cognitive Science, 13,* 259-294. doi:10.1207/s15516709cog1302_7

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39. doi:10.1016/0010-0277(86)90010-7

Cohen, P. R., & Levesque, H. J. (1994). Preliminaries to a collaborative model of dialogue. *Speech Communication, 15,* 265-274. doi:10.1016/0167-6393(94)90077-9

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

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

Dascal, M., & Berenstein, I. (1987). Two modes of understanding: Comprehending and grasping. *Language & Communication, 7,* 139-151. doi:10.1016/0271-5309(87)90004-8

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74-112. doi:10.1037/0033-2909.129.1.74

Drummond, K., & Hopper, R. (1991). Misunderstanding and its remedies: Telephone miscommunication. In N. Coupland, H. Giles, & J. Wiemann (Eds.), *"Miscommunication" and problematic talk* (pp. 301-314). Newbury Park: Sage Publications, Inc.

Edelsky, C. (1979). Question intonation and sex roles. *Language in Society, 8,* 15-32. doi:10.1017/S0047404500005923

Fegyveresi, A. E. (1997). *Vocal cues and pilot/ATC communications.* Paper presented at the Ninth International Symposium on Aviation Psychology, Columbus, OH. Abstract retrieved from

http://md1.csa.com/partners/viewrecord.php?requester=gs&collection=TRD&recid=A9828885AH&q=FEGYVERESI+1997&uid=1000599&setcookie=yes

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language and Speech, 18,* 18-26. doi:10.1177/002383096100400102

Guy, G. R., & Vonwiller, J. (1984). The meaning of intonation in Australian English. *Australian Journal of Linguistics, 4,* 1-17. doi:10.1080/07268608408599317

Hall, J. A. (1984). *Nonverbal sex difference: Communication accuracy and expressive style.* Baltimore: John Hopkins University.

Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. (1994). Repairing conversational misunderstandings and nonunderstandings. *Speech Communication, 15,* 213-229. doi:10.1016/0167-6393(94)90073-6

Howard, J. W. (2008). "Tower; am I cleared to land?": Problematic communication in Aviation discourse. *Human Communication Research, 34,* 370-391. doi:10.1111/j.1468-2958.2008.00325.x

Hunter, M. (2004). The music is the message: Prosody in aviation discourse. In M. A. Turney (Ed.), *Tapping diverse talent in Aviation: Culture, gender, and diversity* (pp. 129-151). Hampshire, England: Ashgate Publishing Ltd.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a 'standard maximum' silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective* (pp. 166-196). Clevedon, UK: Multilingual Matters.

Jones, R. K. (2003). Miscommunication between pilots and air traffic control. *Language Problems and Language Planning, 27,* 233-248. doi:10.1075/lplp.27.3.03jon

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science, 13,* 207-212. doi:10.1111/1467-9280.00439

Kim, A., Park, J., Han, S., & Kim, H. (2010). Development of extended speech act coding scheme to observe communication characteristics of human operators of nuclear power plans under abnormal conditions. *Journal of Loss Prevention in the Process Industries, 23,* 539-548. doi:10.1016/j.jlp.2010.04.005

Krieger, J. L. (2005). Shared mindfulness in cockpit crisis situations: An exploratory analysis. *Journal of Business Communication, 42,* 135-167. doi:10.1177/0021943605274726

McGregor, J., & Palethorpe, S. (2008). High rising tune in Australian English: The communication function of L* and H* pitch accent onsets. *Australian Journal of Linguistics, 28,* 171-193. doi:10.1080/07268600802308766

Morrow, D., Rodvold, M., & Lee, A. (1994). Nonroutine transactions in controller-pilot communication. *Discourse Processes, 17,* 235-258. doi:10.1080/01638539409544868

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustical and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science, 33,* 127-146. doi:10.1111/j.1551-6709.2008.01007.x

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1017-1030. doi:10.1037/0096-1523.34.4.1017

O'Hare, D., Wiggins, M., Williams, M., & Wong, W. (1998). Cognitive task analyses for decision centred design and training. *Ergonomics, 41,* 1698-1718. doi:10.1080/001401398186144

Orasanu J., & Connolly T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 3-20). Norwood, New Jersey: Ablex Publishing Company.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology). Retreived from http://hdl.handle.net/1721.1/16065

Prinzo, O. V., & Britton, T. W. (1993). *ATC/pilot voice communications: A survey of the literature* (Report No. DOT/FAA/AM-93/20). Washington, DC: Office of Aviation Medicine.

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in communication. *Speech Communication, 48,* 1079-1093. doi:10.1016/j.specom.2006.02.001

Schneider, V. I., Healy, A. F., & Barshi, I. (2004). Effects of instruction modality and readback on accuracy in following navigation commands. *Journal of Experimental Psychology: Applied, 10,* 245-257. doi:10.1037/1076-898X.10.4.245

Shimojima, A., Katagiri, Y., Koiso, H., & Swerts, M. (2002). Informational and dialogue-coordinating functions of prosodic features of Japanese echoic responses. *Speech Communication, 36,* 113-132. doi:10.1016/S0167-6393(01)00029-2

Skantze, G. (2005). Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication, 45,* 325-341. doi:10.1016/j.specom.2004.11.005

Smeeton, N. J., Williams, A. M., Hodges, N. J., & Ward, P. (2005). The relative effectiveness of various instructional approaches in developing anticipation skill. *Journal of Experimental Psychology: Applied, 11,* 98-110. doi:10.1037/1076-898X.11.2.98

Streitenberger, K., Breen-Reid, K., & Harris, C. (2006). Handoffs in care - Can we make them safer? *Pediatric Clinics Of North America, 53,* 1185-1195. doi:10.1016/j.pcl.2006.09.010

Weigand, E. (1999). Misunderstanding: The standard case. *Journal of Pragmatics, 31,* 763-785. doi:10.1016/S0378-2166(98)00068-X

Wiggins, M. W., & O'Hare D. (2003). Weatherwise: Evaluation of a cue-based training approach for the recognition of deteriorating weather conditions during flight. *Human Factors, 45,* 337-345. doi:10.1518/hfes.45.2.337.27246

Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied, 8,* 259-270. doi:10.1037/1076-898X.8.4.259

Woods, D. D. (1988). Coping with complexity: The psychology of human behavior in complex systems. In L. P. Goodstein, H. B. Andersen, & S. E. Olsen (Eds.), *Tasks, errors, and mental models* (pp. 128-148). New York: Taylor and Francis.

Paper 1 Appendix

Semantic Content of Communicative Exchanges

Communicative Exchange	Instruction	Readback
1	On X three zero zero nine four slash one, zero one X to X three thirty kV transmission line, I give you permission to carry out steps X three nine to X four one. That is removal of earths at X.	Yes I understand. X three zero zero nine four slash one that I have permission to carry out steps X three nine to X four one, on zero one X to X three thirty kV transmission line. That is removal of earths at X.
2	On X three zero one seven six slash two, X to X power station three thirty kV transmission line, I give you clearance to carry out steps X five to X one six. That is isolate only at X switching station.	Yes I understand. X three zero one seven six slash two, that I have clearance to carry out steps X five to X one six on X to X power station three thirty kV transmission line. That is isolate only at X switching station.

Note. Locations were removed and replaced with 'X' to maintain the confidentiality of participants.

Paper 2: Relationship between Visual and Auditory Cue Utilisation

Paper 2 Aim

The outcomes of Paper 1 suggested that electricity transmission network operators use intonation and fillers as prosodic cues during assessments of nonunderstanding in readback responses. Theoretically, if cue utilisation (irrespective of modality) is the cognitive mechanism that underpins superior performance during situation assessment, those operators who possess relatively greater levels of visual cue utilisation should demonstrate a more nuanced skill in differentiating prosodic cues that are diagnostic of nonunderstanding (e.g., rising intonation, presence of a filler) from those features that are not diagnostic of nonunderstanding (e.g., falling intonation, absence of a filler). Therefore, the aim of Paper 2 was to examine whether a relationship exists between operators' visual cue utilisation performance and their prosodic cue utilisation during assessments of nonunderstanding.

Paper 2 was also conducted within the context of electricity transmission control. This context was selected for two reasons:

(1) Paper 1 established that there is a general use of particular prosodic cues during assessments of nonunderstanding. Since network operators within this context frequently engage the readback/hearback protocol over the telephone, this context would arguably facilitate the development of more nuanced associations between prosodic features and nonunderstandings for those operators with greater experience. Consequently, it represents an ideal context in which to examine differences in the way that operators use prosodic cues.

(2) A primary function of the role of network operator is to monitor Supervisory Control And Data Acquisition (SCADA) systems via computer screens to assess the system state. During the process of situation assessment therefore, network operators rely on the information presented in the form of visual cues displayed by the SCADA machine interfaces to respond to critical situations. For example, a particular system alert may prompt a network operator to seek additional information from a separate substation. Thus, it should be possible

to differentiate network operators on the basis of visual cue utilisation performance using a series of naturalistic tasks involving the application of visual cues during situation assessment.

The study reported in Paper 2 was conducted at the same time, with the same network operators as participants, as the study reported in Paper 1. To ensure that there would be sufficient variation in visual cue utilisation performance across participants within the context of electricity transmission control, Paper 2 extended the sample collected in Paper 1 by incorporating an additional sample of current undergraduate students and recent graduates within the engineering discipline. These additional participants were inexperienced in the domain of electrical transmission control but were expected to have some basic theoretical understanding of the domain due to the nature of their tertiary degree. It was expected therefore, that there would be a sufficient variation in visual cue utilisation amongst the three samples of participants (i.e., students/recent graduates, Australian network operators and New Zealand network operators) to enable reasonable comparisons.

It is important to note that as the studies presented in Papers 1 and 2 were conducted at the same time (with the exception of the additional cohort of engineering students and graduates), the auditory stimuli that participants were asked to listen and respond to, were identical between papers. As a result, the manipulation check for the final intonation contours of the communicative exchanges described in Paper 1, was the same manipulation check described in Paper 2.

Paper 2 Publication History

As a result of the outcomes of Paper 2, it has not been submitted for publication. The author of the present thesis wrote approximately 90% of Paper 2.

Abstract

Within high-risk, high-consequence environments, expertise has typically been associated with a greater capacity to extract and utilise task-relevant visual cues during situation assessment. Much of the empirical evidence to support the notion that skilled cue utilisation distinguishes expert and novice practitioners however, has focussed on the use of *visual* cues, despite the importance of auditory, tactile and olfactory cues within 'real world' environments. Within the context of electricity transmission control, the current study explored whether there was a relationship between visual cue utilisation performance and prosodic cue utilisation during assessments of nonunderstanding of readback responses in radio communication. Although a positive relationship was predicted, no relationship was evident between the interpretation of the prosodic cues to signal nonunderstanding on the part of a speaker and visual cue utilisation. The theoretical significance of these results is discussed.

The Relationship Between Visual Cue Utilisation and Prosodic Cue Utilisation During Assessments of Nonunderstanding in Readback Responses

In the medical domain, a clinician's diagnosis of a patient's physical condition represents an important component in determining the nature of subsequent care (Croskerry, 2009). Within other high-risk, high-consequence environments, such as aviation and electricity transmission control, operators also invest considerable effort in assessing the state of a situation (Klein, 1989, 1993; Randel, Pugh, & Reed, 1996; Rasmussen, 1993; Wellens, 1993). An accurate assessment of a situation can often direct operators towards the most reasonable course of action under the circumstances (Kaempf, Klein, Thordsen, Marvin, & Wolf, 1996; Klein, 1997, 2003; Randel et al., 1996). *Situation assessment* is a process that involves understanding and identifying the current state of a situation by attending to, and interpreting, the available environmental information (Klein & Calderwood, 1991; Wickens & Hollands, 2000).

According to Klein's (1993, 1997) Recognitional-Primed Decision (RPD) model, environmental *cues* underpin the process of situation assessment. Conceptually, cues can be understood as pieces of environmental information that consist of associations in memory between feature/s and events/objects (Ericsson & Kintsch, 1995; Wiggins, 2012). Cue associations are formed when features in the environment are repeatedly paired with an object/event, such that they hold meaning for an operator (Lipshitz, Klein, Orasanu, & Salas, 2001; Wiggins & O'Hare, 2003). As a result of extensive experience within a particular domain, an operator is expected to develop a repertoire of cue associations that represent different situations (Coderre, Mandin, Harasym, & Fick, 2003; Ericsson & Kintsch, 1995; Klein, 1993). These environmental cues are held within Long Term Memory (LTM) and can be auditory, olfactory, tactile or visual in nature (Wiggins, 2006).

The RPD model describes situation assessment as taking place through a process of non-analytical pattern recognition (Kaempf et al., 1996; Klein, Calderwood, & Clinton-

Cirocco, 1986; Klein, 1989, 1997). In any given situation that requires assessment, cues are available for the operator to which to attend and from which to extract meaning. The operator will then draw a cognitive inference about the situation by 'matching' the pattern of cues (or less often, a singular cue; Ebbesen & Konečni, 1975) available in the situation to a mental representation for a similar situation recalled from LTM (Coderre et al., 2003; Kaempf et al., 1996; Simon, 1992; Wickens & Flach, 1988; Wickens & Hollands, 2000). This match between the current situation and the one retrieved from LTM is expected to prompt an 'action script' that includes typical ways of responding in similar situations (Kaempf et al., 1996; Klein, 1989, 2003; Lipshitz et al., 2001). It is this process of pattern matching in high-risk, dynamic and time-pressured environments that requires limited cognitive processing and allows for timely interventions (Lipshitz et al., 2001; Wiggins, 2012).

Consistent with Klein's (1993, 1997) RPD model, Brunswik's (1955) Lens Model also suggests that cognitive judgements, such as situation assessment, take place through a process of cue utilisation. This model assumes that a cognitive judgement is the result of an individual operator attending to many different environmental cues that may or may not be highly predictive of a situation (Newell, Lagnado, & Shanks, 2007; Wigton, Hoellerich, & Patil, 1986). As the RPD model also implies, the way in which an individual operator will extract and integrate particular features to form a judgement, will depend on the associations that an individual operator holds in memory between features and events that have been established through experience (Balzer, Doherty, & O'Connor, 1989; Bisantz & Pritchett, 2003; Karelaia & Hogarth, 2008).

Although distinct, Klein's (1993, 1997) RPD model and Brunswik's (1955) Lens Model both posit that pattern recognition, through the utilisation of associations between features and events, is the cognitive mechanism that underlies situation assessment. Empirical support for the utilisation of cues during situation assessment in operational contexts can be derived from evidence to suggest that experts differ from novices across a range of variables, including the

types of features accessed, the sequence of feature acquisition, and the time spent examining features (Abernethy, 1990; Blignaut, 1979; Coderre et al., 2003; Müller, Abernethy, & Farrow, 2006; North, Williams, Hodges, Ward, & Ericsson, 2009; Wiggins & O'Hare, 2003).

Amongst experts, this superior ability in cue utilisation has been the product of extensive experience or deliberate practice within a domain (typically a decade; Ericsson, Krampe, & Tesch-Römer, 1993), and has been demonstrated in a range of contexts. For example, using gaze data in the domain of mining, Blignaut (1979) found that expert miners could be differentiated from their less experienced counterparts by their superior ability to perceive cues indicative of dangerous rock falls. Similarly, highly skilled batsman have a unique capacity to attend to early movement cues to predict the opposing bowlers' intentions; a skill that is not evident amongst less skilled players (Müller et al., 2006). Within the domain of pediatrics, Loveday, Wiggins, Searle, Festa and Schell (2013) were also able to distinguish two groups of practitioners based on their use of cues during diagnosis. Using an online situational judgement test, EXPERT intensive skills evaluation (*EXPERTise 1.0*; Wiggins, Harris, Loveday, & O'Hare, 2010), Loveday et al. distinguished between two 'typologies' on the basis of their utilisation of visual cues during task-related activities, and found that these two typologies differed in diagnostic accuracy and their ability to sustain performance without the assistance of dynamic information.

There has been reasonably consistent empirical support for the proposition that experienced operators can be reliably discriminated from their less experienced counterparts on the basis of their relatively greater ability to recognise and utilise task-relevant cues during situation assessment (e.g., Calderwood, Klein, & Crandall, 1988; Coderre et al., 2003; de Groot, 1965; Hoffman, Aitken, & Duffield, 2009; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; North et al., 2009; Randel et al., 1996; Schriver, Morrow, Wickens, & Talleur, 2008; Shim, Carlton, Chow, & Chae, 2005). However, this research has focused primarily on how *visual* cues are utilised during situation assessment, to the exclusion of other modalities. For

example, Scialfa et al. (2011) compared the hazard detection skills of novice and experienced drivers using filmed traffic scenes. While experienced drivers were faster and more accurate in identifying hazards, the driving scenes to which participants were asked to respond were silent, restricting participants' situation assessment to visual cues.

The focus on visual cues to the exclusion of other modalities is significant since situation assessment in 'real-world' situations, including aviation, medicine, electricity transmission control and fire fighting, requires operators to attend to cues outside the visual domain. If the underlying mechanism of superior performance during situation assessment is determined by cue utilisation (irrespective of modality), then the same effects should be apparent for situation assessments made on the basis of auditory, olfactory or tactile cues.

The Current Study

The primary aim of the current study was to investigate whether operators from a high-risk, high-consequence environment, categorised on the basis of visual cue utilisation performance, differed in the way that auditory cues were utilised during assessments of *nonunderstanding*. A nonunderstanding is a type of miscommunication that occurs when the receiver of a message has failed to obtain any meaningful interpretation of the message (Hirst, McRoy, Heeman, Edmonds, & Horton, 1994; Skantze, 2005).

Nonunderstandings can either be *disclosed* (i.e., the receiver alerts the sender of the message to the fact that the message has not been understood), or *concealed* (i.e., the receiver deceptively conceals a state of nonunderstanding from the sender; Weigand, 1999). Nonunderstandings are especially evident in the absence of visual cues, since the receiver is reliant solely on the semantic ('what' is said) and prosodic ('how' it is said) cues of the communication to establish meaning (Nygaard & Queen, 2008).

Despite significant advances in technology, communication in the absence of visual cues remains the prevailing means of information transmission and receipt in many environments including aviation, rail, and shipping (Fegyveresi, 1997). Many of these

environments implement the readback/hearback protocol during instructional exchanges over the telephone/radio to reduce the likelihood of miscommunication between operators (Prinzo & Britton, 1993). This is a four-step protocol that begins with a sender transmitting a verbal message to a receiver, to which the receiver is expected to actively listen. During the third step, the receiver repeats the instruction verbatim (known as the *readback*). During the final step, the sender is expected to listen to the readback to ensure that it has been repeated correctly, which is referred to as the *hearback* portion of the exchange (Cushing, 1994; Prinzo & Britton, 1993).

Although the readback/hearback protocol is an effective strategy to identify when the receiver has misheard an instruction, it does not necessarily protect against concealed nonunderstandings, since a nonunderstanding can be concealed, despite an accurate readback (Jones, 2003). In fact, nonunderstandings remain a common cause of communication problems in technical environments (Air Accidents Investigation Branch, 2008; Morrow, Rodvold, & Lee, 1994).

The current study was conducted within the context of electricity transmission control. Electricity transmission control involves the safe and efficient transportation of high voltage electricity from electricity generators to areas of demand, and was selected for three primary reasons: (1) The readback/hearback protocol is a requirement during instructional exchanges between operators over the telephone; (2) The primary role of network operators within this domain is to monitor and maintain the integrity of distribution networks through the use of supervisory screens, allowing for the experimental measurement of visual cue utilisation during task-related activities, and; (3) The consequences of an error are significant.

Arguably, as a product of their experience within a domain, experts will have had a greater opportunity to develop associations between specific auditory features and events that might be indicative of a nonunderstanding. Therefore, if cue utilisation does underpin expert performance, a difference should be apparent in the way that experts and novices use auditory

features to assess nonunderstandings in those environments where communication occurs in the absence of visual cues.

Auton, Wiggins, Searle, Loveday, and Xu (2013) were the first to investigate the auditory cues used by listeners to assess *concealed* nonunderstandings within the context of readback/hearback communicative exchanges. It was important to assess the auditory cues that influenced a listener's *perception* of nonunderstanding on the part of a speaker, rather that the auditory cues *produced* by speakers in a state of nonunderstanding, as it is ultimately the responsibility of the sender (i.e., the listener of the readback) to assess whether or not the receiver (i.e., the speaker of the readback), has understood the instruction during a readback/hearback exchange.

Auton et al. (2013) investigated the prosodic cues that influenced listeners' assessments of nonunderstanding, over and above the semantic content of the readback. This was due to the fact that a readback, by definition, is a repetition of an already uttered text (i.e., the sender's instruction), and therefore, when an accurate readback is produced, the prosodic cues incorporated within a readback response are the only source of information that can indicate a receiver's level of understanding. Auton et al. investigated whether three prosodic features (namely, intonation, inter-turn delay and filler) were associated with listener perceptions of nonunderstanding on the part of the receiver, since these prosodic cues have previously been shown to influence perceptions of uncertainty (Brennan & Williams, 1995).

Auton et al. (2013) required participants to listen to a series of simulated communicative exchanges, where a sender delivered an instruction to a receiver, and a receiver delivered a readback of the instruction. The readback portion of each exchange varied in levels of intonation, inter-turn delay and filler, to assess their relationship with perceptions of nonunderstanding on the part of the participant (i.e., listener).

Auton et al. (2013) found that the prosodic cues of intonation and filler had the capacity to signal a speaker's state of understanding to a listener during radio communication. In the

case of intonation, listeners perceived that receivers had a lower level of understanding of the instruction when they read back the instruction with a rising intonation (i.e., response rising in pitch), rather than a falling intonation (i.e., a response declining in pitch). For filler, listeners perceived readbacks preceded by the interjection *um* as signalling a relatively lower level of understanding, compared to readbacks that were not preceded by a filler. Overall, these findings suggest that participants have developed auditory-based cue associations in memory between features (i.e., intonation and filler, independently) and an event (i.e., an issue of understanding on the part of the speaker).

Research aims and hypotheses.

The present study was designed to establish whether cue utilisation constitutes a trait that influences the acquisition and application of cues, irrespective of modality. Specifically, it was designed to examine whether visual cue utilisation is associated with auditory cue utilisation and namely, the use of prosodic cues during assessments of nonunderstanding. Visual cue utilisation was assessed using *EXPERTise 1.0* (Wiggins et al., 2010), an online situational judgement test that has previously differentiated higher from lower performers in a range of operational contexts. Auditory cue utilisation was assessed using the Auditory Readback Task, similar to the one described in Auton et al. (2013), to assess the use of prosodic features during perceptions of nonunderstandings in radio communication.

In comparison to participants with lesser levels of visual cue utilisation, it was hypothesised that participants with greater levels of visual cue utilisation would perceive a greater difference in levels of understanding on the part of the receiver between:

(H1) Readback responses that ended in rising intonation compared to readback responses that ended in falling intonation.

(H2) Readback responses that were preceded by a shorter compared to a longer inter-turn delay.

(H3) Readback responses that were preceded by a filler compared to an unfilled inter-turn delay of the same duration.

Method

Participants

It was important to recruit participants who had a vast range of experience within electricity transmission control to ensure that there would be variation in performances across the visual cue-based diagnostic tasks. Participants who were likely to have accumulated greater experience within electricity transmission control ('network operators') were recruited from two major electricity transmission networks located both nationally (Australia; N = 22) and internationally (New Zealand; N = 17). Undergraduate students and recent graduates from an Engineering degree from a variety of tertiary institutions also participated (N = 6). These participants had no direct experience within electricity transmission control but were assumed to have some basic theoretical understanding of the domain due to the nature of their tertiary degree. Ten participants were subsequently excluded for missing data (see Results section for further details).

All 35 participants were native English speakers. The network operators from Australia (N = 21) were exclusively male and ranged in age from 26 to 62 years (M = 44.38, SD = 11.37). This cohort had accumulated between two and 40 years of experience within electricity transmission control (M = 24.02, SD = 11.93). The network operators from New Zealand (N = 11) were predominantly male (90.9%), ranged in age from 31 to 58 years (M = 43.10, SD = 7.98), and had accumulated between one and 32 years of experience within electricity transmission control (M = 13.40, SD = 11.62). The students and recent graduates (N = 3) were between 20 and 29 years of age (M = 23.67, SD = 4.73) and reported no prior experience within electricity transmission control.

Materials

EXPERTise 1.0 test battery.

EXPERTise 1.0 is a battery of diagnostic tasks that can be customised to assess operators' utilisation of visual cues during task-related activities in a variety of domains (Wiggins et al., 2010). Typologies of behaviour are calculated based on performance across all tasks, to distinguish between participants with relatively greater levels of cue utilisation and participants with relatively lesser levels of cue utilisation. The validity of EXPERTise 1.0 has been established in power control (Loveday, Wiggins, Harris, O'Hare, & Smith, 2013), pediatric diagnosis (Loveday, Wiggins, Searle, et al., 2013) and aviation decision-making (Wiggins, Azar, Hawken, Loveday, & Newman, 2014). The current study used an abbreviated version of the EXPERTise 1.0 test battery, which comprised three tasks that required the application of cues: the *Paired Association Task*, the *Feature Discrimination Task* and the *Transition Task*.

Stimuli.

Paired Association Task.

During the Paired Association Task, participants were presented with text-based domain-relevant feature-event pairs, and were asked to rate the strength of the perceived relationship between the feature and event (Morrison, Wiggins, Bond, & Tyler, 2012). As experienced operators are expected to possess a greater repertoire of domain-relevant features and events in memory (Schriver et al., 2008), they should demonstrate a greater discrimination for ratings of association between pairs, compared to their less experienced counterparts (Morrison et al., 2012).

In the current study, the Paired Association Task was completed in two stages. In the first stage, participants viewed text-based feature-event pairs relevant to the electricity transmission control domain. Each feature-event pair was presented *sequentially* such that the feature (e.g., *High Frequency*) would appear on-screen shortly before the event (e.g., *Voltage*

Depression). Each stimulus was available on screen for 1500 ms, after which, participants were asked to rate the strength of perceived association between the feature-event on a 6-point Likert scale, from 1 (*extremely unrelated*) to 6 (*extremely related*). During the second stage of the task, participants were asked to rate the strength of perceived association between the feature-event pairs that were presented *simultaneously* on screen for 1500 ms. A total of 29 feature-event pairs were rated in each stage (see Appendix A).

Feature Discrimination Task.

During the Feature Discrimination Task, participants were presented with two complex and ambiguous scenarios within the context of electricity transmission control. Each scenario contained several environmental features incorporated within a brief written synopsis of the situation, as well as a Supervisory Control and Data Acquisition (SCADA) interface screen. Participants were asked to consider the information available in the scenario and select an appropriate response from several possible options. They then rated the relative usefulness of the available features associated with the situation when deciding the appropriate response. Individual ratings were made on a 10-point Likert scale, from 1 (*not important at all*) to 10 (*extremely important*). See Appendices B and C for the two Feature Discrimination Tasks. This task was based on the conceptualisation of the Cochran-Weiss-Shanteau (CWS) index of expertise (Shanteau, Weiss, Thomas, & Pounds, 2002). This index is based on the proposition that, compared to their less experienced counterparts, experienced operators are better able to discriminate consistently between features.

Transition Task.

The objective of the Transition Task was to assess the sequence in which operators extract task-related information from a visual display. Less experienced operators are more likely to acquire information in an order consistent with its visual presentation. In contrast, more experienced operators tend to extract information on the basis of relevance, rather than the sequence of presentation (Wiggins & O'Hare, 1995).

During the Transition Task, participants were presented with an intentionally incomplete descriptive scenario such that additional information would need to be obtained prior to making an accurate assessment of the situation. Additional information could be extracted from a 'drop down' list, containing thirteen distinct information categories. To view the information contained within each category, participants were instructed to click on the relevant button with their computer mouse. The sequence in which participants accessed additional information was recorded (see Figure 1 and Appendix D for the full scenario).

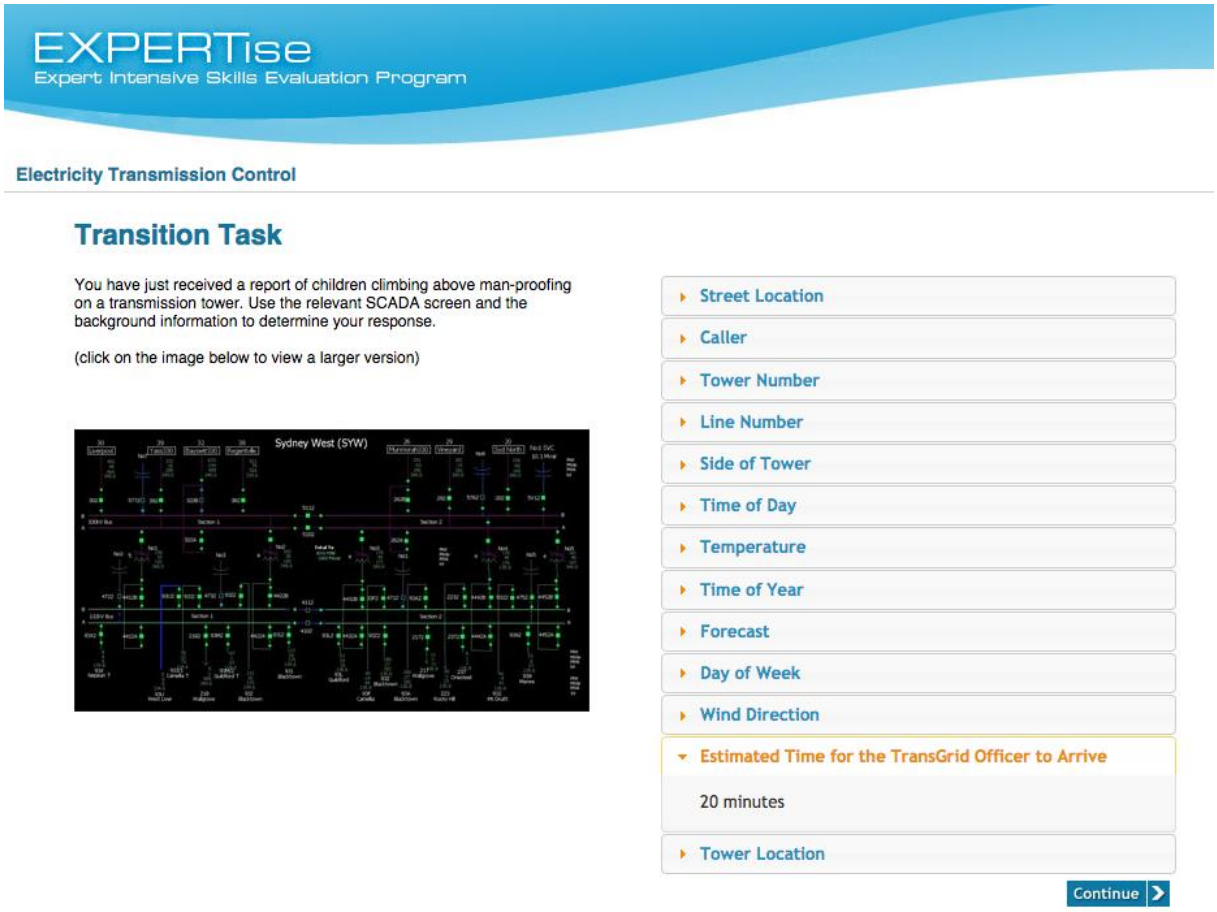


Figure 1. Screen capture of the information presented to participants during the Transition Task. Categories of features are displayed in a list on the right hand side, which can be expanded by a mouse click (currently, the feature 'Estimated Time for the TransGrid Officer to Arrive' has been clicked and expanded).

Stimuli development.

The EXPERTise 1.0 tasks were developed from cognitive interviews conducted with a Subject Matter Expert (SME) from the Australian electricity transmission control company. Selected on the basis of peer reference and industry position, the SME was asked to recall and

recount in detail, three operational critical incidents that had occurred in the past six months. Consistent with the critical incident technique (Flanagan, 1954; Klein, Calderwood, & MacGregor, 1989), it was emphasised that the incidents described should be relatively unusual and difficult, where the SME's experience was considered to influence the outcome. Using a semi-structured interview, various details relating to the decision making process of each incident were then ascertained (Klein et al., 1989). This is consistent with the standard approach for developing stimuli for EXPERTise 1.0 (Loveday, Wiggins, Harris, et al., 2013; Loveday, Wiggins, Searle, et al., 2013; Wiggins, Brouwers, Davies, & Loveday, 2014; Wiggins et al., 2010).

The three incidents described during the cognitive interviews were arranged into chronological timelines and were used to develop the scenarios for the two Feature Discrimination Tasks and the Transition Task. An extensive list of feature-event pairs that an operator may or may not find relevant during decision making were also developed from the cognitive interviews. These feature-event pairs formed the items in the Paired Association Tasks.

Auditory Readback Task.

Design.

The Auditory Readback Task incorporated two levels of the within-groups variable Communicative Exchange (Communicative Exchange 1, Communicative Exchange 2), two levels of the within-groups variable Intonation (Rising, Falling), two levels of the within-groups variable Inter-turn Delay (Shorter, Longer), and two levels of the within-groups variable Filler (Present, Absent), resulting in 16 distinct communicative exchanges.

The variable 'Communicative Exchange' referred to the two semantically distinct auditory communicative exchanges included in the design. Each exchange consisted of an operational instruction (spoken by a 'sender') and a readback of that instruction (spoken by a 'receiver'). The two exchanges differed insofar as they contained different operational

instructions for the sender to execute (see Appendix E for the semantic content of the communicative exchanges). It was important to incorporate two semantically distinct exchanges in the design, such that it could be examined whether the use of prosodic cues during assessments of nonunderstanding were contingent on the semantic content of the exchange.

Intonation refers to the variations in pitch, or the 'melody' of an utterance, where utterances end on either a relatively high or low tone (Hunter, 2004; Pierrehumbert, 1980). 'Intonation', as a variable, specified whether the end of the readback response produced by the 'receiver' concluded on either a relatively high tone (i.e., rising intonation) or a relatively low tone (i.e., falling intonation). 'Inter-turn delay' referred to whether the duration of silence between the sender delivering the instruction, and the receiver producing the readback, was either shorter (600 ms) or longer (1200 ms; Roberts, Francis, & Morgan, 2006). 'Filler' indicated whether or not a receiver produced a filler (i.e., a verbal injection '*um*') prior to the delivery of the readback (Goldman-Eisler, 1961).

After listening to each exchange, participants were asked to rate the extent to which they perceived that the receiver had understood the meaning of the instruction, on a 7-point scale from 1 (*did not understand at all*) to 7 (*understood completely*) (Auton et al., 2013). These ratings represented the dependent variable.

Stimuli development.

Communicative exchanges.

Two Australian English male speakers recorded the audio versions of the communicative exchanges using a digital audio editor and recorder program (*Audacity*™) on a MacBook™ laptop. Each speaker maintained either the role of the sender (i.e., the speaker of the instruction) or the receiver (i.e., the speaker of the readback) across the two exchanges, to ensure that any variations in assessments of nonunderstanding could not be attributed to the voice quality of the receiver. The mean duration of the first exchange was 33.47 seconds (*SD*

= 0.57), while the mean duration of the second exchange was 34.09 seconds (*SD* = 0.57).

The semantic content of the communicative exchanges were devised in consultation with two SMEs and actual documentation from the Australian electricity transmission company. The instruction portion of each communicative exchange represented a typical operational instruction within the Australian electricity transmission company. The readback portion of each exchange was consistent with how instructions were expected to be read back within the organisation, beginning with an acknowledgement of receipt of the instruction (*Yes I understand)* followed by the repetition of the instruction verbatim (see Appendix E).

Prosodic manipulation of readbacks.

The readback portions of the communicative exchanges were systematically manipulated to include the three prosodic cues that were assessed by Auton et al. (2013) during perceptions of nonunderstandings. Each readback was manipulated to vary systematically; (1) the final intonation at the end of the readback; (2) the duration of the inter-turn delay between the offset of the sender delivering the instruction and the onset of the receiver delivering the readback, and; (3) the presence of a filler (*um*) during the inter-turn delay, immediately prior to the delivery of the readback.

Firstly, to generate readbacks that concluded in either rising or falling intonation, the male speaker who played the role of the 'receiver' was asked to produce a readback where his voice either dropped or rose in pitch towards the end of the readback. To validate this naturalistic manipulation of intonation and ensure that listeners perceived the intonation contours accordingly, a manipulation check was conducted (see Manipulation Check section).

The current study also compared the influence of the duration of the inter-turn delay on assessments of nonunderstanding. Auton et al. (2013) found that the inter-turn durations of -300 ms (an overlap in speakers of 300 ms), 300 ms, 600 ms and 1200 ms had no influence on perceptions of nonunderstanding during readback/hearback communicative exchanges. Irrespective of this null finding, the current study included inter-turn delay as the usefulness

of this prosodic cue may have been masked by differences in visual cue utilisation that were not considered in Auton et al.

The current study compared the effects of two inter-turn delays on assessments of nonunderstanding. Shorter (600 ms) and longer (1200 ms) inter-turn delay durations were selected, as previous research has found that such an incremental increase between speaker turns has been associated with perceptions of 'trouble' in communication (Auton, Wiggins, & Loveday, 2012; Roberts et al., 2006). Using the digital audio and recorder software Audacity™, silences of these durations were electronically inserted between the instruction (delivered by the sender) and the readback (delivered by the receiver).

Finally, half of the readbacks presented to participants were preceded by a filler (*um*). The male speaker who played the role of the 'receiver' was asked to produce a singular *um*, which was then inserted digitally within the inter-turn delays as follows: 600 ms (100 ms silence, *um* [400 ms], 100 ms silence) and 1200 ms (700 ms silence, *um* [400 ms], 100 ms silence).

Manipulation check.

To ensure that the naturalistic manipulation of final intonation of readback responses were perceived accordingly (i.e., ending in either rising or falling intonation), a manipulation check was undertaken. Participants comprised 23 undergraduate students from an Australian university. They were native English speakers, predominantly female (86.96%), and ranged in age from 18 to 46 years (M = 21.00, SD = 6.27). They were asked to listen to the readback portion of the communicative exchanges and decide whether the speaker used a rising or falling intonation at the end of his utterance.

Overall, participants listened and responded to, a total of 16 readback responses; eight readbacks derived from Communicative Exchange 1 (Four x Falling Intonation, Four x Rising Intonation) and eight readbacks were derived from Communicative Exchange 2 (Four x Falling Intonation, Four x Rising Intonation). The results indicated that participants were able

to correctly identify rising and falling intonation at the end of each response token with an accuracy rate of at least 90.22% (see Table 1 for detailed results).

Table 1

Accuracy Rate (%) in which Intonation was Correctly Identified Across the Two Communicative Exchanges

Communicative Exchange	Rising Intonation	Falling Intonation
1	93.48	97.83
2	94.57	90.22

Procedure

The participants completed the study in one-hour sessions, either during work-related training sessions (network operators) or during individual experimental sessions (Engineering students and graduates). All of the participants were briefed on the purpose of the study through printed and/or verbal means before electing to participate. After completing a brief demographic questionnaire, participants completed the Auditory Readback Task, followed by the online program EXPERTise 1.0 via a laptop computer. They were asked to wear headphones throughout the duration of the Auditory Readback task, which took approximately twenty minutes to complete. After completing the EXPERTise 1.0 battery of tasks, participants responded to a second questionnaire concerning their perceptions of the recently completed tasks, and complete self-assessments of fatigue, stress, workload and operational performance.

Data Reduction

The data reduction for the EXPERTise 1.0 tasks was consistent with the standard approach to the analysis of these data (e.g., Loveday, Wiggins, Harris, et al., 2013; Wiggins, Azar, et al., 2014). For the Paired Association Task, ratings of perceived association between feature-event pairs were recorded on a scale from one to six. The variance of participants' ratings of perceived association between feature-event pairs was then calculated (an approach similar to Weiss & Shanteau, 2003). As the Paired Association Task comprised two stages,

each participant received two variance scores for this task. For the Feature Discrimination Task, participants' ratings for the utility of each of the features available in the scenario were recorded on a 10-point Likert scale, from 1 (*not important at all*) to 10 (*extremely important*). These ratings were then aggregated into a single discrimination metric based on the variance of the individual's responses. As this task comprised two stages, each participant received two variance scores. For the Transition Task, the ratio was recorded of pairs of features that were selected in sequence, against the total number of pairs of features available. For the Auditory Readback Task, participants made a single rating of their assessment of the receiver's level of understanding of the instruction for each communicative exchange along a 7-point Likert scale. Data were then aggregated for each level of prosodic cue to create a mean rating of nonunderstanding for that variable level.

Results

Overview of Analyses

The primary aim of the current study was to examine whether two typologies of participants, categorised on the basis of operator performance across the EXPERTise 1.0 tasks, differed in the way that prosodic cues were used during assessments of nonunderstandings. The statistical analyses to address this aim comprised two stages. Using a cluster analysis, the first stage established whether typologies could be differentiated on the basis of network operators' performances across the tasks that comprised EXPERTise 1.0. Using planned contrasts, the second stage of the analyses examined whether typology membership was associated with the use of prosodic cues during assessments of nonunderstanding.

Stage 1: Cluster Analysis

A *K*-means cluster analysis was conducted to determine whether participants could be categorised into clear typologies representing distinct levels of visual cue utilisation performance consistent across the EXPERTise 1.0 tasks (Wiggins, Azar, et al., 2014). Before

the cluster analysis could be performed, *z*-scores were computed for the visual cue utilisation measures (namely, the variance in association ratings for both stages of the Paired Association Task, the variance in ratings of the utility of feature-event pairs for both stages of the Feature Discrimination Task and the ratio of sequential feature pairs accessed compared to the total feature pairs available in the Transition Task). Based on the range of years of experience of participants within power generation and the relatively small sample size, K = 2 clusters was predicted as the most likely fit for the data. Only cases that had a complete set of data were included in the cluster analysis. As a result, the ten participants that contained one or more missing value for EXPERTise 1.0 variables were excluded from all subsequent analyses.

Table 2 summarises the results of the cluster analysis, including the mean centroid for each cluster on each of the variables that comprise EXPERTise 1.0. The cluster analysis yielded two distinct typologies that represented relatively greater and lesser levels of cue utilisation. Cluster 1 contained 13 participants who recorded relatively greater variance across both stages of the Paired Association Task, greater variance in the utility of ratings of features in both stages of the Feature Discrimination Task and a relatively lower ratio of sequential pairs of features accessed in the Transition Task. Overall, this pattern of performance is reflective of participants who possess a relatively greater level of cue utilisation while assessing and responding to domain-relevant tasks.

The remaining 22 participants comprised the second cluster, exhibiting lower variance scores in the Paired Association Tasks and Feature Discrimination Tasks, as well as a relatively greater ratio in response to the Transition Task. This performance is consistent with a lesser level of cue utilisation. These two typologies formed the basis of subsequent analyses.

Table 2

Participant Cluster Means for the EXPERTise 1.0 Measures Across the Two Cue Utilisation Typologies

	Typology	
	Greater (*N*=13)	Lesser (*N*=22)
Paired Association Task (Stages 1) – ratings variance	1.08	-0.64
Paired Association Task (Stages 2) – ratings variance	0.85	-0.55
Feature Discrimination Task (Stage 1) – ratings variance	0.69	-0.38
Feature Discrimination Task (Stage 2) – ratings variance	0.74	-0.29
Transition Task – ratio of sequential pairs	-0.14	0.24

Stage 2: Analyses of Hypotheses

Arguably, there could be a potentially confounding relationship between levels of visual and prosodic cue utilisation and years of experience within the domain of electricity transmission control. To account for this possible relationship, years of experience within electricity transmission control was included as a covariate in each analysis.

Prior to testing the hypotheses, it was first important to confirm that the semantic variation between the two communicative exchanges did not influence prosodic cue use. A null finding would support the decision to collapse the data across the two communicative exchanges in subsequent analyses. A series of planned repeated measures analyses of variances (ANOVAs) were conducted to investigate this relationship, with mean ratings of understanding as the dependent variable. To control for family-wise error rate, critical alpha was set at $\alpha = .01$.

Controlling for years of experience within electricity transmission control, a one-way repeated measures ANOVA was conducted for the variable 'Communicative Exchange', with no main effect evident, $F(1, 33) = 1.54$, $p = .224$, $\eta_p^2 = .05$. Further, there were no significant interactions evident between Communicative Exchange and Intonation, Communicative Exchange and Inter-turn Delay or Communicative Exchange and Filler ($p > .01$). These results suggest that the semantic content of each communicative exchange did not influence

ratings of understanding, nor was the semantic content of each exchange related to the use of prosodic cues during assessments of nonunderstanding. As a result, the variable Communicative Exchange was collapsed in subsequent analyses.

Typologies and prosodic cue use.

To examine H1, a two-way mixed design ANOVA was run for Typology (Greater levels of cue utilisation, Lesser levels of cue utilisation) and Intonation (Rising, Falling), with years of experience within high-voltage electricity transmission control as a covariate. There was no significant interaction between Typology and Intonation, $F(1, 32) = 0.12$, $p = .730$, $\eta_p^2 = .00$. This suggested that there was no difference in the way that the two typologies responded to readbacks that ended in rising intonation compared to readbacks that ended in falling intonation (see Figure 1).

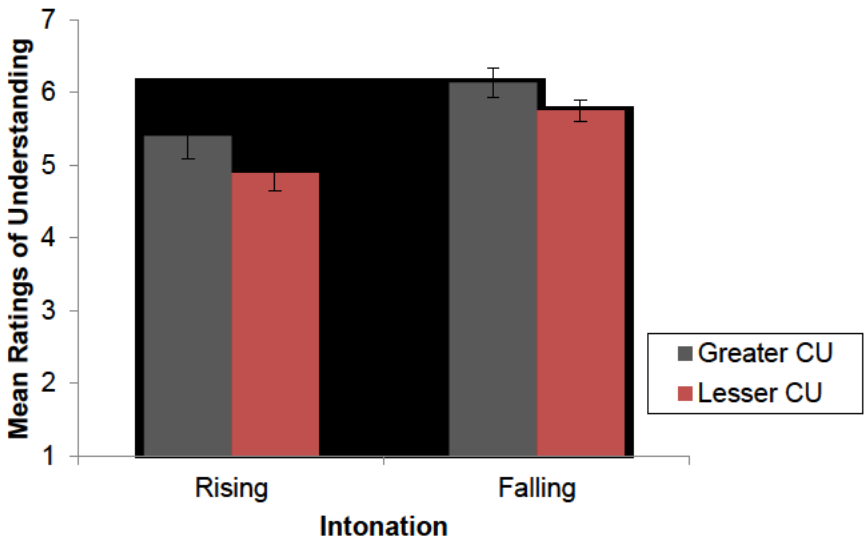


Figure 1. Non-significant two-way interaction between typology and intonation on mean ratings of understanding. Greater CU = Greater levels of cue utilisation; Lesser CU = Lesser levels of cue utilisation. Error bars represent standard error. Value of covariate (years in electricity transmission control) = 18.63 years.

To examine H2, a two-way mixed design ANOVA was run for Typology (Greater levels of cue utilisation, Lesser levels of cue utilisation) and Inter-turn Delay (Shorter, Longer), with years of experience within high-voltage electricity transmission control included as a covariate. There was no significant interaction between Typology and Inter-turn Delay, $F(1, 32) = 0.48$, $p = .493$, $\eta_p^2 = .02$. Participants within both typologies responded to readbacks that preceded a shorter or longer inter-turn delay in a similar fashion (see Figure 2).

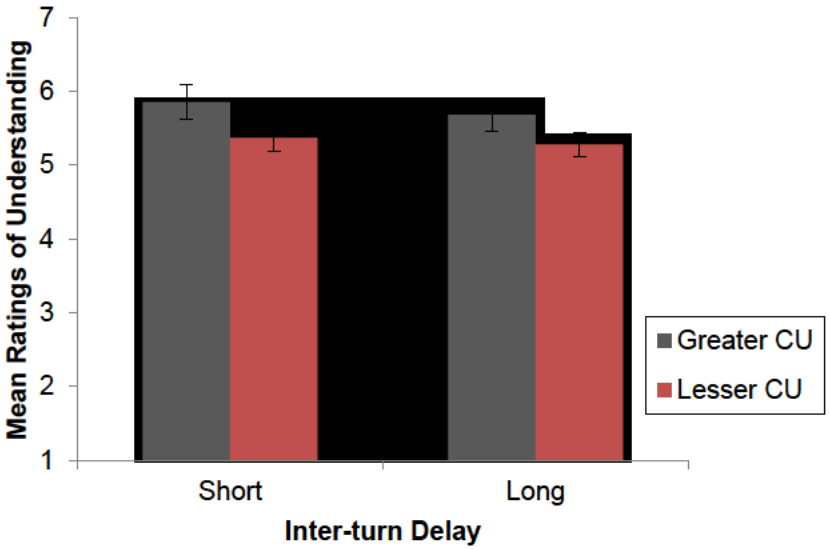


Figure 2. Non-significant two-way interaction between typology and inter-turn delay on mean ratings of understanding. Greater CU = Greater levels of cue utilisation; Lesser CU = Lesser levels of cue utilisation. Error bars represent standard error. Value of covariate (years in electricity transmission control) = 18.63 years.

To examine H3, a two-way mixed design ANOVA was run for Typology (Greater levels of cue utilisation, Lesser levels of cue utilisation) and Filler (Present, Absent), with years of experience within high-voltage electricity transmission control as a covariate. There was no significant interaction between Typology and Filler, $F(1, 32) = 0.11$, $p = .740$, $\eta_p^2 = .00$. Readbacks that followed an inter-turn delay containing a filler compared to readbacks that followed an inter-turn delay containing only silence, were rated in a similar manner by both typologies (see Figure 3). Overall, there was no support provided for H1, H2 nor H3.

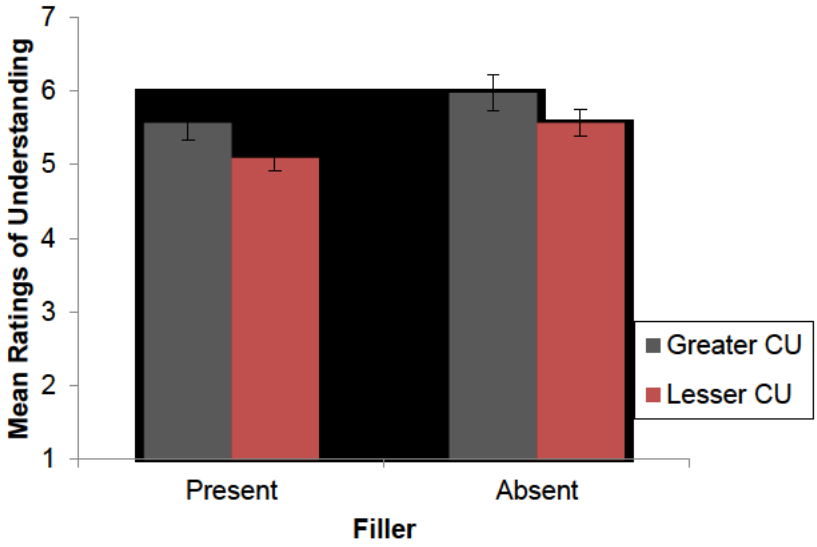


Figure 3. Non-significant two-way interaction between typology and filler on mean ratings of understanding. Greater CU = Greater levels of cue utilisation; Lesser CU = Lesser levels of cue utilisation. Error bars represent standard error. Value of covariate (years in electricity transmission control) = 18.63 years.

Discussion

The overall aim of the current study was to examine whether an association existed between levels of visual and auditory cue utilisation. The current study included three hypotheses that examined visual cue utilisation typologies and prosodic cue use during assessments of nonunderstanding. In comparison to participants with lesser levels of cue utilisation, it was predicted that participants with greater levels of cue utilisation would perceive a greater difference in levels of understanding on the part of the receiver between: readback responses ending in rising intonation and readback responses ending in falling intonation (H1); readback responses preceded by a shorter compared to a longer inter-turn delay (H2), and; readback responses preceded by a filler compared to an unfilled inter-turn delay of the same duration (H3). The results failed to provide any support for H1, H2 or H3. Specifically, there was no difference between the typologies in the way that different levels of intonation, inter-turn delay and filler were rated during assessments of nonunderstanding.

There are two possible explanations for these findings. Firstly, it could be the case that operators who have a relatively superior ability for visual cue utilisation during situation assessment within operational environments, are no more skilled at differentiating prosodic

cues that are diagnostic of nonunderstandings in comparison to their less skilled counterparts, as experience in detecting nonunderstandings through prosodic cues is not unique to the domain of electricity transmission control.

Theoretically, the participants in the current study who were identified as possessing a relatively greater level of visual cue utilisation, were assumed to have established visual feature-event associations in LTM through repeated exposure to the electricity transmission domain (Ericsson & Kintsch, 1995). However, the use of intonation, inter-turn delay and filler as prosodic cues, is likely to represent a generalised skill that is not related to a specific domain. For example, a mundane telephone exchange between two individuals represents a similar context to the readback/hearback protocol, where information is exchanged for mutual understanding without the aid of visual information. Therefore, irrespective of their relative level of visual cue utilisation, network operators within the current study may have acquired the skills necessary to detect nonunderstandings through everyday interactions.

An alternative explanation is that the Auditory Readback Task did not facilitate the use of prosodic cue associations. Cues are expected to represent relatively automatic associations between a feature and an event/object that can be activated even when cognitive resources are limited (Wiggins, 2012). As the Auditory Readback Task simply required participants to listen to a readback/hearback communicative exchange, and then rate the perceived level of understanding of the receiver, theoretically, participants had ample cognitive resources available to consider the significance of the prosodic features in relation to nonunderstandings, even if such associations were not already established in memory. It is possible therefore, that the design of this task did not discriminate those participants who possessed cue associations between prosodic features and levels of understanding (event) and those who adopted a more analytical approach when making judgements of perceived nonunderstanding.

Consistent with this proposition, differences between experts and novices in the visual domain tend become most evident when restrictions are placed on information processing.

For example, Calderwood and colleagues (1988) compared the performance of highly skilled (Masters) and less skilled (Class B) chess players during two game types: regulation and blitz. The blitz game was considered a more cognitively demanding task, as it required players to make a move every six seconds, compared to every 2.25 minutes during the regulation game. While there was no difference in the quality of performance between Masters and Class B players across the regulation game, differences did become evident during the blitz game where time constraints were imposed.

This outcome suggests that the introduction of time pressure during the blitz game did not affect the performance of Masters as they were arguably relying on rapid recognitional processes. In contrast, Class B players performed relatively worse under such time constraints as their analytical approach to situation assessment is carried out over time. The lack of time pressure or strain to cognitive resources imposed in the Auditory Readback Task could potentially explain why no differences were observed in the way the two typologies rated the prosodic cues during assessments of nonunderstanding.

Limitations and Future Research

The main methodological limitation associated with the current study was the way in which prosodic cue use was assessed during the Auditory Readback Task. When engaged in a communicative exchange over the radio/telephone, operators tend to be taxed with additional tasks (such as note taking, or scanning visual screens for information). Such tasks limit the cognitive resources that can be dedicated to attending to prosodic features that might be indicative of nonunderstandings. Therefore, future research needs to consider the imposition of a secondary task that might impose a cognitive load on operators, thereby reducing the cognitive resources available to attend to prosodic cues. In doing so, any differences between greater and lesser levels of task-related cue utilisation would be expected to become more readily apparent.

Conclusion

Overall, the results from this study suggest that differences in visual cue utilisation are not necessarily related to differences in the use of prosodic cues during perceptions of nonunderstanding. Rather, the interpretation of prosodic cues appears more likely to constitute a generalizable ability that emerges irrespective of exposure to electricity control environments. Nevertheless, the lack of cognitive load associated with the interpretation of prosodic cues in the present experiment is such that the proposition remains inconclusive.

Paper 2 References

Abernethy, B. (1990). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sports Sciences*, *8*, 17–34. doi:10.1080/02640419008732128

Air Accidents Investigation Branch. (2008). *Boeing 737-500, SP-LKA* (Report No. EW/C2007/06/02). Retrieved from http://www.aaib.gov.uk/publications/bulletins/june_2008/boeing_737_500__sp_lka.cfm

Audacity (1.2.5) [Free Audio Editor and Recorder Software]. Retrieved from http://audacity.sourceforge.net/

Auton, J. C., Wiggins, M. W., & Loveday, T. (2012). Prosodic cues that signal non-understandings to power control operators during radio communication. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*, 483–487. doi:10.1177/1071181312561049

Auton, J. C., Wiggins, M. W., Searle, B. J., Loveday, T., & Xu Rattanasone, N. (2013). Prosodic cues used during perceptions of nonunderstandings in radio communication. *Journal of Communication*, *63*, 600–616. doi:10.1111/jcom.12033

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, *106*, 410–433. doi:10.1037/0033-2909.106.3.410

Bisantz, A. M., & Pritchett, A. R. (2003). Measuring the fit between human judgments and automated alerting algorithms: A study of collision detection. *Human Factors*, *45*, 266–280. doi:10.1518/hfes.45.2.266.27243

Blignaut, C. J. H. (1979). The perception of hazard: I. Hazard analysis and the contribution of visual search to hazard perception. *Ergonomics*, *22*, 991–999. doi:10.1080/00140137908924673

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398. doi:10.1006/jmla.1995.1017

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217. doi:10.1037/h0047470

Calderwood, R., Klein, G. A., & Crandall, B. W. (1988). Time pressure, skill, and move quality in chess. *The American Journal of Psychology*, *101*, 481–493. doi:10.2307/1423226

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education*, *37*, 695–703. doi:10.1046/j.1365-2923.2003.01577.x

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine*, *84*, 1022–1028. doi:10.1097/ACM.0b013e3181ace703

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

De Groot, A. D. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

Ebbesen, E. B., & Konečni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology*, *32*, 805–821. doi:10.1037/0022-3514.32.5.805

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory, *102*(2), 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406. doi:10.1037/0033-295X.100.3.363

Fegyveresi, A. E. (1997). *Vocal cues and pilot/ATC communication*. Paper presented at the Ninth International Symposium on Aviation Psychology, Columbus, OH. Abstract retreived from

http://md1.csa.com/partners/viewrecord.php?requester=gs&collection=TRD&recid=A9828885AH&q=FEGYVERESI+1997&uid=1000599&setcookie=yes

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327–358. doi:10.1037/h0061470

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language and Speech*, *4*, 18–26. doi:10.1177/002383096100400102

Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. (1994). Repairing conversational misunderstandings and non-understandings. *Speech Communication, 15*, 213–229. doi:10.1016/0167-6393(94)90073-6

Hoffman, K. A., Aitken, L. M., & Duffield, C. (2009). A comparison of novice and expert nurses' cue collection during clinical decision-making: Verbal protocol analysis. *International Journal of Nursing Studies, 46*, 1335–1344. doi:10.1016/j.ijnurstu.2009.04.001

Hunter, M. (2004). The music is the message: Prosody in aviation discourse. In M. A. Turney (Ed.), *Tapping Diverse Talent in Aviation: Culture, gender, and diversity* (pp. 129–151). Hampshire, England: Ashgate Publishing Ltd.

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*, 146–154. doi:10.1016/j.learninstruc.2009.02.019

Jones, R. K. (2003). Miscommunication between pilots and air traffic control. *Language Problems and Language Planning, 27*, 233–248. doi:10.1075/lplp.27.3.03jon

Kaempf, G. L., Klein, G. A., Thordsen, Marvin, L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors, 38*, 220–231. doi:10.1518/001872096779047986

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*, 404–426. doi:10.1037/0033-2909.134.3.404

Klein, G. A. (1989). Recognition-primed decisions (RPD). *Advances in Man-Machine Systems*, *5*, 47–92.

Klein, G. A. (1993). A Recognition-Primed Decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138–147). Norwood, New Jersey: Ablex Publishing Company.

Klein, G. A. (1997). The Recognition-Primed Decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 285–292). Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G. A. (2003). *Intuition at work.* New York: Currency Books.

Klein, G. A., & Calderwood, R. (1991). Decision models: Some lessons from the field. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*, 1018–1026. doi:10.1109/21.120054

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. *Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting* (Vol. 1, pp. 576–580). Norwood, NJ: Ablex.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*, 462–472. doi:10.1109/21.31053

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Focus article: Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*, 331–352. doi:10.1002/bdm.381

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2013). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*, 90–107. doi:10.1177/0018720812450911

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2013). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*, 125–137. doi:10.1177/0018720812448475

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2012). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making*, *7*, 211–226. doi:10.1177/1555343412459192

Morrow, D., Rodvold, M., & Lee, A. (1994). Nonroutine transactions in controller- pilot communication. *Discourse Processes*, *17*, 235–258. doi:10.1080/01638539409544868

Müller, S., Abernethy, B., & Farrow, D. (2006). How do world-class cricket batsmen anticipate a bowler's intention? *The Quarterly Journal of Experimental Psychology*, *59*, 2162–2186. doi:10.1080/02643290600576595

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). *Straight choices: The psychology of decision making*. New York, NY: Psychology Press.

North, J. S., Williams, A. M., Hodges, N., Ward, P., & Ericsson, K. A. (2009). Perceiving patterns in dynamic action sequences: Investigating the processes underpinning stimulus recognition and anticipation skill. *Applied Cognitive Psychology*, *23*, 878–894. doi:10.1002/acp.1581

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1017–1030. doi:10.1037/0096-1523.34.4.1017

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology). Retreived from http://hdl.handle.net/1721.1/16065

Prinzo, O. V., & Britton, T. W. (1993). *ATC/pilot voice communications: A survey of the literature* (Report No. DOT/FAA/AM-93/20). Washington, DC: Office of Aviation Medicine.

Randel, J. M., Pugh, H. L., & Reed, S. K. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human-Computer Studies*, *45*, 579–597. doi:10.1006/ijhc.1996.0068

Rasmussen, J. (1993). Diagnostic reasoning in action. *IEEE Transactions on Systems Man and Cybernetics*, *23*, 981–992. doi:10.1109/21.247883

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication*, *48*, 1079–1093. doi:10.1016/j.specom.2006.02.001

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise difference in attentional strategies related to pilot decision making. *Human Factors*, *50*, 864–878. doi:10.1518/001872008X374974.

Scialfa, C. T., Deschênes, M. C., Ference, J., Boone, J., Horswill, M. S., & Wetton, M. (2011). A hazard perception test for novice drivers. *Accident Analysis and Prevention*, *43*, 204–8. doi:10.1016/j.aap.2010.08.010

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*, 253–263. doi:10.1016/S0377-2217(01)00113-8

Shim, J., Carlton, L. G., Chow, J. W., & Chae, W.-S. (2005). The use of anticipatory visual cues by highly skilled tennis players. *Journal of Motor Behavior*, *37*, 164–175. doi:10.3200/JMBR.37.2.164-175

Simon, H. A. (1992). What is an “explanation” of behavior? *Psychological Science*, *3*, 150–161. doi:10.1111/j.1467-9280.1992.tb00017.x

Skantze, G. (2005). Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication*, *45*, 325–341. doi:10.1016/j.specom.2004.11.005

Weigand, E. (1999). Misunderstanding: The standard case. *Journal of Pragmatics*, *31*, 763–785. doi:10.1016/S0378-2166(98)00068-X

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors*, *45*, 104–114. doi:10.1518/hfes.45.1.104.27233

Wellens, A. R. (1993). Group situation awareness and distributed decision making: From military to civilian applications. In J. Castellan (Ed.), *Individual and Group Decision Making: Current Issues* (pp. 267–291). Hillsdae, NJ: Lawrence Erlbaum Associates.

Wickens, C. D., & Flach, J. M. (1988). Information processing. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in Aviation* (pp. 111–155). San Diego, CA: Academic Press.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Wiggins, M. W. (2006). Cue-based processing and human performance. In W. Karwowski (Ed.), *Encyclopedia of Ergonomics and Human Factors* (2nd ed., pp. 641–645). London, UK: Taylor and Francis.

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science*, 1–10. doi:10.1080/1463922X.2012.724725

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots’ pre-flight and in-flight weather decision-making. *Safety Science*, *65*, 118–124. doi:10.1016/j.ssci.2014.01.006

Wiggins, M. W., Brouwers, S., Davies, J., & Loveday, T. (2014). Trait-based cue utilization and initial skill acquisition: Implications for models of the progression to expertise. *Frontiers in Psychology*, *5*, 1–8. doi:10.3389/fpsyg.2014.00541

Wiggins, M. W., Harris, J., Loveday, T., & O'Hare, D. (2010). EXPERTise (Software Package). Sydney, Australia: Macquarie University.

Wiggins, M. W., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied*, *1*, 305–320. doi:10.1037/1076-898X.1.4.305

Wiggins, M. W., & O'Hare, D. (2003). Expert and novice pilot perceptions of static in-flight images of weather. *The International Journal of Aviation Psychology*, *13*, 173–187. doi:10.1207/S15327108IJAP1302_05

Wigton, R. S., Hoellerich, V. L., & Patil, K. D. (1986). How physicians use clinical information in diagnosing pulmonary embolism: An application of conjoint analysis. *Medical Decision Making*, *6*, 2–11. doi:10.1177/0272989X8600600102

Paper 2 Appendix A

Paired Association Task Items

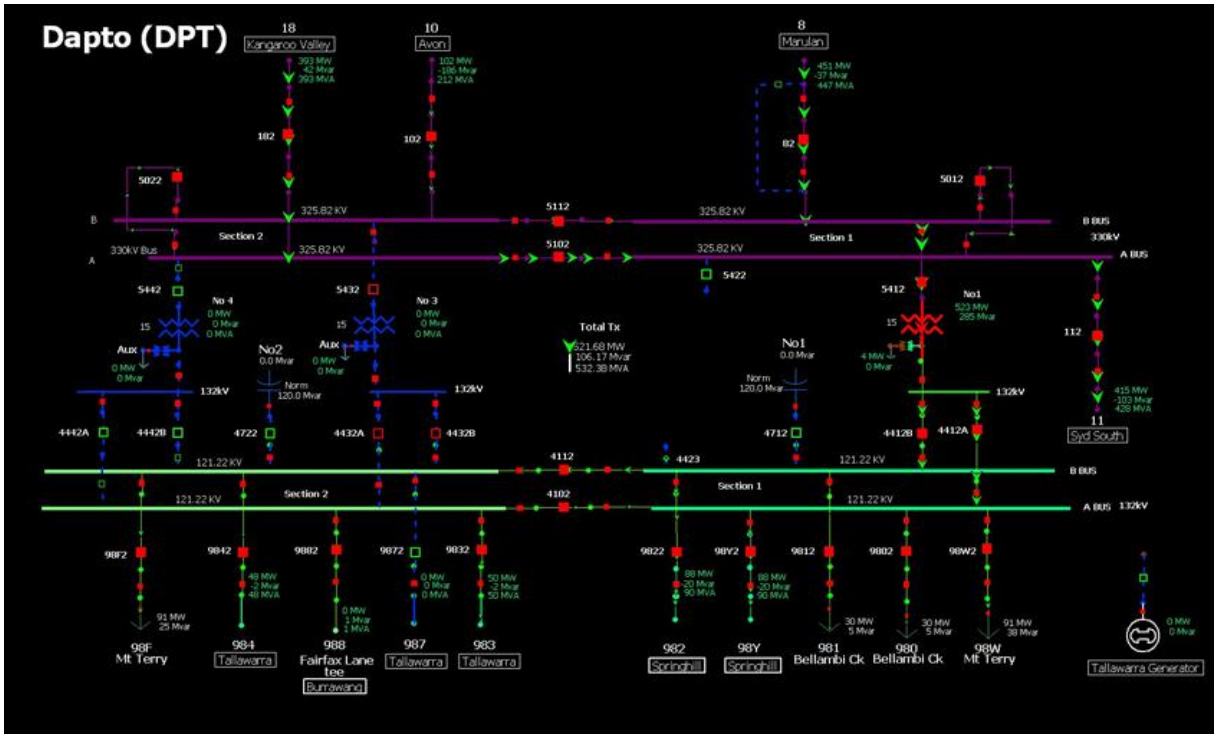
Sequential Pairs		Simultaneous Pairs	
Feature	Event	Feature	Event
Flashover	Transmission Line Trip	Main Inlet Valve	High Temperature Alarm
Energise	Current	High Load	Voltage Depression
High Frequency	Voltage Depression	Transmission Line Trip	Bush Fire
High load	Warm Spring Day	Low Frequency	Increased Load
Voltage Depression	Fault	High Load	Mild Autumn Day
Transmission Line Trip	Ice Storm	Load Increase	High Frequency
Voltage Depression	Warm Spring Day	High Frequency	Loss of Load
High Frequency	High Wind	Voltage Depression	Loss of Generation
Reclose	Bad Weather	Energise	Load Circuit Breaker
Energise	Circuit Breaker Close	Load Increase	Voltage Depression
Transmission Line Trip	Loss of Bus Bar	High Load	Cold Winter Day
High Load	Lost Generation	Voltage Depression	Reactor
Load Increase	Trip Generation	High Frequency	Overspeed of Generator
High Load	Loss of Load	Low Frequency	Reactor
Transmission Line Trip	Dust Storm	High Load	Hot Summer Day
Load Increase	Loss of Line	Spike on the System	Circuit breaker Opening
Transmission Line Trip	High Wind	Voltage Depression	Loss of Interconnector
Voltage Depression	Reclose	High Frequency	Circuit Breaker Close
Load Increase	Bushfire	Low Frequency	Capacitor
Transmission Line Trip	Capacitor	Energise	Loss of Load Trip
Load Increase	Reactor	Load Increase	Uncleared Fault
Low Frequency	Circuit Breaker Close	Load Increase	Loss of Interconnector
Load Increase	Voltage Decrease	Transmission Line Trip	Reactor
Voltage Depression	Peak Demand	Energise	High Load
Transmission Line Trip	Lightning	High Frequency	Open Circuit Breaker
Low Frequency	Loss of Generator	Load Increase	Low Frequency
Voltage Depression	Trip of Line	Transmission Line Trip	Fault
High Frequency	Loss of Generator	Energise	Volts
Voltage Depression	Lightning	Voltage Depression	Low Frequency

Paper 2 Appendix B

Feature Discrimination Task (1)

Scenario

The SCADA screen capture shows a possible load shedding situation. The temperature is 35 degrees and the time is 1400 hours. The transformers are 375 MVA and loaded at 533.5 MVA. Please press Continue when you have determined how you would best manage this situation.



Response Options

From the options available, determine how you would best manage this situation. You may only select one response. If you need to review the scenario again, please click Back. Press Continue when you have made your selection.

- Ask DNSP to load shed
- Load shed at 330
- Load shed at 132
- Do nothing

Ratings of Feature Utility

Rate the relevance of the following features in informing your response.

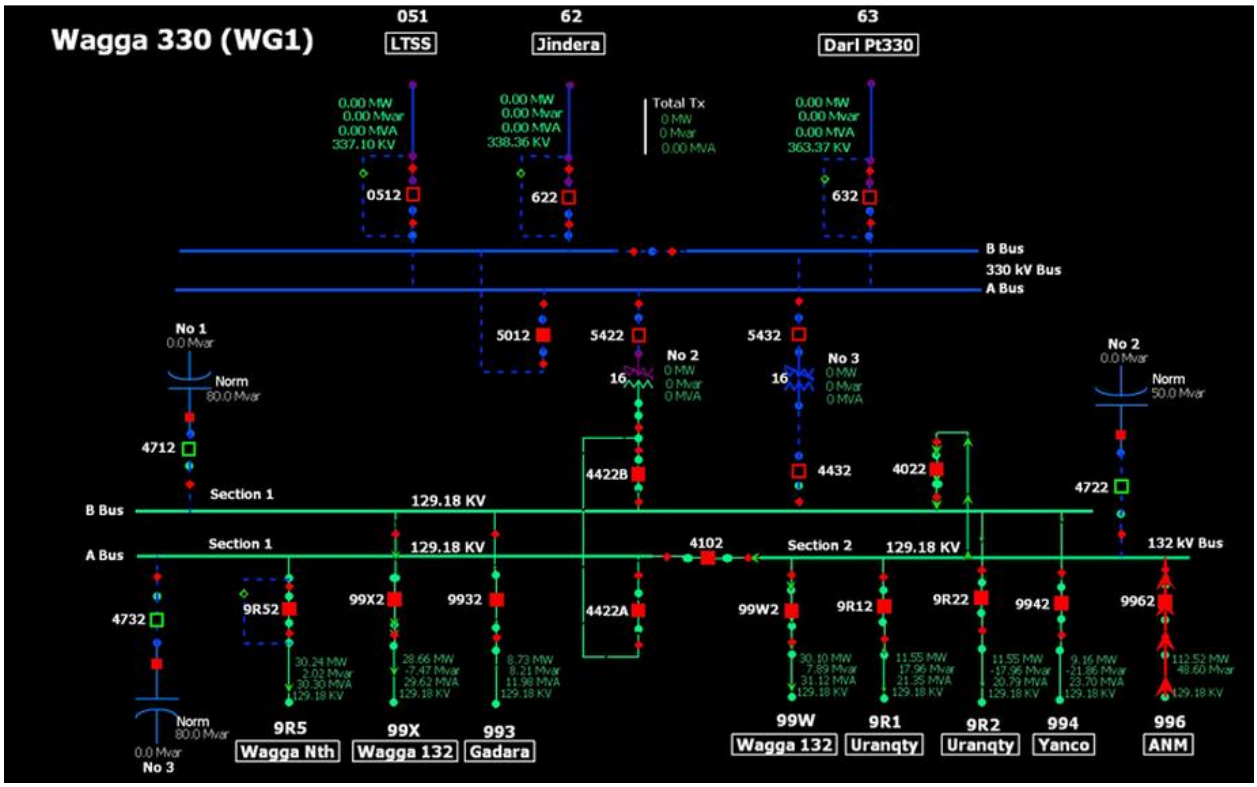
- Time of day
- Temperature
- Over speed of generator
- Reactor
- Capacitor
- Number of people affected
- Amount of load required for shedding
- Transformer rating
- Thermal rating
- Overload
- Pre-condition
- Post-condition

Paper 2 Appendix C

Feature Discrimination Task (2)

Scenario

The SCADA screen capture shows Wagga 330KV Substation, which is a bulk supply point in a 330KV Meshed system. There are no Radial lines on the 132KV BUS. The temperature is 40 degrees and the time is 1800 hours. Please press Continue when you have determined how you would best manage this situation.



Response Options

From the options available, determine how you would best manage this situation. You may only select one response. If you need to review the scenario again, please click Back. Press Continue when you have made your selection.

- Ask customer to load shed
- Open 996 transmission line
- Close No 2 transformer 330KV CB
- Do a power flow and then make a decision

- Do nothing

Ratings of Feature Utility

Rate the relevance of the following features in informing your response.

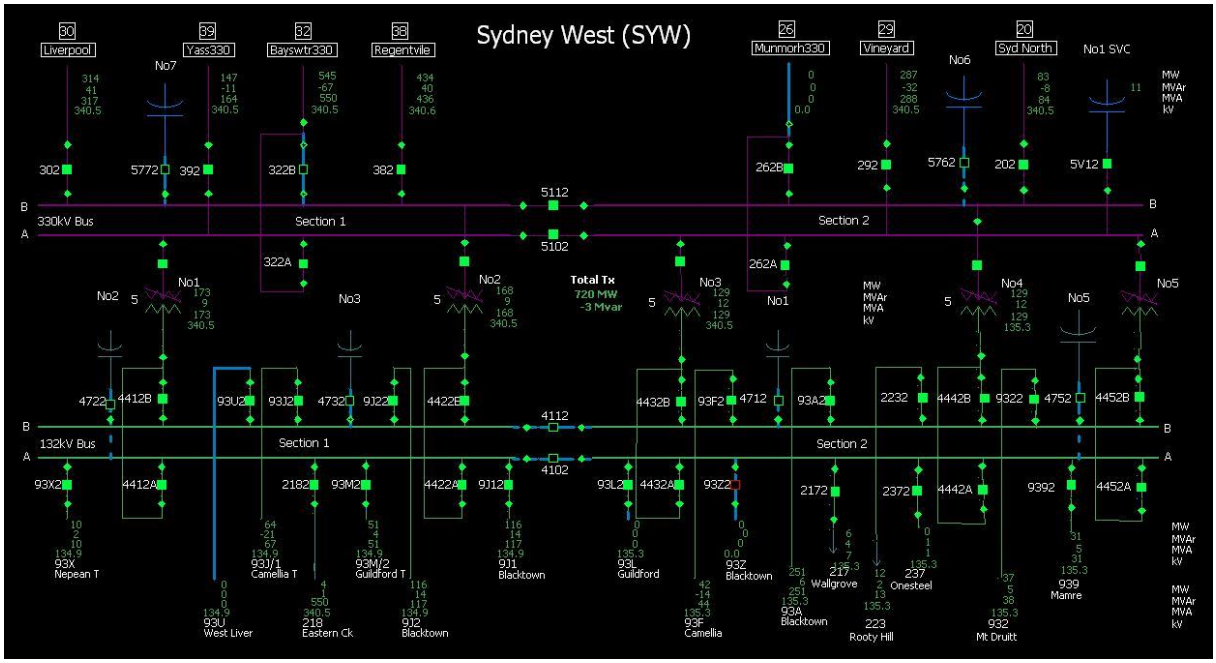
- Time of day
- Temperature
- Over speed of generator
- Reactor
- Capacitor
- Number of people affected
- Amount of load required for shedding
- Transformer rating
- Thermal rating
- Overload
- Pre-condition
- Post-condition

Paper 2 Appendix D

Transition Task Scenario

Scenario

You have just received a report of children climbing above man proofing on a transmission tower. Use the relevant SCADA screen and the background information to determine your response.



Information Categories

- Street location
- Caller
- Tower Number
- Line Number
- Side of Tower
- Time of Day
- Temperature
- Time of Year
- Forecast
- Day of Week

- Wind Direction
- Estimated time for the TransGrid Officer to arrive
- Tower Location

Response Options

- De-energize line a
- De-energize line b
- De-energize line a and line b
- Send a TransGrid officer out
- Do nothing

Paper 2 Appendix E

Semantic Content of Communicative Exchanges

Communicative Exchange	Instruction	Readback
1	On X three zero zero nine four slash one, zero one X to X three thirty kV transmission line, I give you permission to carry out steps X three nine to X four one. That is removal of earths at X.	Yes I understand. X three zero zero nine four slash one that I have permission to carry out steps X three nine to X four one, on zero one X to X three thirty kV transmission line. That is removal of earths at X.
2	On X three zero one seven six slash two, X to X power station three thirty kV transmission line, I give you clearance to carry out steps X five to X one six. That is isolate only at X switching station.	Yes I understand. X three zero one seven six slash two, that I have clearance to carry out steps X five to X one six on X to X power station three thirty kV transmission line. That is isolate only at X switching station.

Note. Locations were removed and replaced with "X" to maintain the confidentiality of participants.

Paper 3: Relationship between Visual and Auditory Cue Utilisation (Extended)

Paper 3 Aims

The outcomes of Paper 2 indicated that there was no relationship between visual cue utilisation performance and the use of prosodic cues during assessments of nonunderstanding. However, it was hypothesised that this null relationship might be accounted for by one of two explanations. Firstly, it may be the case that the use of intonation and filler to assess a speaker's state of understanding is a generic skill that is independent of domain-relevant experience. Secondly, the null relationship was potentially a product of the way prosodic cue utilisation was measured. Specifically, Paper 2 postulated that the Auditory Readback Task, in the absence of a concurrent task, may have been unable to discriminate those participants who had pre-established prosodic cue associations in long-term memory from those who did not possess such associations.

The aim of Paper 3 was two-fold. Firstly, it sought to establish whether the prosodic cues found to be useful in detecting nonunderstandings within the domain of electricity transmission control (Paper 1) have utility in a different operational context. Secondly, Paper 3 sought to test the two explanations for the null outcomes of Paper 2. To address the first explanation, two groups of participants were recruited: those with experience within the operational context and those with no prior experience within the operational context. If skilled prosodic cue utilisation is dependent on domain-relevant experience, a clear difference should be apparent between the two cohorts of participants.

To address the second explanation, Paper 3 incorporated a concurrent task to be completed in conjunction with the Auditory Readback Task. Referred to as the Numeric Digit Span Recall Task, it was designed to impose a cognitive demand on participants, thereby restricting the cognitive resources available during assessments of nonunderstanding. Although succinctly described in Paper 3, a more detailed description of the task and the manipulation check is provided in Appendix C of the thesis.

Paper 3 was conducted within the domain of electrical rail control, which involves the supply of traction energy to trains without local fuel supply. Within this context, ‘electrical control operators’ were selected as the experienced participants. These operators were selected for two reasons including that their primary job function is to monitor the distribution network to ensure efficient operation, and identify and locate, supply faults through Supervisory Control and Data Acquisition (SCADA) visual interfaces and locate irregularities and potential system and operational limitations. Consistent with the network operators recruited in Papers 1 and 2, it should be possible to differentiate experienced and inexperienced participants (undergraduate psychology students) on the basis of visual cue utilisation performance using a series of naturalistic tasks involving the application of visual cues during situation assessment. Secondly, electrical control operators are responsible for communicating clearly and effectively to other relevant parties (such as field supervisors, train controllers and external authorities) to ensure the safety of the high-voltage traction power supply. A large proportion of this communication is undertaken over the telephone or radio, where it is expected that operators follow the readback/hearback protocol during instructional exchanges. Therefore, if more nuanced associations exist between prosodic features and nonunderstandings for operators with greater domain-specific experience, this should become apparent.

Paper 3: Publication History

Due to the outcomes of Paper 3, it has not been submitted for publication. The author of the current thesis wrote approximately 85% of Paper 3.

Abstract

There has been considerable empirical support for the notion that experienced operators can be distinguished from their less experienced counterparts by their relatively greater capacity to utilise environmental cues to facilitate the process of situation assessment in operational environments. A major limitation of this research however, has been the focus on the application of *visual* cues, rather than an examination of how cues from other modalities (such as tactile, auditory and olfactory cues) assist operators in understanding emerging situations. Conducted within the context of electrical rail control, the current study had two aims; (1) to assess whether specific prosodic cues (namely, intonation, inter-turn delay, and filler) are associated with assessments of nonunderstanding during readback responses, and; (2) to examine the extent to which a relationship exists between visual cue utilisation and the use of these prosodic cues during perceptions of nonunderstanding. The results indicated that while intonation and filler as prosodic cues are associated with assessments of nonunderstanding during readback responses, this appears to be a generic skill, independent of domain experience. Limitations and future research directions are discussed.

The Influence of Domain Experience on the Utilisation of Prosodic Cues during Assessments of Nonunderstanding of Readback Responses

The safety and performance of many high-risk operational environments, including medicine, aviation and electrical rail control, relies heavily on operators' skills to rapidly understand and identify the nature of a developing situation (Klein, 1989; Randel, Pugh, & Reed, 1996; Rasmussen, 1993; Wellens, 1993). This process is often referred to as *situation assessment* (Klein & Calderwood, 1991). An accurate situation assessment typically guides experienced operators to the most effective response for that situation (Kaempf, Klein, Thordsen, Marvin, & Wolf, 1996; Randel et al., 1996; Wellens, 1993).

Klein's (1993, 1997) Recognitional-Primed Decision (RPD) model postulates that situation assessment occurs in response to environmental cues. Cues are considered to be associations resident in memory between a feature(s) and an event/object (Lipshitz, Klein, Orasanu, & Salas, 2001; Wiggins, 2012). Features can include any piece of environmental stimuli that is visual, auditory, tactile or olfactory in nature (Wiggins, 2006). Cue associations are formed in memory when a specific feature (or a pattern of features) is consistently experienced with a specific event, such that the presence of a feature in an environment holds meaning for an operator (Lipshitz et al., 2001; Wiggins & O'Hare, 2003a). As a product of extensive experience within a particular domain, an operator is expected to possess a repertoire of cue associations that pertain to different situations (Coderre, Mandin, Harasym, & Fick, 2003; Croskerry, 2009; Ericsson & Kintsch, 1995; Ericsson, Krampe, & Tesch-Römer, 1993; Klein, 1993).

Klein's (1993, 1997) RPD model describes situation assessment as a process of *pattern recognition* through the utilisation of environment cues (Kaempf et al., 1996; Klein, 1989, 1997, 1999). Pattern recognition describes how an operator is expected to 'match' a pattern of cues (or less frequently, an isolated cue; Ebbesen & Konečni, 1975) present in an emerging situation to a template or a similar pattern of cues encountered in a past or prototypical event

retrieved from Long-Term Memory (LTM; Coderre et al., 2003; Kaempf et al., 1996; Wickens & Flach, 1988). This match is expected to activate an *action script* which embodies 'typical' ways of responding that have been developed and refined through extensive experience in a domain (Anderson, 1983; Klein, 1989, 2003). As a result, an experienced operator is expected to implement intuitively, the course of action that has proven successful in the past, requiring the expenditure of minimal cognitive resources and allowing for timely intervention (Kaempf et al., 1996; Lipshitz et al., 2001; Sweller, 1988; Wiggins, 2012). Therefore, the RPD model predicts that experienced operators who possess strong cue associations should be more efficient and more accurate in recognising a pattern of cues indicative of an event, compared to their less experienced counterparts.

Brunswik's (1955) Lens Model also explains how cognitive inferences are made through a process of cue utilisation. This model is based on the proposition that an operator will make a cognitive assessment about the state of a situation by attending to many different pieces of information (cues) available in an environment (Wigton, Hoellerich, & Patil, 1986). As a result of past experiences, different features will have a different contribution to the final assessment for each operator (Balzer, Doherty, & O'Connor, 1989; Bisantz & Pritchett, 2003). Therefore, the Lens Model describes situation assessment as a process whereby operators rely on a range of idiosyncratic cues in the environment to derive meaning from an uncertain situation.

Although the Klein's (1993, 1997) RPD model and Brunswik's (1955) Lens Model are distinct, they both postulate that situation assessment occurs through a rapid process of pattern recognition of associations between features and events. The utility of pattern recognition during situation assessment has been examined extensively using the expert-novice paradigm, in which the performance of experts is compared to the performance of novices (Coderre et al., 2003; de Groot, 1965; Litchfield, Ball, Donovan, Manning, & Crawford, 2010; Norman, Young, & Brooks, 2007; North, Williams, Hodges, Ward, &

Ericsson, 2009; Wiggins & O'Hare, 1995; Williams, Hodges, North, & Barton, 2006). This research has demonstrated that experts differ from novices across a range of variables, including the nature of features extracted, the sequence of feature acquisition, and the time spent examining features (Abernethy, 1990; Blignaut, 1979; Coderre et al., 2003; Müller, Abernethy, & Farrow, 2006; North et al., 2009; Wiggins & O'Hare, 2003a). Moreover, these differences between experts and novices have been demonstrated in a range of contexts, including aviation (Schriver, Morrow, Wickens, & Talleur, 2008; Stokes, Kemper, & Marsh, 1992; Wiggins & O'Hare, 2003b), chess (Chase & Simon, 1973; de Groot, 1965), sport (Abernethy, 1990; Gray, 2004; North et al., 2009) and medicine (Hoffman, Aitken, & Duffield, 2009; Loveday, Wiggins, Searle, Festa, & Schell, 2013; Myles-Worsley, Johnston, & Simons, 1988; Wigton et al., 1986).

Differences in the way that experts and novices use environmental cues is highly domain specific (Cellier, Eyrolle, & Mariné, 1997). For example, Chase and Simon (1973) showed how, compared to less experienced players, chess masters were more capable in reconstructing from memory, chess pieces on a board. Similarly, in tennis, Shim, Carlton, Chow, and Chae (2005) found differences in the way that skilled and novice tennis players extracted visual information from an opponent's movement pattern to predict ball movement, so that skilled players were significantly more accurate than novices in anticipating the future trajectory of the ball.

The expert-novice paradigm has shown reasonably consistent support for the proposition that experienced operators can be reliably discriminated from their less experienced counterparts by their superior ability to recognise and apply task-relevant cues during situation assessment (Calderwood, Klein, & Crandall, 1988; Coderre et al., 2003; de Groot, 1965; Ettenson, Shanteau, & Krogstad, 1987; Hoffman et al., 2009; Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Müller et al., 2006; North et al., 2009; Wiggins, Azar, Hawken, Loveday, & Newman, 2014; Williams, Ward, Knowles, & Smeeton, 2002). This research,

however, has tended to concentrate primarily on operators' use of visual cues during situation assessment. Yet, the multifaceted nature of many real-world environments such as fire fighting, air traffic control and rail control, often require operators to attend to auditory, olfactory and tactile cues *in addition to* visual cues during situation assessment. It is important to examine how cues from other modalities are extracted and utilised by operators during situation assessment, since it will have implications for workplace evaluation and training within technical environments.

Auton, Wiggins, and Searle (2014) examined the relationship between visual cue utilisation and auditory cue utilisation during situation assessment within the context of electricity transmission control. It was predicted that operators with greater levels of visual cue utilisation would possess a superior ability in the use auditory cues to detect a specified communication error during radio communication (i.e., a task that involved the application of auditory cues only). However, no relationship was evident between visual cue utilisation and auditory cue utilisation. Specifically, they found that irrespective of levels of visual cue utilisation, all participants used auditory cues in the same manner to assess instances of miscommunication.

Auton et al. (2014) attributed this finding to one of two possibilities. Firstly, it was possible that the measurement of auditory cue use was unable to discriminate operators who possessed established auditory cue associations and those who did not possess such associations. During their study, Auton et al. (2014) asked participants to complete an *Auditory Readback Task*, which involved operators listening to a series of auditory communicative exchanges and making assessments as to the level of miscommunication. This task was designed to assess the cue associations that operators had developed between specific auditory features (namely intonation, inter-turn delay and filler) that had previously been found to be associated with a particular type of miscommunication (event; Auton, Wiggins, Searle, Loveday, & Xu Rattanasone, 2013). However, the case could be made that,

as participants' cognitive resources were not restricted, it was difficult to distinguish between: (1) those operators that lacked such associations, but nevertheless used the time available to consciously consider the meaning of the auditory features, and; (2) those operators whose responses were based on automatic associations in memory.

Auton et al. (2014) postulated that the null results, if not a by-product of how auditory cue utilisation was measured, could also be due to the fact that the use of auditory cues during assessments of miscommunication represented a general skill, rather than the product of domain-specific experience. It may be the case that the ability to assess miscommunication during radio communication is a skill that can be developed through everyday telephone exchanges. As Auton et al. (2014) only used participants with domain-relevant expertise, and did not restrict their cognitive resources as they evaluated communication exchanges, it remains unclear which explanation is more likely to account for the null association between visual and auditory cues.

The Current Study

The current study extended Auton et al. (2014) by examining whether there was a relationship between visual and auditory cue utilisation, when auditory cue use is measured under conditions that limit cognitive resources. The current study also incorporated two groups of participants that differed in domain-related experience to assess the possibility that auditory cue utilisation during assessments of miscommunication is a general skill, unrelated to domain experience.

The current study took place within the domain of electrical rail control, which involves the supply of high voltage traction electrical power to railway trains. Electrical control operators represented those participants with domain-relevant experience. Electrical control operators are responsible for monitoring and maintaining the supply system by identifying faults and organising fault correction work. In contrast, first year psychology students from a tertiary institution were selected on the basis that they were naïve to the domain. All of the

participants were asked to assess the extent to which an error in communication had taken place during a series of simulated telephone exchanges between two electrical control operators.

A communication error perception task was selected for two reasons. Firstly, within many technical domains, the transmission and receipt of directives takes place over the radio or telephone (Kanki & Smith, 2001). As a telephone or radio exchange is exclusively the exchange of auditory information, the opportunity for operators to detect an error in communication is restricted to the auditory components of the conversation (Nygaard & Queen, 2008). This differs from face-to-face communication where operators have access to visual features (such as the expression on a speaker's face). As the exchange of directives over the telephone is commonplace within electrical rail control, experienced operators would, arguably, have a greater opportunity to strengthen the feature-event (cue) associations that signal errors in communication. Consequently, this task represents an ideal medium in which experienced and less experienced operators' auditory cue utilisation can be examined.

Secondly, this task is largely consistent with that used by Auton et al. (2014), who investigated the relationship between visual and auditory cue utilisation amongst operators within high-risk environments. Maintaining consistency with Auton et al. (2014), the current study examined how participants assessed the extent to which a nonunderstanding had occurred through the use of auditory features. A *nonunderstanding* occurs when the receiver of a message fails to obtain any meaningful interpretation of the message from the sender (Hirst, McRoy, Heeman, Edmonds, & Horton, 1994; Skantze, 2005; Weigand, 1999). During such instances, the receiver will immediately recognise that the message has not been understood (Hirst et al., 1994). At this stage, the receiver can either seek clarification from the sender such that the original message can be fully understood (Clark & Schaefer, 1987) or can deceptively conceal the nonunderstanding and continue with the exchange (Weigand, 1999). The latter is referred to as a *concealed* nonunderstanding.

This type of communication error was selected on the basis that the telephone communication protocol used within this domain (known as the *readback/hearback protocol*) does not protect against the occurrence of nonunderstandings, as it requires evidence that the instruction has been heard accurately, rather than accurately understood. The readback/hearback protocol is a communication procedure that requires the *receiver* of an instruction to repeat the entire message or part thereof back to the *sender* (Cushing, 1994; Prinzo & Britton, 1993). This is referred to as the *readback*. This procedure is designed to minimise the risk of errors during communicative exchanges over the telephone, as it obtains confirmation that the receiver has heard the message correctly (Australian Transport Safety Bureau, 2007; Prinzo & Britton, 1993; Schneider, Healy, & Barshi, 2004; Streitenberger, Breen-Reid, & Harris, 2006).

Although this protocol does protect against instances where a receiver mishears an instruction (as it allows a sender to identify and correct it), reading back an instruction does not necessarily ensure that it has been understood by a receiver (Weigand, 1999). For reasons such as cultural differences or to save face, a receiver can elect to conceal the fact that an instruction has not been understood and proceed to supply the sender with an accurate readback, thereby removing the opportunity for the sender to repair the nonunderstanding.

Arguably, nonunderstandings that occur during a readback exchange can be detected on the basis of auditory cues alone. Auditory cues consist of both 'what' speakers say (the *semantic* cues of the utterance) and 'how' speakers say it (the *prosodic* cues of the voice; (Nygaard, Herold, & Namy, 2009; Nygaard & Queen, 2008; Soskin, 1953). The semantic cues of an utterance incorporate the syllables, words and phrases of an exchange, whereas the prosodic cues incorporate a variety of speech properties including intonation, rhythm, intensity, pitch, timing and fillers (interjections such as *um* or *uh*; Bolinger, 1985; Cruttenden, 1986; Hunter, 2004; Monnot, Nixon, Lovallo, & Ross, 2001; Nygaard et al., 2009). In instances where a receiver has heard but not understood an instruction, a reply containing an

accurate readback can conceal this nonunderstanding from the sender. As a result, there will be no semantic evidence of a lack of understanding. During such instances, the sender must rely on how the receiver delivered the readback (i.e., the prosodic cues of the receiver's voice) to gauge issues of understanding.

Auton et al. (2013) investigated the prosodic features that were useful in the detection of nonunderstandings within the domain of electricity transmission control. They asked participants (i.e., listeners) to listen to a series of simulated communicative exchanges, where a sender delivered an instruction and a receiver repeated the instruction (i.e., the readback). The participants were then asked to rate the extent to which they perceived the receiver (i.e., the speaker) had understood the instruction.

Auton et al. (2013) found that the prosodic cues of intonation and filler (contained within the readback portion of the exchange) signalled a speaker's state of understanding to a listener. In the case of intonation, listeners perceived that receivers had a lower level of understanding of the instruction when they read back the instruction with a rising intonation (i.e., response rising in pitch), rather than a falling intonation (i.e., a response declining in pitch). For filler, listeners perceived readbacks preceded by the interjection '*um*' as signalling a relatively lower level of understanding, compared to readbacks that were not preceded by a filler.

Auton et al. (2014) extended this work by investigating whether participants' levels of visual cue utilisation performance across a series of domain-relevant tasks, influenced the way in which network operators utilised prosodic cues (namely, intonation, inter-turn delay and filler) during assessments of nonunderstanding. The current study builds on Auton et al. (2014) by examining the use of these prosodic cues in a different domain (electrical rail control) as well as imposing a secondary task that restricts the cognitive resources available during prosodic cue use.

Participants were asked to complete an online Situational Judgement Test (SJT),

EXPERT intensive skills evaluation (*EXPERTise 1.0*: Wiggins, Harris, Loveday, & O'Hare, 2010), that comprised a series of visual cue-based diagnostic tasks that has previously been shown to reliably discriminate typologies of behaviour that reflect greater or lesser levels of cue utilisation (Loveday, Wiggins, Harris, O'Hare, & Smith, 2013; Wiggins, Azar, et al., 2014). They also completed an amended version of the *Auditory Readback Task* as originally described in Auton et al. (2013) to assess their use of prosodic cues during perceptions of nonunderstandings in radio communication.

Research aim and hypotheses.

The overall aim of the present study was to examine whether there was a difference in the way that operators rated the importance of the three prosodic cues of nonunderstanding, depending upon their performance across a battery of diagnostic tasks that are associated with the utilisation of visual cues. A secondary aim was to extend the work of Auton et al. (2013) by establishing whether the prosodic cues of intonation, filler (and inter-turn delay[1]) found to be useful by power control operators during perceptions of nonunderstanding, could also be used by participants within a different domain.

In comparison to participants with lesser levels of cue utilisation, it was hypothesised that participants with greater levels of cue utilisation would perceive a greater difference in levels of understanding on the part of the receiver between:

(H1) Readback responses that ended in rising intonation compared to readback responses that ended in falling intonation.

(H2) Readback responses that were preceded by a shorter inter-turn delay compared to a longer inter-turn delay.

(H3) Readback responses that were preceded by a filler compared to an unfilled inter-turn delay of the same duration.

[1] Although both Auton et al. (2013) and Auton et al. (2014) found that an incremental increase in inter-turn delay duration from 600 ms to 1200 ms was not an important cue used by operators to assess nonunderstandings, it is unclear whether this effect would be enhanced with increased demands on working memory or in a different operational context. Therefore, inter-turn delay was included as a variable in the design of the current study.

Consistent with Auton et al. (2013), it was hypothesised that:

(H4) Readback responses ending in rising intonation would receive lower ratings of understanding of the instruction, compared to readback responses ending in falling intonation.

(H5) Readback responses that commenced after a longer inter-turn delay would receive lower ratings of understanding of the instructions compared to readback responses that commenced after a shorter inter-turn delay.

(H6) Readback responses preceded by a filler would receive lower ratings of understanding of the instruction compared to readbacks that were not preceded by a filler.

Method

Participants

To obtain two groups of participants that differed distinctly in terms of their experience within the electrical rail control domain, participants were sought either from a major rail freight operator in Australia (experienced; N = 6) or from an introductory psychology course from an Australian tertiary institution (inexperienced N = 35). All of the participants identified as being fluent in English. The electrical control operators were exclusively male, ranged in age from 33 to 60 years (M = 47.67, SD = 10.69) and had accumulated between one and 40 years of experience within electrical rail control (M = 14.83, SD = 15.64). The Psychology students were predominantly female (77.1%), ranged in age from 18 to 39 years (M = 20.17, SD = 4.93) and had no prior experience within the domain of electrical rail control. Electrical control operators received a US$25 Amazon eVoucher upon completion of the study and psychology students received course credit in exchange for participation.

Materials

This study required participants to complete two distinct tasks. The Auditory Readback Task assessed the extent to which participants used prosodic features during assessments of

nonunderstanding in radio communication. The *EXPERTise 1.0* SJT, categorised participants into typologies based on their scores across a series of tasks that assessed visual cue utilisation within the context of electrical rail control.

Auditory Readback Task.

Design.

The Auditory Readback Task comprised a 3 x 2 x 2 x 2 (Communicative Exchange [1, 2, 3] x Intonation [Rising, Falling] x Inter-turn Delay [Shorter, Longer] x Filler [Present, Absent]) within-groups design, resulting in 24 distinct conditions.

'Communicative Exchange' included three semantically distinct auditory communicative exchanges. Each exchange consisted of two components: a sender delivering an operational instruction to a receiver, and a receiver repeating the instruction back to the sender (i.e., delivering a *readback*). The three exchanges differed in the type of operational task that the sender requested the receiver to execute (see Appendix A for the semantic content of the communicative exchanges). Three different instructions were incorporated into the design to ensure that prosodic cue use was not contingent on the semantic content of a particular communicative exchange.

'Intonation' referred to whether the readback portion of each communicative exchange ended with the receiver's voice rising in pitch (rising intonation) or falling in pitch (falling intonation). 'Inter-turn delay' specified whether the duration between the completion of the sender delivering the instruction and the receiver commencing the readback was either shorter (600 ms) or longer (1200 ms). 'Filler' indicated whether or not the receiver used a verbal interjection (i.e., a filler, specifically *um*) before delivering the readback.

The participants listened to the communicative exchanges in randomised order. After listening to each individual exchange, they were asked to rate the extent to which the receiver was perceived to have had understood the meaning of the instruction on a 7-point scale from 1 (*did not understand at all*) to 7 (*understood completely*; Auton et al., 2013)

Stimuli development.

Communicative exchanges.

Three communicative exchanges were devised in consultation with two Subject Matter Experts (SME) from a large Australian rail freight operator. The SMEs were selected on the basis of peer review and industry position. Each exchange included both an operational instruction typical for that industry and a readback of that instruction in a manner that was also typical for the industry. Specifically, each readback began with an acknowledgement that the instruction had been understood (*I understand*; *It is understood*; *Got that*) followed by the repetition of the operational steps included in the instruction to indicate that the instruction had been heard correctly (the objective of a readback response; Cushing, 1994).

Two Australian English male speakers recorded the auditory versions of the communicative exchanges using a digital audio editor program, *GarageBand*™, on a MacBook™ laptop. One speaker played the role of the sender (who delivered the instruction in each of the three exchanges) and the other speaker played the role of the receiver (who delivered the readback in each of the three exchanges). The speakers maintained the same system sender-receiver roles for all exchanges to ensure that any variation in participant responses was not a product of differences in the voice quality of the receiver. The duration of the complete exchanges (including both the instruction and the readback) ranged from 9 to 11 seconds.

Prosodic manipulation of readbacks.

The readback portion of each exchange was manipulated to vary systematically the final intonation at the end of the readback response, the duration of the inter-turn delay between the offset of the sender's instruction and the onset of the receiver's readback, and whether or not the receiver preceded the readback with a filler.

To generate readbacks that varied in final intonation (either rising or falling), the male Australian English speaker who performed the role of the receiver was directed to deliver the

readback ending with either a rise or fall in pitch. To ensure that listeners perceived the intonation manipulations as intended, a manipulation check was undertaken (see Manipulation Checks section).

The current study compared the effects of a 600 ms and 1200 ms inter-turn delay on assessments of nonunderstanding on the part of the receiver. Using GarageBand™, silences of these two lengths were electronically inserted between the offset of the sender delivering the instruction and the onset of the receiver delivering the readback.

The readback portion of each exchange was either preceded by a filled inter-turn delay (i.e., the delay between the two speakers contained the filler *um*) or by an unfilled inter-turn delay (i.e., the delay between the two speakers contained silence only). To create filled inter-turn delays, the speaker who played the role of the receiver was recorded producing an *um* before delivering each readback. A single recording of *um* was selected for its clarity and timbre, and then this filler was digitally inserted within an inter-turn delay of 600 ms (100 ms silence, *um* [400 ms], 100 ms silence) and 1200 ms (700 ms silence, *um* [400 ms], 100 ms silence).

Numeric Digit Span Recall Task.

The Numeric Digit Span Recall Task represented a concurrent task to be completed alongside the Auditory Readback Task. The purpose of this concurrent task was to address the concern highlighted in Auton et al. (2014) and restrict participants' cognitive resources during assessments of nonunderstanding. This secondary task was expected to differentiate; (1) those operators who have developed strong feature-event cue associations in LTM and are able to access these associations even when cognitive resources are restricted and, (2) those operators who do not possess such refined associations and therefore, rely more heavily on the processes of working memory to understand the meaning of the prosodic features (Coderre et al., 2003; Rasmussen, 1983).

The Numeric Digit Span Recall Task required participants to remember a series of eight

single numeric digits that were visually displayed for the duration of each communicative exchange (see Figure 1). Each series of numeric digits included numerals from one to nine in a randomised order for each exchange, with no repetition of digits. After each communicative exchange finished, the series of numeric digits disappeared, and participants were asked to type in the digits in an open response box. Participants were then asked to rate the level of nonunderstanding on a subsequent page.

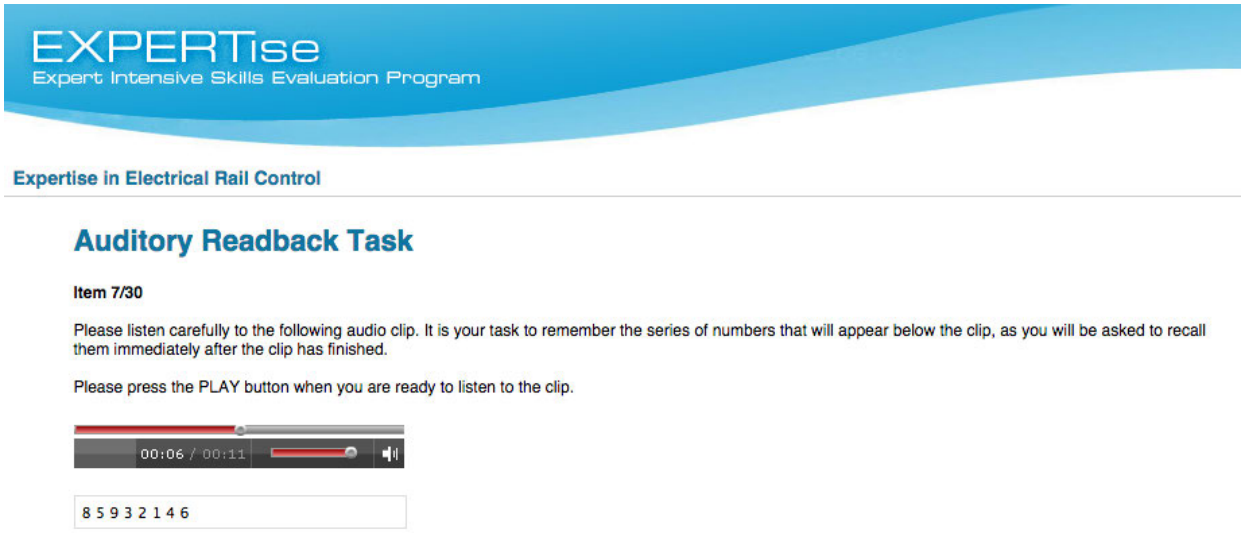


Figure 1. The visual presentation of the Auditory Readback Task and the Numeric Digit Span Recall Task.

To minimise the opportunity to focus exclusively on memorising the numeric digit string and overlooking the communicative exchanges altogether, participants were also presented with four communicative exchanges that required a response to a multiple-choice question regarding the semantic content of the exchange. These four exchanges were prosodic variations of one communicative exchange that was semantically distinct from 'Communicative Exchange 1','Communicative Exchange 2' and 'Communicative Exchange 3'[2]. The exchanges requiring a multiple-choice response were integrated randomly within the 24 communicative exchanges that measured perceptions of nonunderstanding such that participants could not predict when they would be required to respond to multiple-choice

[2] The semantic content of the communicative exchange that was used to generate multiple-choice questions was as follows:
Instruction: *You may now test and apply earths at or within the limits of isolation stated within Part One B of the Form B.*
Readback: *I understand that I am authorised to test and earth the overhead line equipment at or within the limits of isolation stated on the Form B.*

question, and therefore, were more likely to attend to all exchanges, rather than dedicating all of their cognitive resources to remembering the numeric digits.

To ensure that a series of eight numeric digits did not restrict cognitive resources to such an extent that it prevented participants from attending to the exchanges altogether, a manipulation check was conducted (see Manipulation Check section for details). In total, participants were asked to listen and respond to 30 communicative exchanges presented in randomised order, two of which were practice items, 24 of which were communicative exchanges, and four of which were communicative exchanges requiring a response to a multiple-choice question.

Manipulation checks.

Manipulation check for intonation contours.

Eleven post-graduate students from an Australian University participated in the manipulation check. The sample comprised native speakers of English who were predominantly female (81.82%). The participants ranged in age from 23 to 49 years (M = 32.09, SD = 8.37). Using an online experimental format, participants were asked to listen to the readback portion of each of the three communicative exchanges, and were asked to judge whether the speaker used either rising or falling intonation at the end of the readback.

Participants were presented with four repetitions of each of the readbacks from the three communicative exchanges: two examples of the rising intonation version of each readback and two examples of the falling intonation version of each readback. Overall, participants listened to, and categorised, a total of 12 readbacks. The results indicated that the three readbacks that were intentionally spoken by the male speaker with a rising intonation at the end of the utterance were perceived as such by all participants. Similarly, the readbacks that were spoken with a falling intonation at the end of the utterance were all accurately categorised as such by 100% of participants.

Numeric Digit Span Recall Task manipulation check.

The purpose of the Numeric Digit Span Recall Task manipulation check was to establish the ideal length of the numeric digit string that would sufficiently limit participants' cognitive resources, but not to such an extent that it restricted them from attending to the communicative exchanges. Thirteen undergraduate students from an Australian University participated in the manipulation check. All of the participants were native speakers of English and ranged in age from 18 to 55 years (M = 25.31, SD = 11.79). The sample was predominantly male (61.54%).

The Numeric Digit Span Recall Task manipulation check comprised a 4 x 4 (Communicative Exchange [1, 2, 3, 4] x Numeric Digit Span Length [0, 4, 6, 8]) repeated measures design, resulting in 16 distinct conditions. 'Communicative Exchange' referred to the electrical rail control operational exchanges described above (including the addition exchange used to generate multiple-choice questions), which included both an instruction and a readback of the instruction. 'Numeric Digit Span Length' referred to the number of numeric digits that a participant was required to retain while listening to each communicative exchange. This ranged from 0 (i.e., participants were not required to remember any numeric digits at all while listening to each exchange) to eight (i.e., participants were required to remember a series of eight numeric digits as they listened to each exchange).

The participants completed the Numeric Digit Span Recall Task manipulation check using an online survey provider during a self-selected time. They were presented with 16 'trials'. Each trial presented participants with an audio clip containing a communicative exchange. As soon as the clip began, the numeric digit string appeared beneath the audio tool bar. After the audio clip had finished, the numeric digit string disappeared, and the participants were asked to type in the series of digits that they had just viewed into a free response text box. To assess whether the task to retain the numeric digit string restricted cognitive resources, the participants were asked to respond to a multiple choice question

relating to the semantic content of the exchange. They were then asked to rate the difficulty of the task (i.e., subjective measure of task difficulty; Paas & Merriënboer, 1994).

The results of the manipulation check indicated that, compared to the zero, four and six-digit numeric span, the participants subjectively perceived the eight-digit numeric string to be the most difficult. They were also significantly less accurate in recalling the eight digit strings, compared to the six and four digit strings. Nevertheless, the number of digits that they were required to recall did not affect participants' accuracy in the multiple-choice questions concerning the communicative exchanges. In combination, these results suggest that an eight numeric digit string would restrict cognitive resources considerably, but not to the extent that it would deprive participants of the resources necessary to attend to the communicative exchanges.

EXPERTise 1.0 test battery.

EXPERTise 1.0 is an SJT that can be customised to record operator performance during a series of cognitive tasks from a range of industries (Wiggins et al., 2010). The EXPERTise 1.0 test battery comprises cognitive tasks that have been shown to categorise participants into 'typologies' based on their relative performance on the use of task-related features during diagnostic decision making within the power control industry power control (Loveday, Wiggins, Harris, et al., 2013) pediatric diagnosis (Loveday, Wiggins, Searle, et al., 2013) and aviation decision making (Wiggins, Azar, et al., 2014). The test-retest reliability of EXPERTise 1.0 has been demonstrated as satisfactory (Loveday, Wiggins, Festa, Schell, & Twigg, 2013). The four tasks that comprise EXPERTise 1.0 include the *Feature Identification Task*, the *Paired Association Task*, the *Feature Discrimination Task* and the *Transition Task*.

Stimuli.

Feature Identification Task.

Feature identification is based on the observation that experts are able to identify and utilise visual features in the environment that are more diagnostic of the system state

compared to novices (Müller et al., 2006; Schriver et al., 2008; Stokes, Kemper, & Kite, 1997). This skill can be assessed using a target-matching task, requiring an operator to locate a specific domain-relevant target contained within a complex display.

In the current study, feature identification was examined in two stages. Stage One comprised two tasks; each task required participants to identify a domain-relevant target. The first task required participants to identify an open circuit breaker in the abnormal state from within a complex network, while the second required participants to identify the critical open circuit breaker alarm within an Events Summaries List. Stage Two of the Feature Identification Task required participants to identify the location of an open circuit breaker after a brief presentation of a complex network screen.

During Stage One (Part 1), participants were presented with a series of 22 Supervisory Control and Data Acquisition (SCADA) screen captures that incorporated an open circuit breaker in the abnormal state. All of the SCADA screens adhered to the principles of colour propagation that ensured the ecological validity of the task. The participants were asked to click on the open circuit breaker in the abnormal state as quickly as possible (see Figure 2).

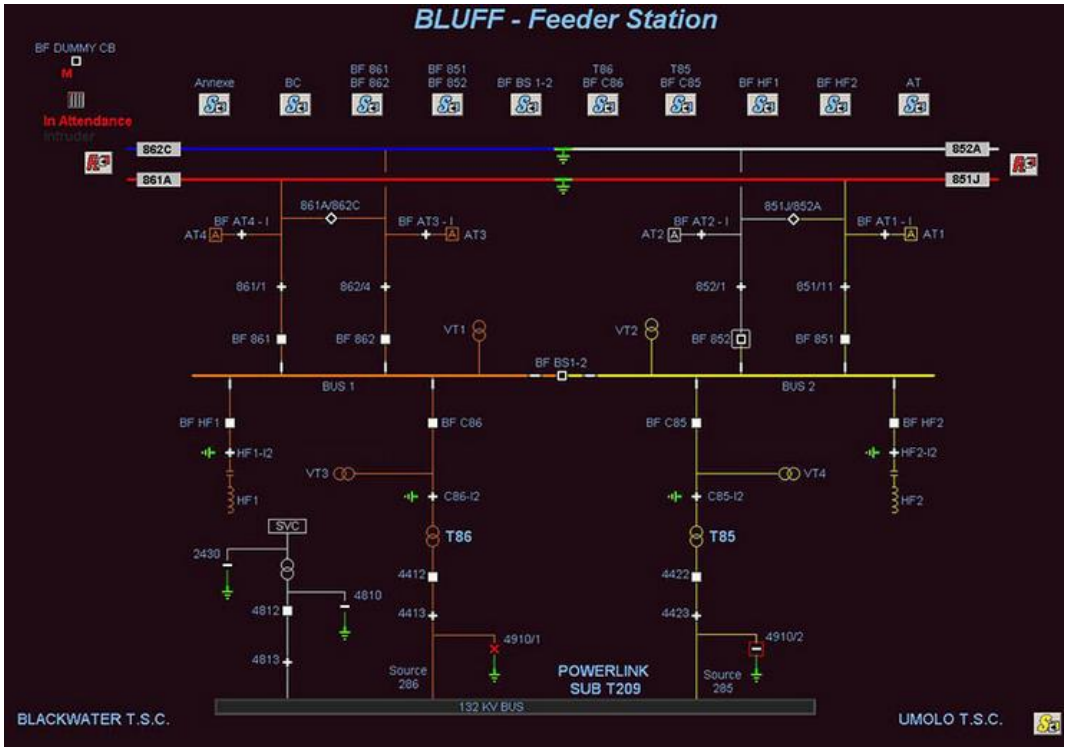


Figure 2. Example of Supervisory Control and Data Acquisition (SCADA) screen capture. The open white square in position BF 852 represents a circuit breaker in the abnormal state.

During Stage One (Part 2), participants were presented with a series of 21 excerpts of an operational Events Summary List that contained a critical alarm (namely, an open circuit breaker alarm). Consistent with Stage One (Part 1), the participants were asked to click, as quickly as possible, on the critical alarm (see Figure 3). Response latencies were collected for both parts of Stage One of this task, where a lower response latency indicated a greater capacity to identify and extract key information from a complex display (Wiggins, 2014).

24/04/2013 18:47:04.736	GRANTLEIGH BATTERY 240V SUPPLY FAILED	N	ON	System	Alarm Server
24/04/2013 18:47:04.724	GRANTLEIGH NO 2 AUX SUPPLY FEED SELECT	I	OFF	System	Alarm Server
24/04/2013 18:47:04.249	GRANTLEIGH NO 1 AUX SUPPLY 240 V FAILED	U	OFF	System	Alarm Server
24/04/2013 18:47:04.234	WESTWOOD WD 111 CB CLOSED	U	ON	System	Alarm Server
24/04/2013 18:47:03	WESTWOOD WD 111 CB Command Sent	O	CLOSE CMD	ECO Rocky	Cit_RokOWS1
24/04/2013 18:43:31	Fault Locator Auto Calculation : 111_112	O	PROCESSED	System	System
24/04/2013 18:43:25.220	GRANTLEIGH FL111_112 DATA SCAN RECEIVED	I	OFF	System	Alarm Server
24/04/2013 18:43:24.910	GRANTLEIGH FL111_112 DATA SCAN RECEIVED	I	ON	System	Alarm Server
24/04/2013 18:43:16.854	WESTWOOD TC AUTO C/O AUX2 NOT HEALTHY	U	OFF	System	Alarm Server
24/04/2013 18:43:16.395	WESTWOOD WD 111 CB X PROTECTION TRIP	U	OFF	System	Alarm Server
24/04/2013 18:43:16.293	GRANTLEIGH DISTANCE PROTECTION TRIP	U	OFF	System	Alarm Server
24/04/2013 18:43:16.269	GRANTLEIGH PLINK PLANT ALARM URGENT ON	U	OFF	System	Alarm Server
24/04/2013 18:43:16.266	GRANTLEIGH BATTERY 240V SUPPLY FAILED	N	OFF	System	Alarm Server
24/04/2013 18:43:16.262	GRANTLEIGH GT 111 CB OPEN	U	ON	System	Alarm Server
24/04/2013 18:43:16.261	GRANTLEIGH AUX SUPPLY FEED SELECT	I	ON	System	Alarm Server
24/04/2013 18:43:16.242	GRANTLEIGH NO 2 AUX SUPPLY FEED SELECT	I	ON	System	Alarm Server
24/04/2013 18:43:16.213	GRANTLEIGH DISTANCE PROTECTION TRIP	U	ON	System	Alarm Server

Figure 3. Example of an operational alarms list, where the alert in red ‘GRANTLEIGH GT 111 CB OPEN’ represents a critical alarm.

During Stage Two of the Feature Identification Task, participants were presented with a series of 20 SCADA screen captures that incorporated an open circuit breaker in an abnormal state, each of which appeared for four seconds. On the subsequent screen, participants were asked to select the correct name of the open circuit breaker that appeared in an abnormal state from four possible options. The accuracy of this response was recorded.

Paired Association Task.

The Paired Association Task presented participants text-based domain-relevant feature-event pairs, where participants were asked to rate the strength of the perceived relatedness between the two items (Morrison, Wiggins, Bond, & Tyler, 2012). This task is designed to assess the strength of association that participants possess in memory between features and events. As experienced operators have a more refined repertoire of cue associations resident in LTM (Schriver et al., 2008), it is expected that they would display greater discrimination for ratings of association between feature-event pairs (Morrison et al., 2012).

The Paired Association Task was completed in two stages. In the first stage, the

participants viewed 27 text-based feature-event pairs relevant to the electrical rail control domain. Each pair was presented sequentially such that the feature (e.g., *Traction Fault*) would appear on-screen shortly before the event (e.g., *Explosion*). Each stimulus was available on screen for one second. After the presentation of each pair, participants were asked to rate the strength of perceived association between the two features on a 6-point Likert scale, from 1 (*extremely unrelated*) to 6 (*extremely related*). During the second stage of the task, the participants were asked to rate the strength of perceived association between feature-event pairs that were presented *simultaneously* on screen for a duration of two seconds. The participants were asked to rate 27 feature-event pairs in each stage. See Appendix B for the catalogue of the feature-event pairs used in the Paired Association Tasks.

Feature Discrimination Task.

The Feature Discrimination Task comprised a complex, yet ambiguous scenario that required a response. The scenario contained several environmental features to aid in an operator's situation assessment. Participants were asked to rate the importance of each of the environmental features to their final assessment of the situation. This task was based on the conceptualisation of the Cochran-Weiss-Shanteau (CWS) index of expertise (Shanteau, Weiss, Thomas, & Pounds, 2002). This index suggests that experts are better able to discriminate consistently between features, compared to their novice counterparts.

During the Feature Discrimination Task, participants responded to two distinct scenarios. In the first scenario (see Appendix C), participants were presented with a detailed scenario regarding a suspicious noise emanating from a 'switch gear' at a switching station. For the second scenario (see Appendix D), participants were presented with a 'trip' at a feeder station, with an associated Events Summary List. For both scenarios, participants were asked to assess the situation and choose a course of action from several possible options. On a second screen, participants were asked to rate the relative usefulness of a series of features associated with the situation when deciding the correct course of action to take in response to

the situation. Individual ratings were made on a 10-point Likert scale, from 1 (*not important at all*) to 10 (*extremely important*).

Transition Task.

The Transition Task is based on the finding that expert and novice operators vary in the approach taken to access task-related information during the initial assessment of a situation (Wiggins & O'Hare, 1995). Novice operators acquire information in a sequence consistent with its visual presentation. In contrast, experts are more discriminating in their approach, extracting information on the basis of relevance, rather than physical accessibility or the sequence in which features are presented (Wiggins & O'Hare, 1995).

To assess operators' information acquisition strategies, participants were presented with an intentionally incomplete scenario, such that that additional information would need to be obtained prior to forming an accurate assessment of the situation. The additional information was presented in a list of categories that could be selected to reveal further information about the event described (see Figure 4 and Appendix E for the full scenario).

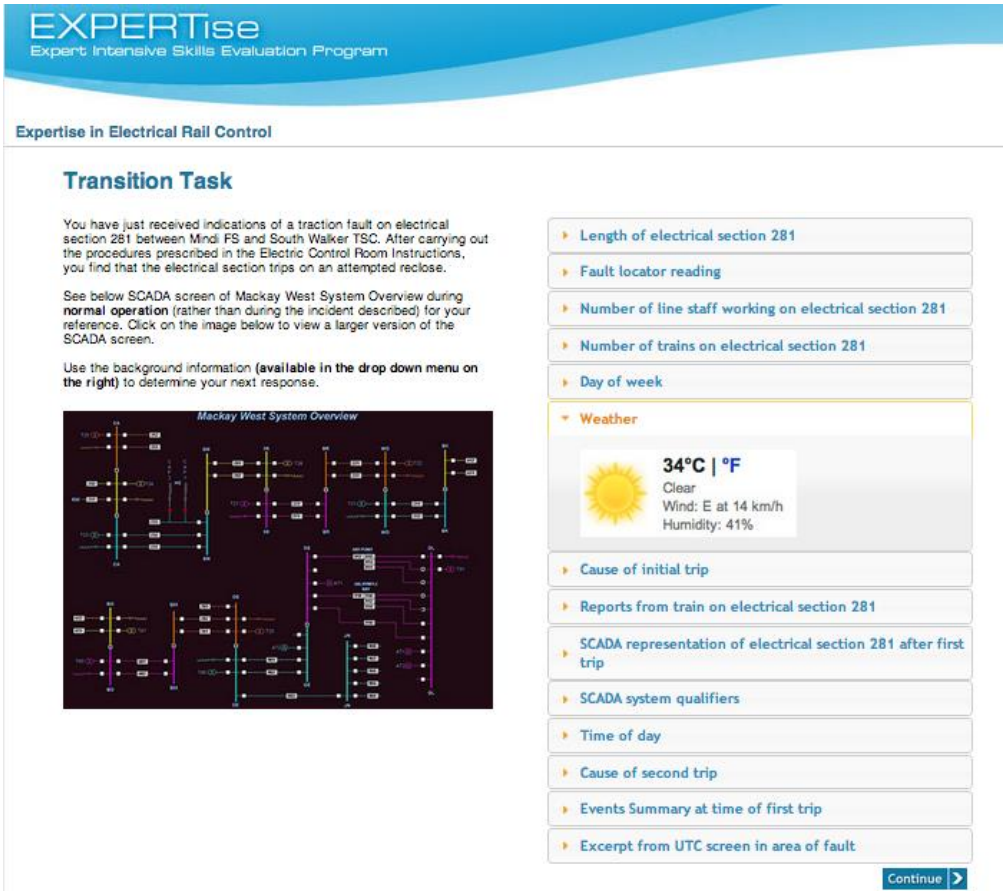


Figure 4. Screen capture of the information presented during the Transition Task. Categories of features are displayed in a list on the right hand side, which can be expanded by a mouse click.

Stimuli development.

Consistent with the standard approach for developing stimuli for EXPERTise 1.0 (Loveday, Wiggins, Harris, et al., 2013), the four tasks for the current study were developed from cognitive interviews conducted with three Subject Matter Experts (SMEs) from Australia's largest rail freight operator. The SMEs were selected on the basis of peer reference and industry position. Based on the critical incident technique (Flanagan, 1954; Klein, Calderwood, & MacGregor, 1989), each SME was asked to recall and describe in detail, an operational incident that was somewhat uncommon and complex, in which the SME's expertise was perceived to have made a critical difference to the outcome. A semi-structured interview was then used to probe details about various aspects of the decision making process (Klein et al., 1989).

Each of the three SMEs described at least one critical incident in detail and was probed for all the information available during the incident. It was important to elicit all of the information, even that which was perceived to be irrelevant, as skilled operators can often find it difficult to articulate the features that influenced their decision making (Kahneman & Klein, 2009). Three critical incidents that were described during the cognitive interviews were arranged into chronological timelines and were used to develop the scenarios for the two Feature Discrimination Tasks and the Transition Task.

An index of feature-event pairs that an operator may or may not find relevant during situation assessment within electrical rail control was also developed from the cognitive interviews. These feature-event pairs formed the items in the Paired Association Task. The images used in the Feature Identification Tasks were screen captures of either circuit breakers in an abnormal state captured during real time (see Figure 2) or an Events List containing a critical alarm (see Figure 3). After the EXPERTise 1.0 battery of tasks was devised, the program was piloted by an electrical control operator from the same organisation who had not been involved in the development of the program. His feedback was incorporated into the final version of the program.

Design

Overall, the design of the current study comprised a 3 x 2 x 2 x 2 x 2 (Communicative Exchange [1, 2, 3] x Intonation [Rising, Falling] x Inter-turn Delay [Shorter, Longer] x Filler [Present, Absent] x Typology Membership [Greater levels of visual cue utilisation, Lesser levels of visual cue utilisation]) mixed-groups design, with communicative exchange, intonation, inter-turn delay and filler as the within-groups variables and typology membership as the between-groups variable, resulting in 24 distinct conditions. Mean ratings of understanding on the part of the receiver comprised the dependent variable.

Data Reduction

For the Auditory Readback Task, ratings of understanding for each readback response

were recorded on a 7-point Likert scale, from 1 (*did not understand at all*) to 7 (*understood complete*ly). Data were then aggregated for each level of each prosodic cue to create a mean rating of understanding for that variable level. Since the aim of the Numeric Digit Span Recall Task was to restrict participants' cognitive resources while listening to the communicative exchanges, data were only included where participants made an attempt to recall at least one of the eight digits. Of the 41 participants who were asked to respond to a total of 24 response items each (resulting in 984 data points), there were two instances, from two separate participants, where no attempt was made to recall any of the digits. These two data points were removed and treated as missing data.

The data reduction for the EXPERTise 1.0 tasks was consistent with the standard approach to the analysis of these data (Loveday, Wiggins, Harris, et al., 2013; Wiggins, Azar, et al., 2014). For the Feature Identification Task, the mean response latency to identify the critical feature was recorded for both parts of Stage One. For Stage Two, participants' accuracy in identifying the location of the open circuit breaker was recorded. For the Paired Association Task, ratings of perceived association between feature-event pairs were recorded on a scale from one to six. These ratings across both stages of the task were then combined into a single discrimination metric based on the mean variance of the participants' responses. Participants' response latency was also recorded for each item in this task. Although this task contained two stages, response latency was calculated across both stages.

For the Feature Discrimination Task, the ratings of utility of each of the nine (or twelve) features was recorded on a 10-point Likert scale, from 1 (*not important at all*) to 10 (*extremely important*). These ratings were then aggregated into a single discrimination metric based on the variance of the individual's responses. As this task comprised two stages, each participant received two variance scores. For the Transition Task, the ratio of pairs of features that were selected in sequence, compared to the total number of pairs of features available, was recorded.

Procedure

Electrical control operators completed the online study during individual work-related training sessions. Psychology students accessed the study online in a self-selected time. All of the participants were directed to the domain address where they were able to access the study. After providing consent, participants completed a brief set of demographic questions and proceeded to the Auditory Readback Task (approximately 20 minutes), and then completed the EXPERTise 1.0 tasks (approximately 40 minutes). Participants were asked to wear headsets during the Auditory Readback Task portion of the study.

Results

Overview of Analyses

The primary aim of the current study was to examine whether two typologies of participants, categorised on the basis of operator performance across the EXPERTise 1.0 tasks, differed in the use of prosodic cues during assessments of nonunderstandings. A secondary aim was to assess the utility of three prosodic cues, namely intonation, inter-turn delay, and filler, during perceptions of nonunderstanding. The statistical analyses to address these questions comprised two stages.

The first stage of the analysis was to confirm that the two typologies of participants, classified *a priori* on the basis of domain-related experience, could be differentiated on the basis of visual cue utilisation performance across the tasks that comprised EXPERTise 1.0. The second stage of the analysis examined the hypotheses. Specifically, it was designed to assess whether: (1) typology membership (as possessing either a greater or lesser level of cue utilisation) predicted the use of prosodic cues during perceptions of nonunderstanding, and whether; (2) the specified prosodic features influenced assessments of nonunderstandings during readback communicative exchanges.

Stage 1: Domain Experience and Visual Cue Utilisation Performance

As greater levels of cue utilisation are arguably a product of domain experience, an

independent samples t-test was conducted, with the vocational background of the participant (electrical control operator or psychology student) as the grouping variable, and the grand mean of the *z*-scores from the EXPERTise 1.0 tasks as the dependent variable.

Z-scores were computed for the following cue utilisation measures: response latency for both parts of Stage One of the Feature Identification Task, accuracy rates for Stage Two of the Feature Identification Task, the variance in association ratings collated across both stages of the Paired Association Task, the variance in the ratings of the utility of features for aggregated across both stages of the Feature Discrimination Task and the ratio of sequential pairs of features accessed compared to features available in the Transition Task. The grand mean has been used previously as a general, standardised measure of overall, visual cue utilisation (Wiggins, Brouwers, Davies, & Loveday, 2014).

Consistent with expectations, the results indicated that the electrical control operators performed at a significantly greater than average level across the battery of EXPERTise 1.0 tasks (M = .56, SE = .47), compared to those participants with no experience within electrical rail control (M = -.12, SE = .09), $t(39) = 3.02$, $p = .004$. As a result, the electrical controllers demonstrated a greater level of visual cue utilisation, while the psychology students demonstrated a relatively lesser level of visual cue utilisation. These classifications were used to constitute the between-groups variable that was employed in subsequent analyses of the hypotheses.

Stage 2: Analysis of Hypotheses

The hypotheses were examined using a 3 x 2 x 2 x 2 x 2 (Communicative Exchange [1, 2, 3] x Intonation [Rising, Falling] x Inter-turn Delay [Shorter, Longer] x Filler [Present, Absent] x Typology [Greater levels of visual cue utilisation, Lesser levels of visual cue utilisation]) mixed-design ANOVA. Mean ratings of understanding on the part of the receiver comprised the dependent variable.

Typologies and prosodic cue use.

Using planned contrasts to investigate the interactions hypothesised, it was evident that there was no significant interaction between typology and intonation, $F(1, 39) = 0.71$, $p = .405$, $\eta_p^2 = .02$ (H1). This suggested that there was no difference in the way that the two typologies responded to readbacks that ended in rising intonation compared to readbacks that ended in falling intonation (see Figure 5).

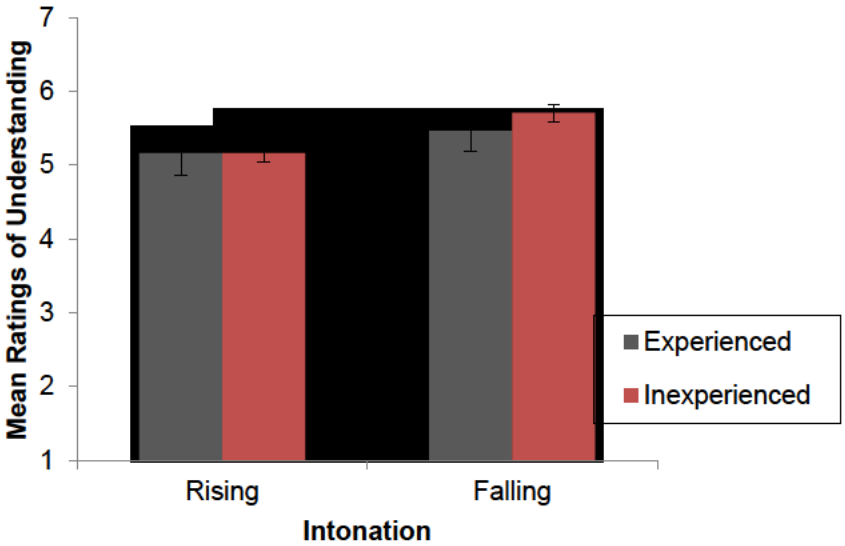


Figure 5. Non-significant two-way interaction between typology and intonation on mean ratings of understanding. Error bars represent standard error.

Consistent with the results pertaining to intonations, there was no statistically significant interaction evident between typology and inter-turn delay, thereby failing to provide support for H2, $F(1, 39) = 0.11$, $p = .740$, $\eta_p^2 = .00$. Participants across both typologies responded to readbacks that preceded a shorter or longer inter-turn delay in a similar fashion (see Figure 6).

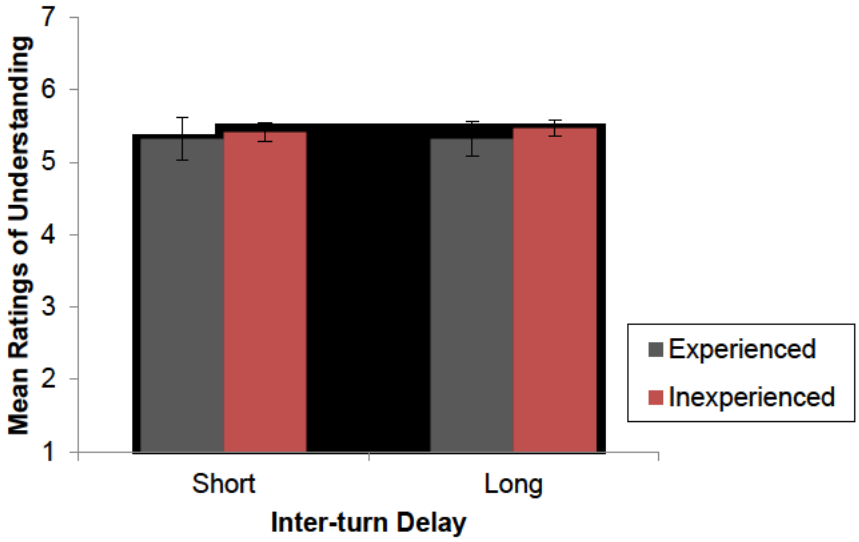


Figure 6. Non-significant two-way interaction between typology and inter-turn delay on mean ratings of understanding. Error bars represent standard error.

Finally, there was no significant two-way interaction between typology and filler, $F(1, 39) = 0.91$, $p = .346$, $\eta_p^2 = .02$, suggesting that both typologies responded similarly to readbacks that were either preceded or not, by a filler (see Figure 7).

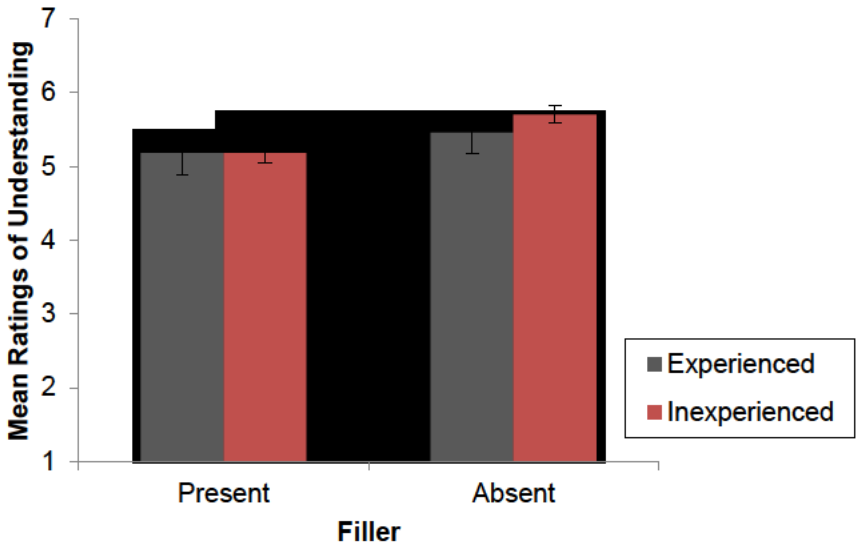


Figure 7. Non-significant two-way interaction between typology and filler on mean ratings of understanding. Error bars represent standard error.

Prosodic cues and assessments of nonunderstanding.

Consistent with expectations, the results revealed a statistically significant main effect for intonation, $F(1, 39) = 8.79$, $p = .005$, $\eta_p^2 = .18$ (H4). In particular, readbacks ending in rising intonation were awarded a lower rating of understanding by participants ($M = 5.17$, *SE*

= 0.17), compared to readbacks that ended in falling intonation (*M* = 5.59, *SE* = 0.15). Similarly, there was a significant main effect for filler, $F(1, 39) = 9.64$, $p = .004$, $\eta_p^2 = .20$ (H6), where participants reported lower ratings of understanding for readbacks that were preceded by a filler (*M* = 5.18, *SE* = 0.16), compared to those readbacks preceded by unfilled delays of the same duration (*M* = 5.58, *SE* = 0.15). Finally, there was no statistically significant main effect for inter-turn delay (H5), $F(1, 39) = 0.11$, $p = .740$, $\eta_p^2 = .00$. This indicates that readbacks that followed a shorter inter-turn delay were rated similarly in terms of perceived understanding (*M* = 5.37, *SE* = 0.16), compared to readbacks that followed a longer inter-turn delay (*M* = 5.39, *SE* = 0.13).

Of some interest in the present study was evidence of a main effect for communicative exchange, $F(2, 78) = 19.26$, $p < .001$, $\eta_p^2 = .33$. Specifically, post-hoc pairwise comparisons using a Bonferroni adjustment indicated that there was a statistically significant difference in ratings of understanding associated with Exchanges 1 and 2, Exchanges 2 and 3, and Exchanges 1 and 3. This suggests that different levels of understanding were associated with the readbacks contained in all three exchanges. However, despite these differences, there were no significant interactions between Communicative Exchange and Intonation, $F(2, 78) = 2.23$, $p = .115$, $\eta_p^2 = .05$, Communicative Exchange and Inter-turn Delay, $F(2, 78) = 0.76$, $p = .473$, $\eta_p^2 = .02$, nor Communicative Exchange and Filler, $F(2, 78) = 1.35$, $p = .266$, $\eta_p^2 = .03$. In combination, these results suggest that, while assessments of nonunderstanding varied in response to the semantic content of each communicative exchange, the semantic content of each exchange was not related to the use of prosodic cues during assessments of nonunderstanding.

Discussion

The overall aim of this study was to examine whether a relationship exists between two typologies, differentiated on the basis of visual cue utilisation, and the use of prosodic cues during assessments of nonunderstanding. A secondary aim was to confirm the utility of three

prosodic features, namely intonation, inter-turn delay, and filler, during perceptions of nonunderstandings in readback/hearback communicative exchanges within the domain of electrical rail control.

Typologies and Prosodic Cue Use

The current study included three hypotheses that examined typologies and prosodic cue use during assessments of nonunderstanding. In comparison to participants with lesser levels of visual cue utilisation, it was predicted that participants with greater levels of visual cue utilisation, would perceive a greater difference in the level of understanding of the receiver, between; readback responses ending in rising intonation compared to falling intonation (H1); readback responses preceded by a shorter compared to a longer inter-turn delay (H2), and; readback responses preceded by a filled compared to an unfilled inter-turn delay (H3).

Overall, the results failed to provide support for the hypotheses. Specifically, there was no relationship evident between the typologies, differentiated on the basis of visual cue utilisation, and perceptions of nonunderstanding in radio communication. This replicates the findings of Auton et al. (2014), even though the current study introduced the Numeric Digit Span Recall Task which, by restricting cognitive resources, was more likely to reveal differences in pre-existing cues. These results suggest that, rather than being related to domain-specific visual cue utilisation, the use of prosodic cues during assessments of nonunderstanding is likely to represent a general skill that is acquired during everyday interactions. This was not necessarily an unexpected finding, considering the opportunity to develop and reinforce cue associations between prosodic features and levels of understanding during everyday interactions.

Prosodic Cues and Assessments of Nonunderstanding

The current study also included three hypotheses that examined the prosodic cues used by participants during assessments of nonunderstanding. The findings replicate the outcomes of Auton et al. (2013), where both intonation and filler constituted prosodic cues that were

used by participants during assessments of nonunderstandings of full readback responses. Collectively, the findings from Auton et al. (2013) and the current study suggest that these prosodic cues are useful to operators in at least three different domains; electricity transmission control, electrical rail control and amongst university students.

Semantic Cues and Assessments of Nonunderstanding

The current study incorporated three semantically distinct communicative exchanges to demonstrate that prosodic cues could be used to signal nonunderstandings, irrespective of the semantic content of the readback. Although the results confirmed that the use of intonation and filler during perceptions of nonunderstanding was used consistently across all three exchanges, the three semantically distinct communicative exchanges did attract different ratings of understanding from participants.

The communicative exchanges used in the current study were not controlled in terms of content, complexity or length. This was consistent with the procedure taken to devise communicative exchanges in Auton et al. (2013) who, by contrast, did not find any differences in assessments of nonunderstanding between their two communicative exchanges. Arguably, a case could be made that the three readbacks contained within the communicative exchanges used in the current study differed in complexity. Message complexity has been defined by the number of distinct elements contained within a single transmission (Cardosi, Brett, & Han, 1996; Prinzo, Hendrix, & Hendrix, 2006). However, as this variable was not controlled *a priori*, it is unclear whether the differences in perceptions of understanding associated with the exchanges was a result of differences in readback complexity.

The findings of the current study, therefore, suggest that semantic cues contained within a readback response, such as message complexity, influence perceptions of nonunderstanding. It is important to identify the specific contexts in which the prosodic cues of intonation and fillers are used by operators to detect nonunderstandings, as it may be the case that prosodic cue use is constrained to particular semantic requirements. For example, both Auton et al.

(2013) and the current study investigated how prosodic cues were used by listeners when attending to full readback responses. However, there are many high-risk, high-consequence environments that do not embody a readback requirement or, more commonly, do not adhere to the requirement for a readback (as in the case of partial readbacks; Garzone et al., 2010; Hanna, Griswold, Leape, & Bates, 2005). Future research should investigate whether the utility of intonation and filler as prosodic cues vary, depending on the semantic cues of the readback response (such as during partial readback responses).

Limitations and Future Research

The main limitation of the current study was that it compared prosodic cue utilisation during perceptions of nonunderstanding between two typologies that were largely unequal in size. There were only six electrical control operators that comprised the typology with greater levels of visual cue utilisation, compared to 39 undergraduate psychology students that comprised the typology with lesser levels of visual cue utilisation. Due to the small sample size of electrical control operators, it may be the case that there was not enough power to detect any significant interactions between typology and prosodic cue utilisation. The issue of inadequate sample size needs to be considered in future research to determine whether there is a genuine association between visual and prosodic cue utilisation within the domain of electrical rail control, or whether, as the findings from the current study support, prosodic cue utilisation is a generic skill that requires no domain experience. Future research should also consider whether there is a difference between typologies and semantic cue use (such as message complexity) during perceptions of nonunderstanding, as arguably, the development of semantic cue associations requires domain experience.

Conclusion

Overall, the outcomes of the current study indicated that participants used the prosodic cues of intonation and filler during assessments of nonunderstanding in radio communication, replicating the results of Auton et al. (2013) but in a different domain. There was no

difference evident between typologies of participants based on visual cue utilisation performance and prosodic cue use, suggesting that the use of prosodic cues to assess understanding is a generic capacity possessed by all individuals. Future research should consider how prosodic cue use is influenced by the semantic content of the instruction response type (such as a partial readback response).

Paper 3 References

Abernethy, B. (1990). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sports Sciences*, *8*, 17–34. doi:10.1080/02640419008732128

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295. doi:10.1016/S0022-5371(83)90201-3

Apple Inc. (2014). GarageBand (6.0.4) [Digital Audio Workstation and Music Sequencer]. Retrieved from https://itunes.apple.com/au/app/garageband/id682658836?mt=12&ls=1

Australian Transport Safety Bureau. (2007). *Radiotelephony readback compliance and its relationship to surface movement control frequency congestion* (Investigation Report No. B2006/0053). Retrieved from http://www.atsb.gov.au/media/29962/b20060053.pdf

Auton, J. C., Wiggins, M. W., & Searle, B. J. (2014). The relationship between visual cue utilisation and prosodic cue utilisation during assessments of nonunderstanding in readback responses. *Unpublished Manuscript, Department of Psychology, Macquarie University, Sydney, Australia.*

Auton, J. C., Wiggins, M. W., Searle, B. J., Loveday, T., & Xu Rattanasone, N. (2013). Prosodic cues used during perceptions of nonunderstandings in radio communication. *Journal of Communication*, *63*, 600–616. doi:10.1111/jcom.12033

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, *106*, 410–433. doi:10.1037/0033-2909.106.3.410

Bisantz, A. M., & Pritchett, A. R. (2003). Measuring the fit between human judgments and automated alerting algorithms: A study of collision detection. *Human Factors*, *45*, 266–280. doi:10.1518/hfes.45.2.266.27243

Blignaut, C. J. H. (1979). The perception of hazard: I. Hazard analysis and the contribution of visual search to hazard perception. *Ergonomics*, *22*, 991–999. doi:10.1080/00140137908924673

Bolinger, D. (1985). *Intonation and its parts*. London, UK: Edward Arnold.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217. doi:10.1037/h0047470

Calderwood, R., Klein, G. A., & Crandall, B. W. (1988). Time pressure, skill, and move quality in chess. *The American Journal of Psychology*, *101*, 481–493. doi:10.2307/1423226

Cardosi, K., Brett, B., & Han, S. (1996). *An analysis of TRACON (Terminal Radar Approach Control) controller-pilot voice communications* (DOT Report No. DOT/FAA/AR-96/66). Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Cellier, J. M., Eyrolle, H., & Mariné, C. (1997). Expertise in dynamic environments. *Ergonomics*, *40*, 28–50. doi:10.1080/001401397188350

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81. doi:10.1016/0010-0285(73)90004-2

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, *2*, 19–41. doi:10.1080/01690968708406350

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education*, *37*, 695–703. doi:10.1046/j.1365-2923.2003.01577.x

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine*, *84*, 1022–1028. doi:10.1097/ACM.0b013e3181ace703

Cruttenden, A. (1986). *Intonation*. Cambridge, UK: Cambridge University Press.

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

De Groot, A. D. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

Ebbesen, E. B., & Konečni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology*, *32*, 805–821. doi:10.1037/0022-3514.32.5.805

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory, *102*(2), 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406. doi:10.1037/0033-295X.100.3.363

Ettenson, R., Shanteau, J., & Krogstad, J. (1987). Expert judgment: Is more information better? *Psychological Reports*, *60*, 227–238. doi:10.2466/pr0.1987.60.1.227

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 327–358. doi:10.1037/h0061470

Garzone, G., Catino, M., Gobo, G., Bait, M., Catenaccio, P., Degano, C., & Rozzi, S. (2010). Towards an integrated model for the understanding of communication failures in aviation accidents: Tenuous identities under pressure. In G. Garzone & J. Archibald (Eds.), *Discourse, identities and roles in specialized communication* (pp. 209–244). Bern, Switzerland: International Academic Publishers.

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*, 42–54. doi:10.1037/1076-898X.10.1.42

Hanna, D., Griswold, P., Leape, L. L., & Bates, D. W. (2005). Communicating critical test results: Safe practice recommendations. *Joint Commission Journal on Quality and Patient Safety*, *31*(2), 68–80.

Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. (1994). Repairing conversational misunderstandings and non-understandings. *Speech Communication*, *15*, 213–229. doi:10.1016/0167-6393(94)90073-6

Hoffman, K. A., Aitken, L. M., & Duffield, C. (2009). A comparison of novice and expert nurses' cue collection during clinical decision-making: Verbal protocol analysis. *International Journal of Nursing Studies*, *46*, 1335–1344. doi:10.1016/j.ijnurstu.2009.04.001

Hunter, M. (2004). The music is the message: Prosody in aviation discourse. In M. A. Turney (Ed.), *Tapping Diverse Talent in Aviation: Culture, gender, and diversity* (pp. 129–151). Hampshire, England: Ashgate Publishing Ltd.

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*, 146–154. doi:10.1016/j.learninstruc.2009.02.019

Kaempf, G. L., Klein, G. A., Thordsen, Marvin, L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors*, *38*, 220–231. doi:10.1518/001872096779047986

Kahneman, D., & Klein, G. A. (2009). Conditions for intuitive expertise: A failure to disagree. *The American Psychologist*, *64*, 515–526. doi:10.1037/a0016755

Kanki, B. G., & Smith, G. M. (2001). Training Aviation communication skills. In E. Salas, C. A. Bowers, & E. Edens (Eds.), *Improving Teamwork in Organizations* (pp. 31–54). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Klein, G. A. (1989). Recognition-primed decisions (RPD). *Advances in Man-Machine Systems*, *5*, 47–92.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138–147). Norwood, NJ: Ablex.

Klein, G. A. (1997). The Recognition-Primed Decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 285–292). Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G. A. (1999). *Sources of power*. Cambridge, Mass: MIT Press.

Klein, G. A. (2003). *Intuition at work.* New York: Currency Books.

Klein, G. A., & Calderwood, R. (1991). Decision models : Some lessons from the field. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*, 1018–1026. doi:10.1109/21.120054

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*, 462–472. doi:10.1109/21.31053

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Focus article: Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*, 331–352. doi:10.1002/bdm.381

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied*, *16*, 251–262. doi:10.1037/a0020082

Loveday, T., Wiggins, M. W., Festa, M., Schell, D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In P. Latorre Carmona, J. S. Sánchez, & A. L. N. Fred (Eds.), *Pattern recognition - Applications and methods* (Vol. 204, pp. 1–11). Berlin, Heidelberg: Springer Berlin Heidelberg.

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2013). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*, 90–107. doi:10.1177/0018720812450911

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2013). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners

in pediatric diagnosis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*, 125–137. doi:10.1177/0018720812448475

Monnot, M., Nixon, S., Lovallo, W., & Ross, E. (2001). Altered emotional perception in alcoholics: Deficits in affective prosody comprehension. *Alcoholism: Clinical and Experimental Research*, *25*, 362–369. doi:10.1111/j.1530-0277.2001.tb02222.x

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2012). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making*, *7*, 211–226. doi:10.1177/1555343412459192

Müller, S., Abernethy, B., & Farrow, D. (2006). How do world-class cricket batsmen anticipate a bowler's intention? *The Quarterly Journal of Experimental Psychology*, *59*, 2162–2186. doi:10.1080/02643290600576595

Myles-Worsley, M., Johnston, W. A, & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 553–557. doi:10.1037/0278-7393.14.3.553

Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: the role of experience. *Medical Education*, *41*(12), 1140–1145.

North, J. S., Williams, A. M., Hodges, N., Ward, P., & Ericsson, K. A. (2009). Perceiving patterns in dynamic action sequences: Investigating the processes underpinning stimulus recognition and anticipation skill. *Applied Cognitive Psychology*, *23*, 878–894. doi:10.1002/acp.1581

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*, 127–146. doi:10.1111/j.1551-6709.2008.01007.x

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1017–1030. doi:10.1037/0096-1523.34.4.1017

Paas, F. G. W. C., & Merriënboer, J. J. G. Van. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122–133. doi:10.1037/0022-0663.86.1.122

Prinzo, O. V., & Britton, T. W. (1993). *ATC/pilot voice communications: A survey of the literature* (Report No. DOT/FAA/AM-93/20). Washington, DC: Office of Aviation Medicine.

Prinzo, O. V., Hendrix, A. M., & Hendrix, R. (2006). *The outcome of ATC message complexity on pilot readback performance* (Report No. DOT/FAA/AM-06/25). Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

Randel, J. M., Pugh, H. L., & Reed, S. K. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human-Computer Studies*, *45*, 579–597. doi:10.1006/ijhc.1996.0068

Rasmussen, J. (1983). Skills, rules, knowledge: Signals, signs, and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*, 257–267. doi:10.1109/TSMC.1983.6313160

Rasmussen, J. (1993). Diagnostic reasoning in action. *IEEE Transactions on Systems Man and Cybernetics*, *23*, 981–992. doi:10.1109/21.247883

Schneider, V. I., Healy, A. F., & Barshi, I. (2004). Effects of instruction modality and readback on accuracy in following navigation commands. *Journal of Experimental Psychology: Applied*, *10*, 245–257. doi:10.1037/1076-898X.10.4.245

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise difference in attentional strategies related to pilot decision making. *Human Factors*, *50*, 864–878. doi:10.1518/001872008X374974.

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*, 253–263. doi:10.1016/S0377-2217(01)00113-8

Shim, J., Carlton, L. G., Chow, J. W., & Chae, W.-S. (2005). The use of anticipatory visual cues by highly skilled tennis players. *Journal of Motor Behavior*, *37*, 164–175. doi:10.3200/JMBR.37.2.164-175

Skantze, G. (2005). Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication*, *45*, 325–341. doi:10.1016/j.specom.2004.11.005

Soskin, W. F. (1953). Symposium: Communication in the counseling situation. *American Psychologist*, *8*, 272.

Stokes, A. F., Kemper, K., & Kite, K. (1997). Aeronautical decision making, cue recognition, and expertise under time pressure. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 183–196). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Stokes, A. F., Kemper, K. L., & Marsh, R. (1992). *Time-stressed flight decision making: A study of expert and novice aviators* (Report No. ARL-93-1/INEL-93-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Streitenberger, K., Breen-Reid, K., & Harris, C. (2006). Handoffs in care - Can we make them safer? *Pediatric Clinics of North Americalinics of North America*, *53*, 1185–1195. doi:10.1016/j.pcl.2006.09.010

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi:10.1016/0364-0213(88)90023-7

Weigand, E. (1999). Misunderstanding: The standard case. *Journal of Pragmatics*, *31*, 763–785. doi:10.1016/S0378-2166(98)00068-X

Wellens, A. R. (1993). Group situation awareness and distributed decision-making: From military to civilian applications. In J. Castellan (Ed.), *Individual and group decision making: Current issues* (pp. 267–291). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wickens, C. D., & Flach, J. M. (1988). Information Processing. In E. L. Wiener & D. C. Nagel (Eds.), *Human Factors in Aviation* (pp. 111–155). San Diego, CA: Academic Press.

Wiggins, M. W. (2006). Cue-based processing and human performance. In W. Karwowski (Ed.), *Encyclopedia of Ergonomics and Human Factors* (2nd ed., pp. 641–645). London, UK: Taylor and Francis.

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science*, 1–10. doi:10.1080/1463922X.2012.724725

Wiggins, M. W. (2014). Differences in situation assessments and prospective diagnoses of simulated weather radar returns amongst experienced pilots. *International Journal of Industrial Ergonomics*, *44*, 18–23. doi:10.1016/j.ergon.2013.08.006

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science*, *65*, 118–124. doi:10.1016/j.ssci.2014.01.006

Wiggins, M. W., Brouwers, S., Davies, J., & Loveday, T. (2014). Trait-based cue utilization and initial skill acquisition: implications for models of the progression to expertise. *Frontiers in Psychology*, *5*, 1–8. doi:10.3389/fpsyg.2014.00541

Wiggins, M. W., Harris, J., Loveday, T., & O'Hare, D. (2010). EXPERTise (Software Package). Sydney, Australia: Macquarie University.

Wiggins, M. W., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied*, *1*, 305–320. doi:10.1037/1076-898X.1.4.305

Wiggins, M. W., & O'Hare, D. (2003a). Expert and novice pilot perceptions of static in-flight images of weather. *The International Journal of Aviation Psychology*, *13*, 173–187. doi:10.1207/S15327108IJAP1302_05

Wiggins, M. W., & O'Hare, D. (2003b). Weatherwise: Evaluation of a cue-based training approach for the recognition of deteriorating weather conditions during flight. *Human Factors*, *45*, 337–345. doi:10.1518/hfes.45.2.337.27246

Wigton, R. S., Hoellerich, V. L., & Patil, K. D. (1986). How physicians use clinical information in diagnosing pulmonary embolism: An application of conjoint analysis. *Medical Decision Making*, *6*, 2–11. doi:10.1177/0272989X8600600102

Williams, A. M., Hodges, N. J., North, J. S., & Barton, G. (2006). Perceiving patterns of play in dynamic sport tasks: Investigating the essential information underlying skilled performance. *Perception*, *35*, 317–332. doi:10.1068/p5310

Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied*, *8*, 259–270. doi:10.1037/1076-898X.8.4.259

Paper 3 Appendix A

Semantic Content of Communicative Exchanges

Communicative Exchange	Instruction	Readback
1	Part four of the form is to be completed by sixteen hundred hours on the third of June, two thousand and thirteen.	It is understood that part four of the Form B must be completed by sixteen hundred hours.
2	Electrical section one oh two is de-energised; you may now open Isolator one oh two slash three.	I understand that I am authorised to open isolator one oh two slash three.
3	I want you to proceed to Rockland's Feeder Station and give me a call when you arrive.	Got that John; I will call you when we get to Rockland's Feeder Station.

Paper 3 Appendix B

Paired Association Task Items

Sequential Pairs		Simultaneous Pairs	
Feature	Event	Feature	Event
Outgoing Feeder	Fault	Qualitrol	Return Conductor
Traction Bonds	Train	Live Wire	Danger
Re-close	Emergency	Switch Gear	Evacuation
Wildlife	Power Supply	Midpoint Anchor	Power Transformer
High Winds	De-energize	Traction Fault	Explosion
High Winds	Safety	Double Trip	Earth
RC Bar	Earth Grid	Earth Wire	Bonding
Blockage	Voltage	Switch	Current
Return Conductor	Qualitrol	Fault Locators	Traction Fault
Autotransformer System	40kV-55kV	Switch	Re-energize
Catenary Wire	Contact Wire	Auxiliary Supply	Reclose
Cable Fault	Injury	Dewirement	Traction Fault
Displaced Conductors	Injury	Isolate	Explosion
Catenary Wire	De-energize	Track Coupler Unit	Return Current Conductor
Circuit Breaker	Tripped	Earth Wire	Autotransformer Feeder Wire
Permit to Work	Tripped	Arching	Fire
Cable Fault	Power Supply	Pantograph	Power
Permit to Work	Safety	Speed Restriction	Electrical Shock
Overheads	Earth Grid	Fault Locators	Current
Overheads	Energized	Traction Earthing	Autotransformer Feeder Wire
Cable Differential Fault	Distance Protection Trip	Pantograph	Hard Fault
Displaced Conductors	40kV-55kV	Dewirement	Power
Track Block	Harmonic Filler	Traction Fault	Fire
RC Bar	Trip	Isolate	Earth
Autotransformer System	Energized	Arching	Re-energize
Wildlife	Trip	Double Trip	Hard Fault
Circuit Breaker	Contact Wire	Traction Earthing	Bonding

Paper 3 Appendix C

Feature Discrimination Task (1)

Scenario

It is 1402 on a Tuesday afternoon in late November. You receive a phone call from one of your experienced substation electricians who is in attendance at a switching station. This particular switching station has only been recently commissioned and you know that there have been some difficulties in the commissioning process. The electrician tells you that he is standing adjacent to the switchgear and can hear a loud noise emanating from it. He describes it as a distinctive crackling or buzzing sound, like a swarm of bees. The volume of the noise appears to vary in conjunction with the train movement. He also mentions that he has noticed a harsh sort of smell. You ascertain from him that there are several substation staff present on the scene. He is nonchalant in the way he relays this information to you, but seeks your advice on how to proceed.

Response Options

From the options available, determine the **first step** you would take to manage this situation. You may only select one response. If you need to review the scenario again, please click Back. Press Continue when you have made your selection.

- Instruct the substation electrician to investigate the cause of the sound and report back any findings
- Make necessary arrangements to stop all trains on the associated electrical sections to determine if that makes any difference to the sound
- Immediately take the switching station out of service and arrange for alternative feeding arrangements for the affect electrical sections
- Do nothing

Ratings of Feature Utility

Rate the relevance of the following features in informing your response.

- Time of day
- Risk of commercial loss to company as a result of action taken
- Month of year
- Experience of substation electrician on the phone
- Recency of substation commissioning
- Description of the noise emanating from the switch gear
- Presence of other substation staff on the scene
- Day of week
- Position of the substation electrician when he heard the noise from the switchgear
- Description of the smell emanating from the switch gear

Paper 3 Appendix D

Feature Discrimination Task (2)

Scenario

At 1410 on a Saturday in mid-July, you receive SCADA indications of a Harmonic Filter Overcurrent trip on HF1 at Rangal Feeder Station. Below is an excerpt of the Events Summary at the time of the incident for your reference. The weather on site is fine with a temperature of approximately 26 degrees centigrade. The coalmines are in full production and there are three electric trains operating on the associated outgoing feeders at the time of the fault. It is noted that bus voltage and incoming feeder current monitoring equipment has been out of service since May this year after a 132kV fault at Rangal Feeder Station. Overhead callout line staff are in the area inspecting overhead line equipment after an earlier outgoing feeder trip. Please press Continue when you have determined the first step you would take to manage this situation.

Rockhampton

System ▾ Section ▾ Outstation ▾ Comms ▾ Alarms ▾ Hist Events ▾ Events Trends ▾ Event Log

RANGAL Urgent NonUrgent Information

Date	Time	Description	Type	Action
13/07/2013	14:16:26.575	RANGAL HF1 CB CLOSE INHIBIT ALARM	N	OFF
13/07/2013	14:10:58.719	RANGAL RL HF1 CB OPEN	U	ON
13/07/2013	14:10:58.718	RANGAL HF1 CB CLOSE INHIBIT ALARM	N	ON
13/07/2013	14:10:58.705	RANGAL HF OVERCURRENT TRIP	U	OFF
13/07/2013	14:10:58.671	RANGAL HF1 TRIP RELAY LATCHED ALARM	N	ON
13/07/2013	14:10:58.669	RANGAL HF OVERCURRENT TRIP	U	ON
13/07/2013	11:07:28.700	RANGAL NO 2 AUX SUPPLY FEED DESELECTED	I	OFF
13/07/2013	11:07:28.679	RANGAL RL 153 CB CLOSED	U	ON
13/07/2013	11:07:28.676	RANGAL NO 2 AUX SUPPLY 240 V FAILED	U	OFF
13/07/2013	11:07:28	RANGAL RL 153 CB Command Sent	O	CLOSE CMD
13/07/2013	11:07:23	RANGAL RL 153 CB SCA Issued	O	CLOSE CMD
13/07/2013	09:36:14.893	RANGAL RL 152 CB CLOSED	U	ON
13/07/2013	09:36:14	RANGAL RL 152 CB Command Sent	O	CLOSE CMD
13/07/2013	09:35:19.806	RANGAL RL 152 CB OPEN	U	ON
13/07/2013	09:35:19	RANGAL RL 152 CB Command Sent	O	OPEN CMD
13/07/2013	09:34:04.191	RANGAL NO 2 AUX SUPPLY FEED DESELECTED	I	ON
13/07/2013	09:34:04.186	RANGAL NO 2 AUX SUPPLY 240 V FAILED	U	ON
13/07/2013	09:34:04.143	RANGAL RL 153 CB OPEN	U	ON
13/07/2013	09:34:03	RANGAL RL 153 CB Command Sent	O	OPEN CMD
13/07/2013	06:50:47.690	RANGAL FL151_152 COMMUNICATIONS FAILURE	N	OFF
13/07/2013	06:50:43.690	RANGAL FL151_152 COMMUNICATIONS FAILURE	N	ON
13/07/2013	06:50:37.780	RANGAL FL151_152 COMMUNICATIONS FAILURE	N	OFF
13/07/2013	06:50:33.780	RANGAL FL151_152 COMMUNICATIONS FAILURE	N	ON
12/07/2013	19:43:10	RANGAL TFMR T15 IF AMPS HIGH	N	OFF
12/07/2013	19:42:35	RANGAL TFMR T15 IF AMPS HIGH	N	ON

Response Options

From the options available, determine the **first step** you would take to manage this situation. You may only select one response. If you need to review the scenario again, please

click Back. Press Continue when you have made your selection.

- Immediately arrange for substation electrician staff to attend the site to inspect the harmonic filter and interrogate the protection relays for fault data.
- Wait 5 minutes and then attempt to reclose the Harmonic Filter Circuit Breaker, RL/HF1.
- Allow for affected rail traffic to clear the section without the Harmonic Filter, HF1 in service and then apply a block to electric trains on the associated outgoing feeders.
- Immediately apply a control inhibit to the tripped circuit breaker RL/HF1 and establish alternative feeding arrangements.
- Arrange for Overhead line staff to inspect the harmonic filter. Do nothing else but continue to monitor the situation pending further reports.

Ratings of Feature Utility

Rate the relevance of the following features in informing your response.

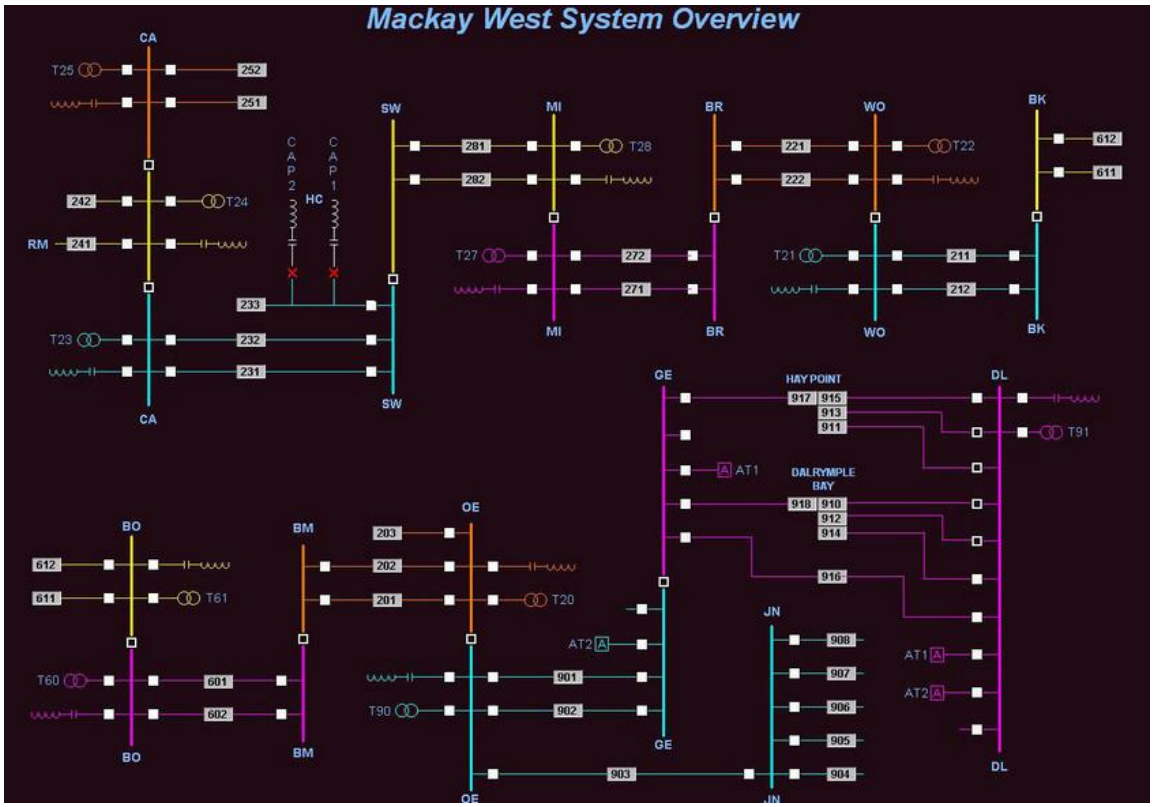
- Type of fault
- Weather conditions at time of fault
- Number of trains on the section at the time of fault
- Presence of Overhead line staff in the area
- Current coal mine production
- The availability of bus voltage and incoming feeder current fault data
- Recent fault history at Rangal Feeder Station
- Time fault occurred
- High alarms that occurred at 19:42:35
- Open circuit breaker at Rangal at 14:10:58
- Day of week fault occurred
- Month of year fault occurred

Paper 3 Appendix E

Transition Task

Scenario

You have just received indications of a traction fault on electrical section 281 between Mindi FS and South Walker TSC. After carrying out the procedures prescribed in the Electric Control Room Instructions, you find that the electrical section trips on an attempted reclose. See below SCADA screen of Mackay West System Overview during normal operation (rather than during the incident described) for your reference. Click on the image below to view a larger version of the SCADA screen. Use the background information to determine your next response.



Information Categories

- Length of electrical section 281
- Fault locator reading
- Number of line staff working on electrical section 281
- Number of trains on electrical section 281

- Day of week
- Weather
- Cause of initial trip
- Reports from train on electrical section 281
- SCADA representation of electrical section 281 after first trip
- SCADA system qualifiers
- Time of day
- Cause of second trip
- Events Summary at time of first trip
- Excerpt from UTC screen in area of fault

Response Options

From the options available, determine the **first step** you would take to manage this situation. You may only select one response. If you need to review the scenario again, please click Back. Press Continue when you have made your selection.

- Wait six minutes and attempt a second reclose.
- Dispatch call-out staff to patrol the faulted electrical section.
- Commence sectionalising the faulted area.
- Contact Powerlink and request the de-energisation of all traction power supplies to the area.
- Immediately stop all trains within the faulted section and arrange for a blockage to all traffic to be applied to the affected area.
- Advise the call-out engineer of the trip on reclose and request further instructions.
- Do nothing. Reassess the situation in six minutes.

Paper 4: Prosodic and Semantic Cue Utilisation in Perceptions of Nonunderstandings

Paper 4 Aims

The outcomes of Paper 3 indicated that intonation and filler had utility as prosodic cues during assessments of nonunderstanding within electrical rail control. In conjunction with the results of Paper 1, these findings support the notion that intonation and filler are used to assess speaker understanding of a full readback response. A primary aim of Paper 4 was to assess whether such cues have utility within the domain of hydroelectric power generation control (a domain which involves the production of electrical power through the use of flowing water). Such a finding would provide further support for the function of particular prosodic cues to aid in the recognition of concealed nonunderstandings during full readback responses in a variety of operational environments.

An additional outcome of Paper 3 suggested that the semantic content of the full readback influenced perceptions of nonunderstanding. As many operational environments, such as hydroelectric power generation, use a combination of full and partial readbacks, Paper 4 also examined whether prosodic cue use was associated with the semantic content of the instruction response type (i.e., partial readback vs. full readback). This would provide a greater understanding of the contexts in which prosodic (and semantic) cues are used to assess nonunderstandings in readback response that could have implications for training and system design in distributed environments.

Paper 3 found no relationship between visual cue utilisation performance and prosodic cue utilisation during perceptions of nonunderstanding *even when* cognitive resources were restricted. In combination with the outcomes of Paper 2, these results suggest that prosodic cue utilisation during assessments of nonunderstanding is more likely to represent a general skill acquired through non-operational interactions. A final aim of Paper 4 was to examine whether there was any relationship between visual cue utilisation performance and prosodic cue use in a context more representative of the operational environment (i.e., with the use of

full and partial readback responses). Paper 4 also incorporated the revised version of the Auditory Readback Task (including the concurrent Numeric Digit Span Recall Task). See Appendix C of the thesis for a more detailed description of the development of this task for the auditory stimuli used in Paper 4.

As with the domains selected for Papers 1, 2 and 3, hydroelectric power generation control was selected as the domain of investigation as it is arguably a context in which auditory and visual cue associations are expected to develop on the basis of experience and where such cue utilisation can be measured with a degree of ecological validity in an experimental context. Power generation operators were selected as participants. The nature of this role requires operators to follow the readback/hearback procedure when communicating over the radio/telephone using a combination of full and partial readbacks, and to monitor and diagnose system events through Supervisory Control And Data Acquisition (SCADA) visual interfaces and intervene where necessary to safeguard the supply of electricity.

Papers 3 and 4 were similar insofar as they both required manipulation checks to be carried out to validate the final intonation contours of the readback responses used in the Auditory Readback Task within each domain, in addition to the number of single digits to be used in the Numeric Digit Span Recall Task within each domain. While the stimuli and analyses used in the intonation manipulation checks and the digit span manipulation checks were distinct across the two papers, the intonation manipulation check (for both Papers 3 and 4) was conducted with the same post-graduate cohort. Similarly, the digit span manipulation check for both Papers 3 and 4, was conducted with the same undergraduate cohort.

Paper 4 Publication History

Paper 4 was submitted for publication in the Journal of Experimental Psychology: Applied on the 5th September 2014 and is currently under review. The Journal of Experimental Psychology: Applied has an impact factor of 2.426 and ranks 12th of 75 journals in the Psychology: Applied category for the 2013 Journal Citation Report. Paper 4 presented

in the current thesis is the manuscript that was submitted for publication. Due to length constraints for the submitted manuscript, Appendix D (to the thesis) includes additional resources to accompany this Paper (namely, the *EXPERTise 1.0* stimuli).

A modified version of Paper 4 was also accepted for publication in the Proceedings of the *58th Annual Meeting of the Human Factors and Ergonomics Society (HFES), 2014*. The author of the current thesis will present the paper at the HFES meeting in Chicago on October 28th 2014. This modified version investigated the relationship between native language, intonation and instruction response type on assessments of nonunderstanding and included a slightly larger sample of hydroelectric power generation operators compared to the sample described in Paper 4. Although the proceedings have yet to be published, the formatted version for publication is provided in Appendix E (N.B. the original submission has been subsequently removed from the Appendices for copyright purposes). Both Paper 4 and the HFES paper use American, rather than Australian, English.

As the focus of Paper 4 and the HFES Paper is largely on communication, rather than the theoretical process of situation assessment, neither paper includes an extensive discussion of the role of cue utilisation in the assessment of nonunderstandings. As such, the term ‘prosodic cue’ is used exclusively rather than the more nuanced distinction between prosodic features and cues. The author of the current thesis wrote approximately 80% of Paper 4 and 90% of the HFES paper.

Utilization of Prosodic and Semantic Cues during Perceptions of Nonunderstanding in Radio Communication

Jaime C. Auton[1], Mark W. Wiggins[1], Ben J. Searle[1] & Nan Xu Rattanasone[2]

1 Department of Psychology, Macquarie University, Sydney, Australia

2 Department of Linguistics, Macquarie University, Sydney, Australia

Author Note

Correspondence concerning this article should be addressed to Jaime Auton, Department of Psychology, Macquarie University, Sydney, New South Wales, Australia, 2109. E-mail: jaime.auton@mq.edu.au

Abstract

The readback/hearback loop is a communicative protocol used in many high-risk environments to ensure that a verbal instruction has been *heard* correctly by a receiver. However, it does not necessarily ensure that an instruction has been *understood*. Using an international sample of hydroelectric power generation controllers, the current study investigated whether particular semantic and prosodic cues contained within a readback response could signal to listeners, the extent to which speakers had understood an instruction. A secondary aim was to establish whether the use of such cues was associated with participants' relative levels of visual cue utilization. The results indicated that different prosodic cues are used to detect nonunderstandings, depending upon the accuracy of the semantic content of the readback. Further, greater levels of visual cue utilization were associated with relatively greater emphasis on the use of a 'filler' and the semantic content of the readback during perceptions of nonunderstanding. The results have implications for training and system design in distributed environments.

Keywords: prosodic cues, semantic cues, perceptions of nonunderstanding, spoken communication, cue utilization.

Utilization of Prosodic and Semantic Cues during Perceptions of Nonunderstanding in Radio Communication

The safe and efficient management of many high-risk technical environments is reliant upon effective communication between operators (Australian Transport Safety Bureau, 2007). These environments, including air traffic control, hydroelectric power generation and emergency services, are characterized by time pressure, high stakes, and uncertain and dynamic conditions, in which any error has the potential for a significant loss (Orasanu & Connolly, 1993). Communication has been a contributing factor to a significant proportion of errors within these industries (Kim, Park, Han, & Kim, 2010). For example, Billings and Cheaney (1981) reported that information transfer problems between operators were a contributing factor in over 70% of the incidences reported to the National Aeronautics and Space Administration (NASA) Aviation Safety Reporting System. Similarly, Greenberg et al. (2007) noted that almost a quarter of surgical malpractice claims resulting in patient harm, were a result of erroneous communication.

Many of these industries have implemented various defence mechanisms in an attempt to reduce operator error and maintain the integrity of communication processes. The readback/hearback loop is one of these mechanisms that specifies how verbal instructions are to be delivered and confirmed over the radio/telephone (Prinzo & Britton, 1993). This is a four-step protocol that is initiated when a sender transmits a verbal message to a receiver. For example, an Air Traffic Controller (ATC) might send the following instruction to a pilot; "*Aircraft X, turn left to a heading of zero three zero*" (Barshi & Farris, 2013, p. 2). The receiver is then required to actively listen to the instruction before repeating the instruction verbatim back to the sender (known as the *readback*). In the aviation context, a pilot would respond to the ATC with "*Left to zero three zero, Aircraft X*". The sender is then expected to listen to the readback to ensure that it has been repeated correctly, which is referred to as the *hearback* portion of the exchange (Cushing, 1995; Prinzo & Britton, 1993).

The main purpose of the readback/hearback loop is to ensure that the receiver has accurately heard the instruction, as the sender is able to identify and correct any errors made in the readback during the hearback portion of the exchange (Schneider, Healy, & Barshi, 2004). However, this protocol does not directly prevent errors that can arise from a sender not understanding an instruction, since a receiver may have the ability to repeat words in the correct sequence but derive no meaning from these words (Jones, 2003).

Instances where a receiver does not obtain any meaningful interpretation of an instruction or is unable to select from two or more interpretations of an instruction is referred to as a *nonunderstanding* (Hirst, McRoy, Heeman, Edmonds, & Horton, 1994; Skantze, 2005). Distinct from a *misunderstanding*, a nonunderstanding involves conscious awareness of a lack of understanding. In the case of a misunderstanding, receivers are under the incorrect assumption that they have understood the instruction as intended and in such cases, receivers remain unaware that an error in discourse has occurred (Hirst et al., 1994; Skantze, 2005).

There is an explicit expectation in many high-risk industries that a receiver will signal instances of nonunderstanding to the sender, thereby allowing the sender to clarify the instruction. For example, in the Aeronautical Information Manual (AIM; Federal Aviation Administration, 2014, Section 5-5-2), it specifies that the receiver must "request clarification or amendment, as appropriate, any time a clearance is not fully understood or considered unacceptable from a safety standpoint". However, receivers can elect to deceptively conceal their state of nonunderstanding from the sender (Weigand, 1999), particularly to save face in those situations where there is an expectation that the instruction would normally be understood. Although there is a common belief in many industries that the receipt of an accurate readback is sufficient evidence of understanding on the part of the receiver (International Civil Aviation Organization, 2007), nonunderstandings, in spite of an accurate readback, do occur (Air Accidents Investigation Branch, 2008; Billings & Cheaney, 1981; Morrow, Rodvold, & Lee, 1994).

The identification of effective strategies that could improve the detection of nonunderstandings during readback responses is an important step in improving the integrity of communication systems within technical environments. Auton, Wiggins, Searle, Loveday, and Xu Rattanasone (2013) initially addressed this issue by investigating whether there were particular prosodic cues contained within a receiver's accurate full readback that could signal to the sender the extent to which the receiver had understood the instruction. Prosodic cues refer to the range of paralinguistic characteristics of the voice, such as tempo and intonation, that influence *how* something is said (Bolinger, 1985). In contrast, semantic cues refer to the syllables, words or phrases of an utterance that comprise *what* is said (Nygaard & Queen, 2008).

Situated within the context of electricity transmission control, Auton et al. (2013) asked network operators to listen to a series of auditory-only, simulated readback/hearback communicative exchanges. During each exchange, participants heard a sender delivering an instruction to a receiver, followed by the receiver delivering an accurate readback back to the sender. Auton et al. systematically manipulated three prosodic cues contained within the readback (specifically intonation, filler and inter-delay, which had been previously associated with perceptions of uncertainty; Brennan & Williams, 1995), to determine whether these cues signalled to a participant (i.e., 'listener'), the extent to which a receiver (i.e., 'speaker') had understood an instruction. After listening to each exchange, participants were then asked to rate the extent to which they perceived that the receiver had understood the instruction based on the way the readback had been spoken.

Auton et al. (2013) found that when the prosodic cues of intonation and filler were contained within the readback portion of the exchange, they signalled a receiver's state of understanding to a participant. Intonation describes the pitch patterns or 'melody' of an utterance, where there is a basic division between tunes falling in pitch (i.e., falling intonation) and those rising in pitch (i.e., rising intonation; Cruttenden, 1981; O'Connor &

Arnold, 1961). A filler is an interjection in speech, such as *um* or *uh* (Clark & Fox Tree, 2002). In the case of intonation, participants perceived that the receiver had a lower level of understanding of the instruction when the readback ended with rising intonation in comparison to falling intonation. For filler, participants perceived readbacks that were preceded by the filler *um* as signalling a relatively lower level of understanding on the part of the receiver, compared to readbacks that were not preceded by a filler.

Although Auton et al. (2013) demonstrated that prosodic cues such as intonation and filler were useful to network operators in detecting varying levels of understanding on the part of the receiver when a full readback was articulated, they did not consider whether such cues would be relied upon when a receiver responded with a partial readback. In a partial readback response, the receiver does not repeat all the information in the instruction provided by the sender (Barshi & Farris, 2013).

Although there is an explicit expectation in many high-risk environments that operators provide a full readback to an issued instruction (i.e., a receiver is to repeat the message back to the sender in full), procedural variations occur frequently (Garzone et al., 2010). For example, Morrow, Lee, and Rodvold (1993) observed that up to 13% of the routine pilot communication they analyzed contained partial, rather than full, readbacks. Similarly, Prinzo, Hendrix, and Hendrix (2006) found that pilots used full readbacks on 82.7% of occasions. If the prosodic cues, such as intonation and filler, have the potential to be used as a training tool to improve operator detection of nonunderstandings, it is important to establish whether the use of such cues can be generalized across non-routine responses to instructions (i.e., during partial readback responses).

Compared to a full readback, the receipt of a partial readback provides the sender with less semantic evidence that an instruction has been understood. It is unclear from Auton et al. (2013) whether the utility of intonation, filler, and less so inter-turn delay, only exists when the precondition of a full readback is met. Evidence from Brennan and Williams' (1995)

investigation into the usefulness of prosodic cues during perceptions of uncertainty suggests that intonation, filler and inter-turn delay are still used by listeners to gauge a speaker's level of certainty, even when the listener does not have evidence that the speaker's semantic response is accurate. Although Brennan and Williams investigated uncertainty rather than nonunderstandings associated with communication exchanges, their evidence suggests that intonation, filler and inter-turn delay can still be utilized by listeners to gauge a speaker's state of understanding even in the absence of supporting semantic evidence. However, this proposition has not been empirically tested within the framework of nonunderstandings and the readback/hearback protocol.

In addition to examining the usefulness of prosodic cues during perceptions of nonunderstanding, it is also important to consider how semantic cues contained within a response influence perceptions of nonunderstanding. Although there is a common assumption that the receipt of a full readback from a receiver can be considered sufficient evidence of understanding (International Civil Aviation Organization, 2007), it is important to empirically test this assumption, as the readback procedure has been criticized as unnecessary and time-consuming (Schneider et al., 2004). If it is evident that a full readback does not provide a listener with additional evidence that the receiver has understood the instruction, it supports the position that the requirement for a full readback may need to be reconsidered.

The aim of this study was the examine how both semantic and prosodic cues contained within a receiver's response to an instruction, influence how the receiver is perceived to have understood the instruction. It was hypothesized that full readback responses to instructions would receive higher ratings of understanding compared to partial readback responses to instructions (H1). It was also predicted that full readback responses ending in rising intonation would attract lower ratings of understanding of the instruction, compared to readback responses ending in falling intonation (H2a). Similarly, it was predicted that partial readback responses ending in rising intonation would receive lower ratings of understanding

of the instruction compared to partial readback responses ending in falling intonation (H2b).

For the prosodic cue ‘filler’, it was hypothesized that readback responses preceded by a filler would attract lower ratings of understanding of the instruction compared to full readbacks that were not preceded by a filler (H3a). Similarly, partial readback responses preceded by a filler were predicted to attract lower ratings of understanding of the instruction compared to partial readbacks that were not preceded by a filler (H3b).

Although Auton et al. (2013) found that inter-turn delay was not a salient cue that electrical network operators used during perceptions of nonunderstanding of full readback responses, it was unclear whether inter-turn delay would be useful in a different domain and/or when the semantic cues associated with a full readback were unavailable (i.e., during partial readback responses). Therefore, it was hypothesized that readback responses that commenced after a longer inter-turn delay would attract lower ratings of understanding of the instructions compared to readback responses that commenced after a shorter inter-turn delay (H4a). Finally, it was predicted that partial readback responses that commenced after a longer inter-turn delay would attract lower ratings of understanding of the instruction compared to partial readback responses commencing after a shorter inter-turn delay (H4b)

Auton et al. (2013)’s investigation was based on the assumption that there was no variation in the way that operators utilize prosodic cues when assessing potential instances of nonunderstanding. However, research into the use of cues in operational contexts has shown that experienced operators can be reliably discriminated from their less experienced counterparts by their greater ability to recognize and utilize task relevant cues during situation assessment (Coderre, Mandin, Harasym, & Fick, 2003; Müller, Abernethy, & Farrow, 2006; Schriver, Morrow, Wickens, & Talleur, 2008; Stokes, Kemper, & Marsh, 1992). *Situation assessment* is a process that involves understanding and identifying the current state of a situation by attending to, and interpreting the available environmental cues (Klein & Calderwood, 1991; Wickens & Hollands, 2000).

Experienced operators are better able to recognize patterns of cues that are indicative of an event, as they have accumulated greater domain-relevant experience or deliberate practice that has strengthened the association between a cue and an event/object in long term memory (Wickens & Hollands, 2000; Wiggins, 2006). For example, experienced miners have been shown to possess a superior ability to discriminate between dangerous and safe rock conditions compared to less experienced miners (Blignaut, 1979). Similarly, highly skilled batsmen show a superior ability to attend to specific, early cues (such as bowling hand and arm cues) to predict the opposing bowlers' intentions compared to intermediate and low-skilled batsmen (Müller et al., 2006).

This evidence would suggest that those operators who have a greater capacity to utilize visual cues during situation assessment, would also have developed a more nuanced repertoire of auditory (particularly prosodic and semantic) cues to assess instances of nonunderstanding, compared to those operators with a lesser capacity for cue utilization. It is important to investigate the way in which operators with greater levels of cue utilization attended to both semantic and prosodic cues during the assessment of nonunderstandings in radio communication, as it would highlight the most salient auditory cues to assist in the recognition of nonunderstandings.

It was, therefore, predicted that, in comparison to participants with lesser levels of cue utilization, participants with greater levels of cue utilization would perceive a greater difference in levels of understanding on the part of the receiver between a full readback response compared to a partial readback response (H5). It was also predicted that, in comparison to participants with lesser levels of cue utilization, participants with greater levels of cue utilization would perceive a greater difference in levels of understanding on the part of the receiver between readback responses ending in rising intonation compared to responses ending in falling intonation (H6). In comparison to participants with lesser levels of cue utilization, it was hypothesized that participants with greater levels of cue utilization would

perceive a greater difference in levels of understanding on the part of the receiver between readback responses preceded by a filler compared to those responses not preceded by a filler (H7). Finally, it was predicted that, compared to participants with lesser levels of cue utilization, those participants with relatively greater levels of cue utilization would perceive a greater difference in levels of understanding on the part of the receiver between a shorter compared to a longer inter-turn delay (H8).

The Present Study

The present study was undertaken within the domain of hydroelectric power generation, and involved controllers whose role was to monitor and diagnose system events and intervene where necessary to safeguard the supply of electricity. Power generation control was selected on the basis that operators in this environment generally follow the readback/hearback procedure when communicating over the radio/telephone. However, unlike aviation, the communication process involves a combination of full and partial readbacks, depending upon the circumstances.

Participants were asked to listen to a series of scripted audio communicative exchanges through an online experimental platform that reflected different forms of the readback/hearback protocol. Each exchange included a sender delivering an instruction and a receiver responding with either a full or partial readback of the instruction. The readbacks systematically varied in the type of intonation heard at the end of the response, the presence of a filler before the commencement of the response, and the duration of the inter-turn delay between speakers. After listening to each exchange, participants were asked to indicate the extent to which they perceived the receiver (who delivered the full or partial readback) had understood the instruction. Power generation controllers also completed the EXPERT intensive skills evaluation (*EXPERTise 1.0*) situational judgement test, which evaluated and classified them into one of two typologies based on their aggregated scores across four tasks that were associated with visual cue utilization (Wiggins, Harris, Loveday, & O'Hare, 2010).

Method

Participants

Twenty-four hydroelectric power generation operators from companies worldwide participated in the study. Participants were recruited from either online advertisements within a major United States (US) power generation magazine or via direct contact through a major hydroelectric power generation company in New Zealand. Overall, a significant proportion (37.5%) of participants were from the US, with a further 29.2% from New Zealand. The remaining participants resided in various countries including England, Australia, Italy, and Croatia.

The sample was predominantly male (91.7%) and native English speaking (83.3%). The participants ranged in age from 29 to 74 years (M = 51.83, SD = 9.47) and had accumulated between four and 44 years of experience within power generation (M = 24.65, SD = 11.45). Within the specialized field of hydroelectric power generation, participants had accumulated between two and 44 years of experience (M = 19.79, SD = 11.79). Participants received a US$25 Amazon eVoucher in exchange for their completion of the study.

Materials

The participants were asked to complete two distinct tasks. The *Auditory Readback Task* assessed the influence of prosodic and semantic cues on participants' perceptions of nonunderstandings across a series of readback/hearback exchanges. The *EXPERTise 1.0* battery required participants to complete a series of visual, cue-based tasks within hydroelectric power generation, and based on their performance across the tasks, participants were categorized into two typologies; those with relatively greater levels of cue utilization and those with relatively lesser levels of cue utilization.

Auditory Readback Task.

Design.

The Auditory Readback Task incorporated two levels of each of the following within-

groups variables: communicative exchange (Exchange 1, Exchange 2); Instruction Response Type (Full Readback, Partial Readback); Intonation (Rising, Falling); Filler (Present, Absent), and, Inter-turn Delay (Shorter, Longer). This resulted in 32 distinct communicative exchanges to which participants listened in random order.

After listening to each exchange, participants were asked to rate, on a 7-point scale from 1 (*did not understand at all*) to 7 (*understood completely*), the extent to which they perceived that the receiver had understood the meaning of the instruction. This type of scale was consistent with Brennan and Williams' (1995) measure of perceptions of uncertainty and Auton et al.'s (2013) measure of perceptions of nonunderstanding. These ratings represented the dependent variable.

The design incorporated two semantically distinct communicative exchanges to ensure that the semantic content of the exchanges did not influence prosodic cue use. Each communicative exchange consisted of a sender delivering an operational instruction to a receiver and a receiver responding to the instruction. The two communicative exchanges used in the current study differed in the type of operational tasks that the sender requested the receiver to execute. The Appendix lists the semantic content of the two communicative exchanges.

'Instruction Response Type' referred to the nature of the response to the operational instruction in each of the two communicative exchanges. Upon hearing the instruction delivered by the 'sender', the receiver either responded with a full readback (i.e., the entire message was repeated back to the sender) or a partial readback (i.e., receipt of the instruction was acknowledged). For example, the instruction contained within Communicative Exchange 1, read; *Hi Steve, Can you go to Rangipo Station and check the status of the bypass valves on the two generators and give me a call?* The response to this instruction was either a full readback (*Thanks Pete. I'm on my way to Rangipo right now and will check the status of the bypass valves and give you a call*) or a partial readback, (*Thanks Pete, I'll do that now*).

‘Intonation’ referred to the end of the receiver’s response, and specifically, whether it ended in a rising pitch (rising intonation) or falling pitch (falling intonation). ‘Filler’ referred to the presence or not of a verbal interjection (i.e., *um*) prior to the commencement of the response. Finally, ‘Inter-turn delay’ referred to whether the delay between the sender delivering the instruction and the receiver responding to the instruction was either shorter (600 ms) or longer (1200 ms).

Stimuli development.

Communicative exchanges.

The semantic content of the two communicative exchanges was developed in consultation with two Subject Matter Experts (SME) from a major hydroelectric power generation company in New Zealand. The SMEs were selected on the basis of peer assessment and industry position. Each exchange consisted of both an instruction that was likely to occur within the hydroelectric power generation industry, specifying the operational steps to be executed by the receiver, as well as a response (either a full readback or a partial readback).

The auditory versions of the communicative exchanges were recorded by two Australian English male speakers using a digital audio editor program (*GarageBand*™) on a MacBook™ laptop. One speaker assumed the role of the sender, who delivered the instruction in both exchanges and the other speaker assumed the role of the receiver, who delivered the instruction responses associated with both exchanges. Speakers maintained their allocated roles as either sender or receiver across both exchanges to prevent the influence of differences in voice timbre affecting ratings of understanding. The duration of the communicative exchanges (including both the instruction and response) ranged from 10 to 13 seconds.

Prosodic manipulation of instruction response types.

The instruction response portion of each communicative exchange was manipulated to systematically vary the type of intonation at the end of the response, whether the receiver used

a filler (i.e., *um*) before responding to the instruction, and the duration of the inter-turn delay between the offset of the sender's instruction and the onset of the receiver's response.

To generate responses that either ended with rising or falling intonation, the speaker who played the role of the receiver, was instructed to end the instruction response token with either rising or falling intonation. A manipulation check was conducted to ensure that participants perceived the manipulation of intonation accordingly (see Manipulation Checks section). To create inter-turn delays of either a shorter (600 ms) or a longer (1200 ms) duration, artificial silences of the two durations, were electronically inserted between the offset of the sender delivering the instruction and the onset of the receiver's response, using the digital audio editor program GarageBand™.

The receiver's response was either preceded by a filled inter-turn delay (i.e., the duration between the sender's instruction and receiver's response contained silence as well as the filler *um*) or an unfilled inter-turn delay (i.e., the duration between the sender's instruction and the receiver's response was silent). The male speaker who played the role of the receiver was recorded producing the filler *um* several times. One instance of the filler *um* was selected based on its duration (400ms) and timbre, and was used across all filled inter-turn delays to maintain consistency. The filler was then digitally inserted within an inter-turn delay of 600 ms (100 ms silence, *um* [400 ms], 100 ms silence) and 1200 ms (700 ms silence, *um* [400 ms], 100 ms silence).

Numeric Digit Span Recall Task.

The Auditory Readback Task was designed to assess cognitive associations that operators have developed between prosodic cues (namely intonation, filler and inter-turn delay) and the extent to which they perceive that a speaker has understood an instruction. Operators who have developed refined associations between prosodic cues and instances of nonunderstanding can access these associations even when cognitive resources are limited. In contrast, operators who have not developed such associations in memory rely more heavily on

working memory processes to determine the meaning of the prosodic cues and are more hindered when cognitive resources are limited (Coderre et al., 2003; Sweller, 1988).

It was important to devise a task that would restrict participants' cognitive resources and be completed while listening to the communicative exchanges. However, it was also important to ensure that it did not impose such a load that it prevented participants from attending to the exchanges altogether. Consequently, the Numeric Digit Span Recall Task was incorporated into the Auditory Readback Task, where participants were presented with a string of eight, single numeric digits to hold in memory (and subsequently report back) while listening to each communicative exchange (see Figure 1). Each series of eight digits included numerals from one to nine in a randomized order for each exchange, with no repetition of digits.

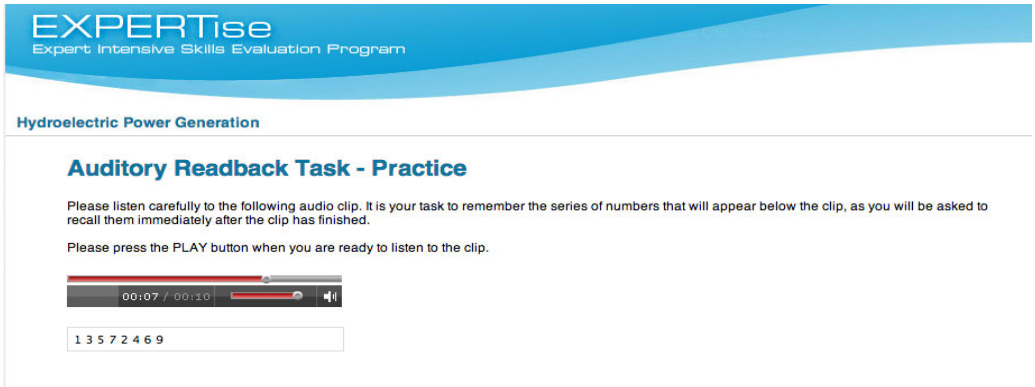


Figure 1. The visual presentation of the Auditory Readback Task and the Numeric Digit Recall Task.

To minimize the risk that participants would focus solely on retaining the numeric digit string, rather than listening to the exchanges, they were presented with four communicative exchanges that required a response to a multiple-choice question regarding the semantic content of the exchange. These four exchanges were prosodic variations of one exchange that was semantically distinct from 'Communicative Exchange 1' and 'Communicative Exchange 2'[1]. The exchanges requiring a multiple-choice response were incorporated randomly within the 32 communicative exchanges that measured perceptions of nonunderstanding such that

[1] The semantic content of the communicative exchange that was used to generate multiple-choice questions was as follows:
Instruction: *Hey Steve, can you just shut down Unit 2?*
Full Readback: *Sure Pete. I'll shut down Unit 2 now.*

participants were unaware at what intervals they would be required to respond to multiple-choice questions, and therefore, were more likely to attend to all exchanges rather than focusing solely on remembering the numeric digit strings.

As working memory is capacity-limited (Miller, 1956), a manipulation check was conducted to ensure that a string of eight numeric digits would restrict cognitive resources but not to an extent that participants could not attend to the communicative exchange (see Manipulation Checks section for details). In total, the Auditory Readback Task required participants to listen and respond to 38 communicative exchanges presented in randomized order: two practice items, 32 communicative exchanges, and four communicative exchanges requiring a response to a multiple-choice question.

Manipulation checks.

Manipulation check for intonation contours.

A manipulation check was conducted to validate that the intonation contours produced by the male speaker of the instruction responses were perceived as intended (i.e., either ending with rising or falling intonation). Participants included eleven post-graduate students from an Australian University who were all native English speakers, predominantly female (81.82%), and ranged in age from 23 to 49 years (M = 32.09, SD = 8.37).

The participants accessed the study through an online survey, where they were asked to listen to the instruction response tokens of each exchange and decide whether the receiver used rising or falling intonation at the end of each response. Although the manipulation of intonation was validated with a predominantly female cohort (81.82%), whereas the majority of power generation operators who participated in the current study were male (91.7%), Edelsky (1979) found that there was no difference in the way that males and females categorized intonation contours.

The participants listened and responded to, a total of 16 instruction response tokens; eight full readbacks (four x Communicative Exchange 1, four x Communicative Exchange 2)

and eight partial readbacks (four x Communicative Exchange 1, four x Communicative Exchange 2). Half of the instruction responses ended in rising intonation and half in falling intonation. The results indicated that participants were able to correctly identify rising and falling intonation at the end of each response token with an accuracy rate of at least 95.45% (see Table 1 for detailed results).

Table 1

Accuracy Rate (%) in which Intonation was Correctly Identified Across Instruction Response Types

	Instruction Response Type			
	Full Readback		Partial Readback	
Communicative Exchange	Rising Intonation	Falling Intonation	Rising Intonation	Falling Intonation
1	100	100	95.45	95.45
2	100	100	100	95.45

Numeric Digit Span Recall manipulation check.

The purpose of the Numeric Digit Span Recall Task manipulation check was to identify the number of single numeric digits that would sufficiently tax participants' working memory, but not to such an extent that they would be unable to attend to the information contained within the communicative exchanges. Thirteen undergraduate students from an Australian University participated in the manipulation check. All of the participants were native speakers of English, predominantly male (61.54%) and ranged in age from 18 to 55 years (M = 25.31, SD = 11.79).

The Numeric Digit Span Recall Task manipulation check comprised a 2 x 2 x 4 (Communicative Exchange [1, 2, 3] x Instruction Response Type [Readback, Partial Readback) x Numeric Digit Span Length [0, 4, 6, 8]) repeated measures design, resulting in 24 distinct conditions. 'Communicative Exchange' and 'Instruction Response Type' were identical to those described above. 'Numeric Digit Span Length' referred to the number of numeric digits that a participant was required to retain in memory while listening to each

communicative exchange. This ranged from zero (i.e., participants were not required to remember any numeric digits at all while listening to each exchange) to eight (i.e., participants were required to remember a eight numeric digits as they were listening to each exchange).

Participants completed the study online via a survey provider. Each trial involved the presentation of an audio clip containing a communicative exchange. As soon as the audio clip commenced, the numeric digit string appeared underneath the audio tool bar and remained on screen for the duration of the clip. After the completion of the audio clip, participants were asked to type the series of single numeric digits they had just viewed into a free response text box on the computer. To assess whether the task to retain the numeric string restricted cognitive resources, participants were asked to respond to a multiple choice question relating to the semantic content of the exchange. Participants were then asked to rate the difficulty of the task (i.e., subjective measure of task difficulty; Paas & Merriënboer, 1994).

The results of the manipulation check indicated that, compared to the four and six-digit numeric span, participants perceived the eight numeric digit span to be most difficult. Participants were also significantly less accurate in eight numeric digits strings, compared to four numeric digits. Nevertheless, the number of digits that they were required to recall did not affect participants' accuracy in the multiple-choice questions. In combination, these results suggest that a numeric digit span of eight digits would restrict cognitive resources considerably, but not to the extent that it would deprive participants of the resources necessary to attend to the communicative exchanges.

EXPERTise 1.0 test battery.

EXPERTise 1.0 is a situational judgement test that can be customized to assess operators' utilization of cues during task-related activities (Wiggins et al., 2010). The EXPERTise 1.0 test battery comprises a series of four tasks, performance on which requires the application of cues. Typologies of behavior that reflect greater or lesser levels of cue

utilization are calculated, the validity of which has been established in power control (Loveday, Wiggins, Harris, O'Hare, & Smith, 2013), pediatric diagnosis (Loveday, Wiggins, Searle, Festa, & Schell, 2013) and aviation decision making (Wiggins, Azar, Hawken, Loveday, & Newman, 2014). The test-retest reliability of EXPERTise 1.0 has been demonstrated as satisfactory (Loveday, Wiggins, Festa, Schell, & Twigg, 2013). The four tasks that comprise EXPERTise 1.0 included the *Feature Identification Task*, the *Paired Association Task*, the *Feature Discrimination Task* and the *Transition Task*.

Stimuli.

Feature Identification Task.

The Feature Identification Task is based on the reasonably consistent observation that experienced operators are faster and more accurate in identifying and extracting diagnostic information from an environment than their less experienced counterparts (Schriver et al., 2008; Stokes et al., 1992). Feature identification can be assessed using a target-matching task, whereby an operator is required to identify a specific domain-relevant target within a display containing other domain-relevant targets.

In the current study, participants were presented with a series of hydraulic overview screens of a New Zealand hydroelectric power generation network, each containing an alarm (indicated by either a yellow, blue or red figure; see Figure 2 for an example). The hydraulic overview was visible to participants for 1500 ms, after which they were asked to indicate, from four possible options, the geographic location of the alarm. The Feature Identification Task comprised 27 trials, where accuracy was recorded.

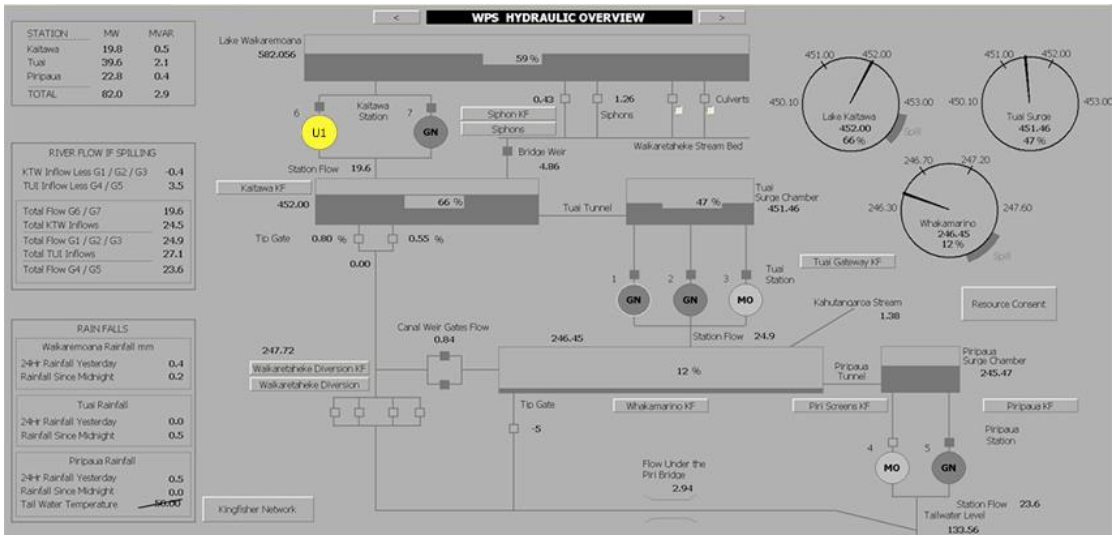


Figure 2. Example of a hydraulic overview screen capture. The alarm indicating generator outage at Kaitawa Station (represented by the yellow filled in circle) was the location target participants were required to identify.

Paired Association Task.

The Paired Association Task involved the presentation of pairs of text-based domain-relevant features and events, where participants were asked to rate the strength of the perceived relatedness between the feature and event pair (Morrison, Wiggins, Bond, & Tyler, 2012). This task is designed to assess the cue-based relationships that participants possess in memory. As experienced operators possess a greater repertoire of domain-relevant features and their relationship with events (Schriver et al., 2008), they tend to display greater speed and discrimination for ratings of association between pairs (Morrison et al., 2012).

The Paired Association Task was completed in two stages. During the first stage, participants viewed 26 text-based feature-event pairs relevant to the hydroelectric power generation domain. Each feature-event pair was presented sequentially such that the feature (e.g., *Operator Alarm*) would appear on-screen shortly before the event (e.g., *Increase Flow*). Each stimulus was available on the screen for one second. After the presentation of each pair, participants were asked to rate the strength of the perceived association between the feature-event pair on a 6-point Likert scale, from 1 (*extremely unrelated*) to 6 (*extremely related*). During the second stage of the task, participants were asked to rate the strength of perceived

association between 26 feature-event pairs that were presented *simultaneously* on screen for a duration of two seconds.

Feature Discrimination Task.

The Feature Discrimination Task involved two complex and ambiguous scenarios to which a solution was required. Each scenario contained several environmental features and participants were asked to assess the situation presented and select an appropriate course of action from several possible options. Having completed this phase, participants were asked to rate the relative usefulness of the available features associated with the situation when deciding the course of action. Individual ratings were made on a 10-point Likert scale, from 1 (*not important at all*) to 10 (*extremely important*). This task was based on the conceptualization of the Cochran-Weiss-Shanteau (CWS) index of expertise (Shanteau, Weiss, Thomas, & Pounds, 2002). This index suggests that experts are better able to discriminate consistently between features, compared to their novice counterparts.

Transition Task.

The Transition Task is based on evidence to suggest that experienced and less experienced operators differ in their acquisition of task-related information. Less experienced operators tend to acquire information in a sequence consistent with its visual presentation. However, more experienced operators tend to extract information on the basis of relevance, rather than physical accessibility (Wiggins & O'Hare, 1995).

During the Transition Task in the current study, participants were presented with an intentionally incomplete scenario, whereby additional information could be accessed through a list of thirteen 'drop down' information categories. Participants could view the information contained within each category by selecting a tab (see Figure 3). The sequence in which participants accessed additional information was recorded.

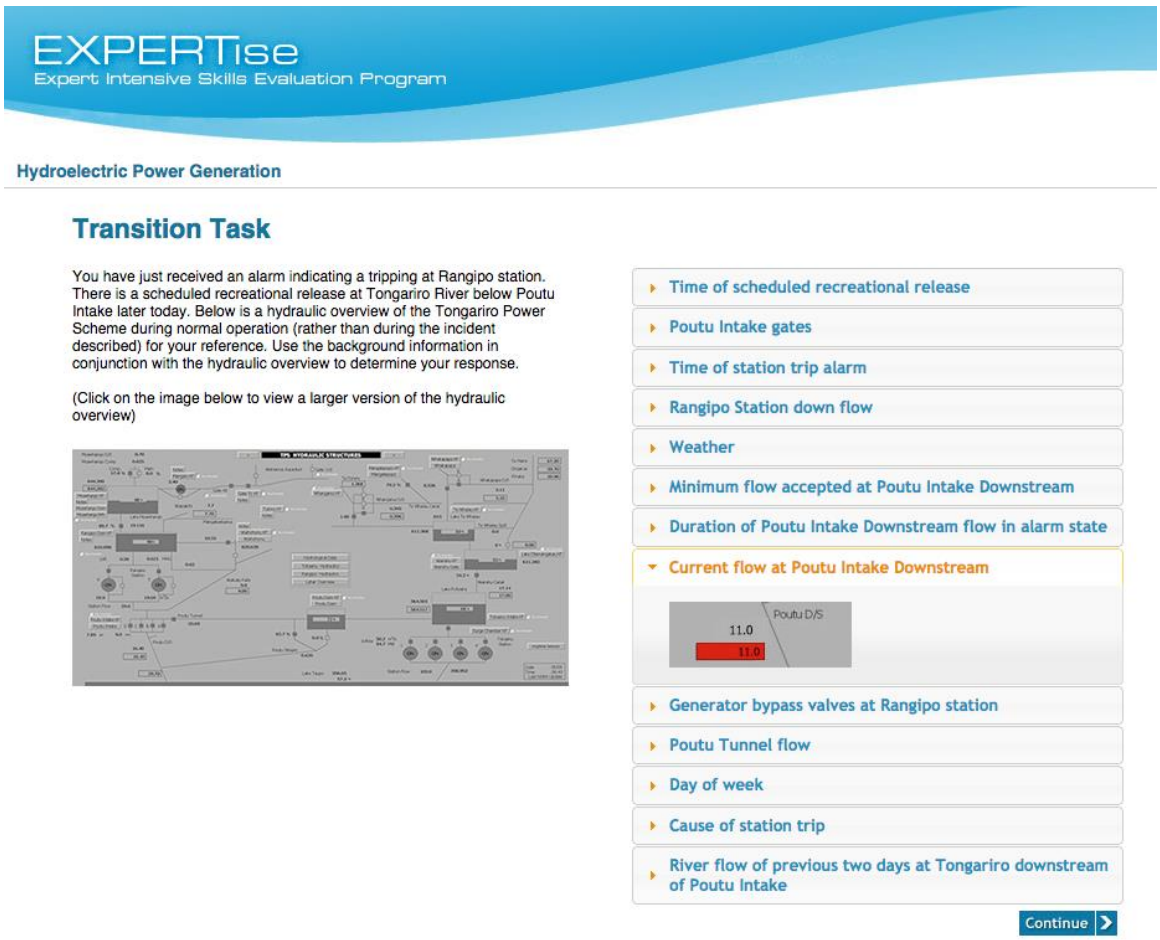


Figure 3. Screen capture of the information presented during the Transition Task. Categories of features are displayed in a list on the right hand side, which can be expanded by a mouse click. The category 'Current flow at Poutu Intake Downstream' has been expanded.

Stimuli development.

Consistent with the standard approach to developing stimuli for EXPERTise 1.0 (Loveday, Wiggins, Harris, et al., 2013), the four tasks for the current study were created from cognitive interviews conducted with two SMEs from a major New Zealand hydroelectric power generation company. The SMEs were selected on the basis of peer reference and industry position. Based on the critical incident technique (Flanagan, 1954; Klein, Calderwood, & MacGregor, 1989), each SME was asked to recall and describe in detail, an operational incident that was somewhat unusual and difficult, and in which the SMEs believed that their expertise made a critical difference to the outcome. A semi-structured interview was then used to probe details about various aspects of the decision making process (Klein et al., 1989). Three critical incidents were described during the cognitive interviews and were arranged into chronological timelines and used to develop the scenarios for the two

Feature Discrimination Tasks and the Transition Task.

An index of feature-event pairs that an operator may or may not find relevant during diagnostic decision making were also developed from the cognitive interviews and formed the items in the Paired Association Tasks. The images used in the Feature Identification Task were screen captures of hydraulic overviews, where different alarm locations were manipulated using Adobe Photoshop CS6.

Data Reduction

For the Auditory Readback Task, participants made a single rating of their perception of the receiver's level of understanding of the instruction for each communicative exchange along a 7-point Likert scale, from 1 (*did not understand at all*) to 7 (*understood completely*). Data were then aggregated for each level of prosodic and semantic cue to create a mean rating of nonunderstanding for that variable.

Since the aim of the Numeric Digit Span Recall Task was to restrict participants' cognitive resources while listening to the communicative exchanges, data were only included where participants made an attempt to recall at least one of the eight digits. Of the 24 participants who were asked to respond to a total of 32 response items each (resulting in 768 data points), there were four instances where no attempt was made to recall any of the digits. These four data points were removed and treated as missing data.

The data reduction for the EXPERTise 1.0 tasks was consistent with the standard approach to the analysis of these data (e.g., Loveday, Wiggins, Harris, et al., 2013; Wiggins et al., 2014). For the Feature Identification Task, participants' accuracy in identifying the location of the alarm was recorded. For the Paired Association Task, the speed and variance of participants' ratings of perceived association between feature pairs were calculated across both stages of the task. Each participant received a variance and latency score for this task. For the Feature Discrimination Task, participants rated the utility of each of the features on an 10-point Likert scale, from 1 (*not important at all*) to 10 (*extremely important*). These ratings

were then aggregated into a single metric based on the mean variance of the individual's responses. As this task comprised two stages, each participant received two variance scores. For the Transition Task, the ratio of pairs of features that were selected in sequence, against the total number of pairs of features available, was recorded.

Procedure

The participants were invited to take part in the study through either an organizational email distributed to operators within the New Zealand power generation company or through an online advertisement published in a major US power generation magazine. Through both sources, participants were directed to the domain address where they were able to access the study. Participants completed a number of brief demographic items that ascertained their years of experience within the domain, as well as the extent to which they give/receive instructions over the telephone/radio in their position. They first completed the Auditory Readback Task (approximately 20 minutes), and then completed the EXPERTise 1.0 tasks (approximately 40 minutes).

Results

It was important to establish initially whether participants' native language influenced the dependent variables. An independent samples *t*-test was run with native language as the grouping variable (Native Language: English, Other) and the aggregated means of the ratings of understanding for each semantic and prosodic cue used as the dependent variables. There were no significant differences between how native English speakers and non-native English speakers rated prosodic and semantic cues for understanding ($p > .05$) and, therefore, Native Language was not included as a covariate in subsequent analyses.

Influence of Semantic and Prosodic Cues on Ratings of Understanding

A series of one-way repeated measures analyses of variances (ANOVAs) were conducted to investigate the outcome of the first four hypotheses, with ratings of understanding as the dependent variable. To control for the family wise error rate, critical

alpha was set at $\alpha = .01$. To examine H1, a one-way repeated measures ANOVA was run for Instruction Response Type (Full Readback, Partial Readback). There was a significant effect of Instruction Response Type on ratings of understanding, $F(1, 23) = 78.24$, $p < .001$, $\eta_p^2 = .77$. This suggests that receivers who responded to an instruction with a full readback, were rated as having a greater level of understanding of the instruction ($M = 5.87$, $SD = .81$), compared to receivers who responded to an instruction with a partial readback ($M = 4.03$, $SE = 1.17$), providing support for H1.

To examine H2a, a one-way repeated measures ANOVA was conducted for the effects of intonation on full readback responses. There was no significant difference for ratings of understanding on the type of intonation used at the end of full readback responses, $F(1, 23) = .328$, $p = .572$, $\eta_p^2 = .01$, providing no support for H2a (see column cluster 1, Figure 4).

To examine H2b, a one-way repeated measures ANOVA was conducted for the effects of intonation within partial readback responses on ratings of understanding. There was a significant effect of intonation at the end of partial readback responses for ratings of understanding, $F(1, 23) = 14.21$, $p = .001$, $\eta_p^2 = .382$. Consistent with H2b, it was evident that partial readbacks that ended with rising intonation attracted a lower mean rating of understanding ($M = 3.84$, $SD = 1.26$) compared to partial readbacks that ended with falling intonation ($M = 4.22$, $SD = 1.13$; see column cluster 2, Figure 4).

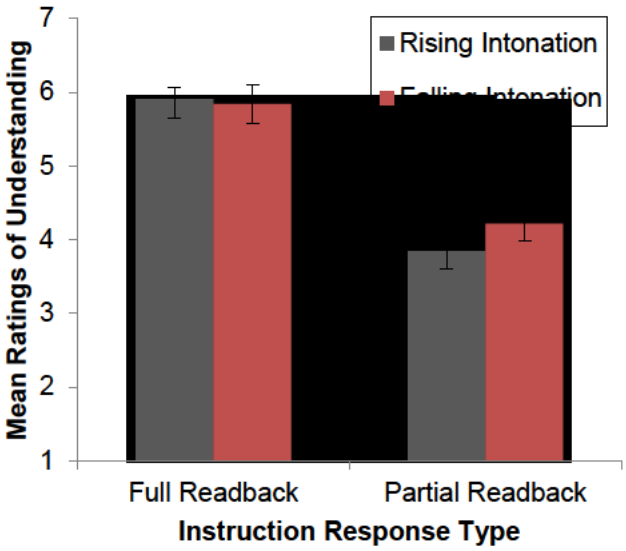


Figure 4. Non-significant effect of intonation within full readback responses for mean ratings of understanding (column cluster 1); significant effect of intonation within partial readback responses for mean ratings of understanding (column cluster 2). Error bars represent standard error.

To investigate H3a and H3b, two separate one-way repeated measures ANOVA were conducted for a filler preceding full and partial readback responses. Consistent with H3a, a significant effect was evident for the presence of a filler before a full readback response on ratings of understanding, $F(1, 23) = 8.05$, $p = .009$, $\eta_p^2 = .26$. When a full readback was preceded by a filler, participants rated the receiver as having a lower level of understanding of the instruction ($M = 5.77$, $SD = 0.77$), compared to a full readback that was not preceded by a filler ($M = 5.98$, $SD = 0.88$; see column cluster 1, Figure 5).

There was also a significant effect for the presence of a filler before a partial readback response on ratings of understanding, $F(1, 23) = 11.33$, $p = .003$, $\eta_p^2 = .330$. Providing support for H3b, participants rated a partial readback response preceded by a filler with a lower rating of understanding ($M = 3.81$, $SD = 1.31$), compared to a partial readback response that was not preceded by a filler ($M = 4.26$, $SD = 1.11$; see column cluster 2, Figure 5).

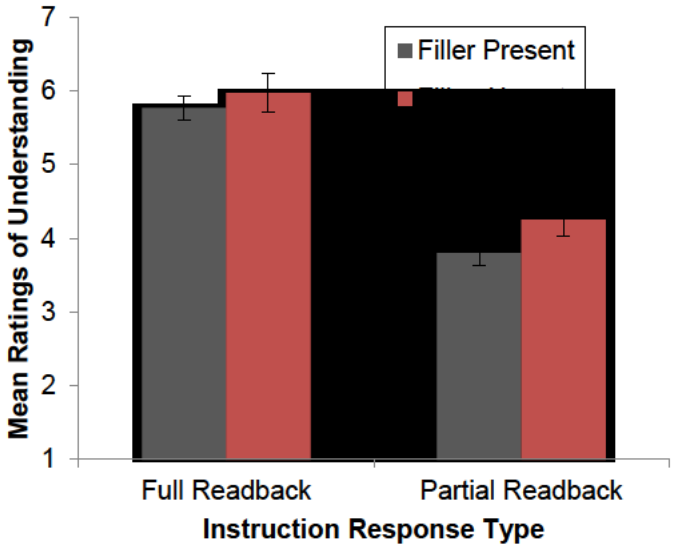


Figure 5. Significant effect of filler within both full and partial readback responses for mean ratings of understanding. Error bars represent standard error.

To investigate H4a and H4b, two separate one-way repeated measures ANOVA were conducted to assess the influence of inter-turn delay for both full and partial readback responses. No support was evident for H4a; no significant effect was evident for the duration of the inter-turn delay between the instruction and full readback response on ratings of understanding, $F(1, 23) = .568$, $p = .459$, $\eta_p^2 = .02$. That is, irrespective of whether a full readback was preceded by a shorter ($M = 5.84$, $SD = .82$) or a longer ($M = 5.90$, $SD = .84$) inter-turn delay, ratings of understanding did not vary.

There was also no support for H4b since no statistically significant effect was evident for the duration of the inter-turn delay between the instruction and partial readback response on ratings of understanding, $F(1, 23) = 2.41$, $p = .135$, $\eta_p^2 = .10$. Therefore, partial readbacks that followed a shorter inter-turn delay ($M = 4.10$, $SD = 1.23$) were rated similarly in terms of the receiver's understanding, compared with partial readbacks that followed a longer inter-turn delay ($M = 3.96$, $SD = 1.16$).

Typologies and Cue Use During Ratings of Understanding

Cluster analysis.

The overall aim of H5, H6 and H7 was to examine how typologies that were differentiated on the basis of performance across the visual cue-based diagnostic tasks,

recorded differences in the use of prosodic and semantic cues during perceptions of nonunderstanding. A K-means cluster analysis was conducted to determine whether participants could be categorized into two typologies representing distinct levels of visual cue utilization based on their composite performance across the tasks (Wiggins et al., 2014). Before the cluster analysis could be performed, *z*-scores were computed for the visual cue utilization measures (namely, the accuracy scores for the Feature Identification Task, the variance and response latency in association ratings averaged across both stages of the Paired Association Task, the variance in ratings of the utility of features for both stages of the Feature Discrimination Task and the ratio of sequential pairs of features accessed compared to features available in the Transition Task). Based on the range of years of experience of participants within power generation and the relatively small sample size, K = 2 clusters was predicted as the most likely fit for the data.

The cluster analysis yielded two distinct typologies that broadly represented relatively greater and lesser levels of visual cue utilization. Cluster 1 contained seven participants who recorded relatively greater accuracy in response to the Feature Identification Task, greater variance and lower response latency in the Paired Association Tasks and greater variance in the ratings of the utility of features in both the Feature Discrimination Tasks. Although this cluster did exhibit a relatively high ratio in response to the Transition Task that is indicative of lesser levels of cue utilization, overall, this pattern of performance across the EXPERTise 1.0 tasks is reflective of a relatively greater level of cue utilization.

The remaining 17 participants comprised the second typology with a relatively lower accuracy rate in response to the Feature Identification Task, lower variance and higher response latency in the Paired Association Tasks, and lower variance in the ratings of the utility of features in the Feature Discrimination Tasks. Although this cluster did record a relatively lower ratio in response to the Transition Task (indicative of greater levels of cue utilization), overall, performance was consistent with a lesser level of cue utilization. Table 2

summarizes the results of the cluster analysis.

Table 2

Participant Cluster Means for the EXPERTise 1.0 Measures Across the Two Cue Utilization Typologies

	Typology	
	Greater (*N*=7)	Lesser (*N*=17)
Feature Identification Task - accuracy	.67	-.07
Paired Association Task (Stages 1 + 2) – ratings variance	1.10	-.61
Paired Association Task (Stages 1 + 2) – response latency	-.46	-.10
Feature Discrimination Task (Stage 1) – ratings variance	1.24	-.48
Feature Discrimination Task (Stage 2) – ratings variance	1.16	-.51
Transition Task – ratio of sequential pairs	.57	-.26

[a] The results from the Transition Task did not correspond with the greater and lesser cue utilization typologies.

Typologies and semantic and prosodic cue use.

To examine how the two performance typologies used semantic cues during perceptions of nonunderstanding (H5), a 2 (Typology; Greater levels of cue utilization, Lesser levels of cue utilization) x 2 (Instruction Response Type; Full Readback, Partial Readback) mixed design ANOVA was conducted, with Typology as the between subjects variable and mean ratings of understanding as the dependent variable. The hypotheses were tested using planned contrasts. There was a significant interaction between Typology and Instruction Response Type, $F(1, 22) = 5.63$, $p = .027$, $\eta_p^2 = .20$, suggesting that the ratings of understanding given to full and partial readbacks were different for the two performance typologies. Specifically, in comparison to participants with lesser levels of cue utilization, participants with greater levels of cue utilization rated a full readback response as reflecting a relatively higher level of understanding ($M = 6.12$, $SE = .31$), compared to a partial readback response ($M = 3.58$, $SE = .44$), providing support for H5 (see Figure 6).

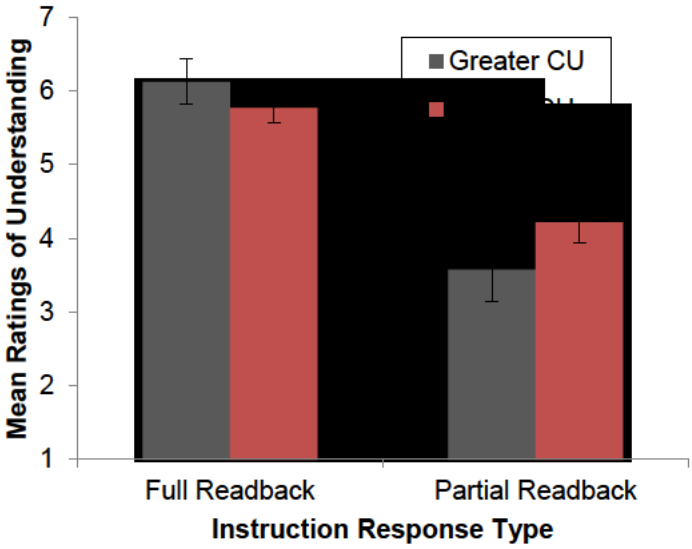


Figure 6. Significant two-way interaction between typology and instruction response type on mean ratings of understanding. Greater CU = Greater levels of cue utilization; Lesser CU = Lesser levels of cue utilization. Error bars represent standard error.

To investigate this interaction further, a simple main effects analysis using a Bonferroni adjustment was conducted. There was no significant difference in the way that participants with greater levels of cue utilization rated a full readback response on levels of understanding, compared to participants with lesser levels of cue utilisation, $p = .339$ (see column cluster 1 in Figure 6). Similarly, there was no significant difference between the ratings of understanding attributed to partial readbacks from participants with higher levels of cue utilisation and lower levels of cue utilisation, $p = .234$ (see column cluster 2 in Figure 6).

To examine whether the two performance typologies differed in their use of intonation during perceptions of understanding (H6), a 2 (Typology; Greater levels of cue utilization, Lesser levels of cue utilization) x 2 (Intonation: Rising, Falling) mixed design ANOVA was conducted, with Typology as the between subjects variable and mean ratings of understanding as the dependent variable. Using planned contrasts, the interaction between Typology and Intonation was non-significant, $F(1, 22) = 0.66$, $p = .426$, $\eta_p^2 = .03$, providing no support for H6.

To investigate whether there was any support for H7, a 2 (Typology; Greater levels of cue utilization, Lesser levels of cue utilization) x 2 (Filler: Present, Absent) mixed design

ANOVA was conducted, with Typology as the between subjects variable and mean ratings of understanding as the dependent variable. Using planned contrasts, a statistically significant interaction was evident between Typology and Filler, $F(1, 22) = 4.36$, $p = .049$, $\eta_p^2 = .17$. Compared to those participants with lesser levels of cue utilization, participants with greater levels of cue utilization rated the presence of a filler before an instruction response as indicative of a relatively lower level of understanding of the instruction ($M = 4.55$, $SE = .35$), compared to the absence of a filler ($M = 5.11$, $SE = .35$; see Figure 7).

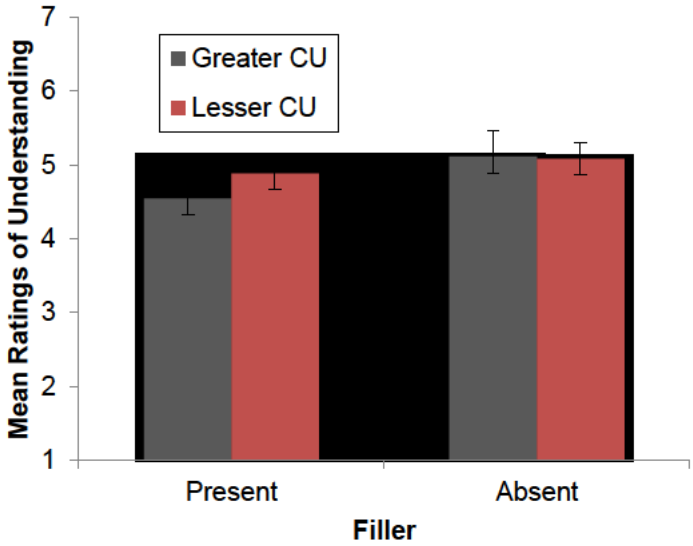


Figure 7. Significant two-way interaction between typology and filler on mean ratings of understanding. Greater CU = Greater levels of cue utilization; Lesser CU = Lesser levels of cue utilization. Error bars represent standard error.

This interaction was further investigated using simple main effects analysis, with a Bonferroni adjustment. It was found that participants' relative levels of cue utilisation did not significantly affect the ratings given to instruction responses preceded by a filler ($p = .421$; see column cluster 1 in Figure 7). The same results were found for instruction responses that were not preceded by a filler; ratings were not significantly different for those participants with greater levels of cue utilisation compared to those with lesser levels of cue utilisation ($p = .943$; see column cluster 2 in Figure 7).

Finally, H8 was tested using a 2 (Typology; Greater levels of cue utilization, Lesser levels of cue utilization) x 2 (Inter-turn delay: Shorter, Longer) mixed design ANOVA with

Typology as the between subjects variable and mean ratings of understanding as the dependent variable. Using planned contrasts, a non-significant interaction between Typology and Inter-turn delay, $F(1, 22) = 0.10$, $p = .756$, $\eta_p^2 = .00$, suggested that participants within both typologies responded in a similar fashion to instruction responses that followed either a short or long inter-turn delay, thereby failing to provide support for H8.

Discussion

There were two general aims of the current study, the first of which was to investigate how particular semantic (namely a full or partial readback) and prosodic (namely intonation, filler and inter-turn delay) cues were used by hydroelectric power generation operators when judging the extent to which a receiver had understood an instruction during a series of simulated readback/hearback exchanges. The second aim was to examine whether relatively greater or lesser levels of visual cue utilization were associated with differences in the use of prosodic and semantic cues during perceptions of nonunderstanding.

Influence of Semantic and Prosodic Cues on Perceptions of Nonunderstanding

Consistent with H1, the amount of semantic information contained within an instruction response influenced how operators judged a receiver's state of understanding. Specifically, when a response contained a full readback (containing all semantic information contained within the instruction), operators perceived the receiver as possessing a greater understanding of the instruction compared to response containing a partial readback (containing only an acknowledgement that the instruction would be executed).

In the case of full readback responses, the results indicated that only filler was useful as a prosodic cue for power generation operators when assessing a receiver's level of understanding. Consistent with H3a, when operators heard a full readback that was preceded by a filler, they rated the receiver as having a lower level of understanding of the instruction compared to a full readback that was not preceded by a filler. H2a and H4a were not supported since the prosodic cues of intonation and inter-turn delay did not influence

operators' perceptions of nonunderstanding during full readback responses.

During partial readback responses, both intonation and filler were useful prosodic cues for power generation operators when gauging the extent to which a receiver had understood an instruction. Specifically, when an operator heard a partial readback response that either ended with rising intonation or was preceded by a filler, a lower level of understanding of the instruction was ascribed to the receiver, compared to a partial readback that ended in falling intonation or was not preceded by a filler (consistent with H2b and H3b). Again, the duration of the inter-turn delay between the instruction and partial readback response did not influence how operators rated receivers' level of understanding (inconsistent with the predictions of H4b).

The outcomes of H1 established that when receivers responded with a full readback, they were perceived to possess a greater level of understanding of the instruction compared to a partial readback This suggests that the semantic information associated with a full readback does signal to listeners that the instruction has been better understood compared to an instruction that is acknowledged with a partial readback. This outcome suggests that operators who have heard an instruction correctly, but have failed to understand it, could potentially conceal their lack of understanding by using an accurate readback.

This finding does not necessarily imply that replacing the readback requirement with acknowledgements (such as *Roger that* in the aviation industry) is the solution to curtail concealed nonunderstandings during radio communication. The readback requirement is critical insofar as it provides the sender with an opportunity to monitor the accuracy of a receiver's readback, thereby reducing errors of mishearing (Schneider et al., 2004). The outcomes of the current study, however, suggest that operators should not rely solely on a full readback as evidence of a receiver's understanding. Operators should be trained to consider both an accurate readback and the prosodic cues contained within that readback when judging a receiver's state of understanding, particularly in high-risk manoeuvres.

The results also indicated that prosodic cues influence perceptions of nonunderstanding, depending upon the type of response. Specifically, a filler influenced perceptions of nonunderstanding during *full* readback responses, and both intonation and filler influence perceptions of nonunderstanding during *partial* readback responses. These results are not entirely consistent with Auton et al. (2013), who found that, during full readback responses, *both* intonation and filler were used by operators to gauge a receiver's level of understanding. This discrepancy could be accounted for by slight methodological variations between the two studies.

The present study required operators to make judgements of understanding on both full and partial readbacks, whereas Auton et al. (2013) presented full readbacks to participants in isolation. As a result, the participants in Auton et al. were assessing responses that did not vary in the amount of semantic information available, nor did they have full readbacks constantly contrasted with partial readbacks. This prevented any investigation as to whether the use of prosodic cues differed according to the amount of semantic information available. The results from the current study suggest that participants do use intonation as an important cue during perceptions of understanding, but only when semantic information from a full readback is unavailable. In their instructions to participants, Auton et al. also primed participants to attend to the prosodic cues of the readback when making judgments of nonunderstanding, whereas no such instructions were given in the present study. Such a methodological difference could have inflated the importance that participants ascribed to intonation in Auton et al.

Although Auton et al. (2013) found that the duration of the inter-turn delay between speakers is not a useful prosodic cue for operators during perceptions of nonunderstanding during full readback responses, inter-turn delay was included in the current design to investigate whether it became useful to operators within a different domain and/or when the semantic information of a full readback is unavailable (i.e., during partial readback responses).

The results of the current study demonstrate that, irrespective of the instruction response type, inter-turn delay is not a prosodic cue used by power generation operators to gauge a receiver's level of understanding.

Typologies and Cue Use during Perceptions of Nonunderstanding

The second aim of the current study was to assess whether operators, who were categorized on the basis of visual cue utilization, would differ in the way that they used prosodic and semantic cues during perceptions of nonunderstanding. Consistent with H5, compared to those participants with lesser levels of visual cue utilization, participants with greater levels of visual cue utilization perceived a full readback response as reflecting a greater level of understanding on the part of the receiver, compared to a partial readback response. Consistent with H7, participants with greater levels of visual cue utilization perceived the presence of a filler before an instruction response as more indicative of a lower level of understanding of the instruction, compared to the absence of a filler. The results failed to support H6 and H8, since there was no difference in the relationship between the two typologies and intonation at the end of instruction responses (H6) and the duration of inter-turn delays between speakers (H8). In combination, these results suggest that, while all operators utilize fillers and semantic cues during perceptions of nonunderstanding, operators with relatively higher levels of visual cue utilization accord greater importance to these cues as indicators of the understanding of the receiver.

Limitations and Future Research

An advantage of the current study was the recruitment of hydroelectric power generation operators from distinct organisations and countries. The collection of data from a heterogeneous, rather than homogeneous sample enabled greater control for the influence of culture during cue utilisation and perceptions of nonunderstanding. Organisational and cultural differences have been found to mediate operator performance, especially within high-risk, high-consequence domains (Raghunathan, 2012; Strauch, 2010), and by recruiting

participants from several different organisations and countries, there can be a higher degree of confidence that the results obtained were not specific to one particular organisation or culture. Rather, the current study demonstrates that the importance of specific auditory cues during perceptions of nonunderstanding can be applied to a range of organisational cultures.

There were, however, two main limitations of the current study. First, perceptions of nonunderstanding were not made by a participant within the communicative exchange, but rather, by a third party participant listening to the exchange. During an actual readback/hearback exchange, the sender is also required to speak when issuing instructions. This means that senders are burdened with the additional responsibility of speaking, which may impose a greater cognitive demand compared to the demands placed on participants in the current study. In an attempt to control for this issue, participants in the current study were required to remember a series of eight numeric digits while listening to each exchange with the purpose of increasing cognitive demands. However, it is not clear whether the cognitive demands incurred during the actual experience of a sender speaking and listening to a readback are comparable to those participants experienced in the current study. This potential discrepancy in cognitive load may reduce a 'real life' sender's ability to perceive nonunderstandings on the part of the receiver, as fewer cognitive resources are available to attend to the prosodic and semantic cues contained within a readback response (Keysar & Henly, 2002). Nevertheless, the current study demonstrates that, under the cognitive load imposed by a secondary task rather than the task of speaking, instruction response type, intonation and filler were important semantic and prosodic cues attended to by hydroelectric power generation operators to gauge a receiver's level of understanding during instruction responses.

Secondly, the current study did not investigate whether semantic and prosodic cues incorporated within an instruction response were sufficiently salient to motivate operators to ensure that the instruction had been properly understood. If future research finds that the

critical semantic and prosodic cues highlighted in the current research not only influence operators' perceptions of nonunderstanding on the part of the receiver but also their behaviour subsequent to the receipt of such information (such as clarifying the instruction), these cues could be incorporated within cue-based learning approaches to assist the recognition of nonunderstandings on the part of the sender, thereby improving performance.

Learning approaches that have focused on the training of critical cues have been successful in improving cue detection in a variety of industries, including aviation (Wiggins & O'Hare, 2003) and sport (Abernethy, Wood, & Parks, 1999; Smeeton, Williams, Hodges, & Ward, 2005; Williams, Ward, Knowles, & Smeeton, 2002). The results arising from the current study suggest that a filler is important when contained within full readback responses, and that both intonation and filler are important when contained within partial readback responses. These results might provide a specific framework for a cue-based learning approach. However, as operators who exhibited greater levels of visual cue utilization showed a greater tendency to rely on a filler during perceptions of nonunderstanding, this may suggest that this cue should be the focus of future training endeavours.

At an applied level, the outcomes of the present study also suggest that it is important to ensure that operators are aware that relying on an accurate readback is not sufficient to ensure that a receiver has understood the instruction. This is especially important for operators with greater levels of cue utilization who rely more heavily on the availability of semantic information to gauge a receiver's level of understanding. Rather, there should be a consideration of both an accurate readback and prosodic cues (namely, a filler) contained within that readback when judging a receiver's state of understanding.

Conclusion

Using an international sample of hydroelectric power generation operators, the current study identified that a filler, when contained within a full readback response, and filler and intonation, when contained within partial readback responses, are important prosodic cues

used during perceptions of nonunderstanding. It was also evident that semantic cues provide operators with information pertaining to the level of nonunderstanding. Finally, operators with a relatively greater level of visual cue utilization pay particular attention to both filler and semantic cues in detecting levels of nonunderstanding. The outcomes of the current study have implications for training and system design in technical environments.

Paper 4 References

Abernethy, B., Wood, J. M., & Parks, S. (1999). Can the anticipatory skills of experts be learned by novices? *Research Quarterly for Exercise and Sport*, *70*, 313–318. doi:10.1080/02701367.1999.10608050

Air Accidents Investigation Branch. (2008). *Boeing 737-500, SP-LKA* (Report No. EW/C2007/06/02). Retrieved from http://www.aaib.gov.uk/publications/bulletins/june_2008/boeing_737_500__sp_lka.cfm

Apple Inc. (2014). GarageBand (6.0.4) [Digital Audio Workstation and Music Sequencer]. Retrieved from https://itunes.apple.com/au/app/garageband/id682658836?mt=12&ls=1

Australian Transport Safety Bureau. (2007). *Radiotelephony readback compliance and its relationship to surface movement control frequency congestion* (Investigation Report No. B2006/0053). Retrieved from http://www.atsb.gov.au/media/29962/b20060053.pdf

Auton, J. C., Wiggins, M. W., Searle, B. J., Loveday, T., & Xu Rattanasone, N. (2013). Prosodic cues used during perceptions of nonunderstandings in radio communication. *Journal of Communication*, *63*, 600–616. doi:10.1111/jcom.12033

Barshi, I., & Farris, C. (2013). *Misunderstandings in ATC communication: Lanugage, cognition, and experimental methodology*. Burlington, USA: Ashgate Publishing Group.

Billings, C. E., & Cheaney, E. S. (1981). The information transfer problem: Summary and comments. In C. E. Billings & E. S. Cheaney (Eds.), *Information transfer problems in the Aviation system* (Report No. NASA TP-1875). Washington, DC: National Aeronautics and Space Administration.

Blignaut, C. J. H. (1979). The perception of hazard: II. The contribution of signal detection to hazard perception. *Ergonomics*, *22*, 1177–1183. doi:10.1080/00140137908924692

Bolinger, D. (1985). *Intonation and its parts*. London, UK: Edward Arnold.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398. doi:10.1006/jmla.1995.1017

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111. doi:10.1016/S0010-0277(02)00017-3

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education*, *37*, 695–703. doi:10.1046/j.1365-2923.2003.01577.x

Cruttenden, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics*, *17*, 77–91. doi:10.1017/S0022226700006782

Cushing, S. (1995). Pilot-Air traffic control communications: It's not (only) what you say, it's how you say it. *Flight Safety Digest*, *14*(7), 1–10.

Edelsky, C. (1979). Question intonation and sex roles. *Language in Society*, *8*, 15–32. doi:10.1017/S0047404500005923

Federal Aviation Administration. (2014). *Aeronautical information manual*. Washington, DC: Government Printing Office. Retrieved from http://www.faa.gov/air_traffic/publications/atpubs/aim/index.htm

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 327–358. doi:10.1037/h0061470

Garzone, G., Catino, M., Gobo, G., Bait, M., Catenaccio, P., Degano, C., & Rozzi, S. (2010). Towards an integrated model for the understanding of communication failures in aviation accidents: Tenuous identities under pressure. In G. Garzone & J. Archibald (Eds.), *Discourse, identities and roles in specialized communication* (pp. 209–244). Bern, Switzerland: International Academic Publishers.

Greenberg, C. C., Regenbogen, S. E., Studdert, D. M., Lipsitz, S. R., Rogers, S. O., Zinner, M. J., & Gawande, A. A. (2007). Patterns of communication breakdowns resulting in

injury to surgical patients. *Journal of the American College of Surgeons*, *204*, 533–540. doi:10.1016/j.jamcollsurg.2007.01.010

Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. (1994). Repairing conversational misunderstandings and non-understandings. *Speech Communication*, *15*, 213–229. doi:10.1016/0167-6393(94)90073-6

International Civil Aviation Organization. (2007). *Manual of Radiotelephony* (Report No. Doc 9432 AN/925). Retrieved from http://www.aeroclubvercelli.it/download/Manual%20of%20RT%20DOC%209432%20-%204%20ed.%202007.pdf

Jones, R. K. (2003). Miscommunication between pilots and air traffic control. *Language Problems and Language Planning*, *27*, 233–248. doi:10.1075/lplp.27.3.03jon

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, *13*, 207–212. doi:10.1111/1467-9280.00439

Kim, S., Park, J., Han, S., & Kim, H. (2010). Development of extended speech act coding scheme to observe communication characteristics of human operators of nuclear power plants under abnormal conditions. *Journal of Loss Prevention in the Process Industries*, *23*, 539–548. doi:10.1016/j.jlp.2010.04.005

Klein, G. A., & Calderwood, R. (1991). Decision models : Some lessons from the field. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*, 1018–1026. doi:10.1109/21.120054

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*, 462–472. doi:10.1109/21.31053

Loveday, T., Wiggins, M. W., Festa, M., Schell, D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In P. Latorre Carmona, J. S. Sánchez, & A. L. N.

Fred (Eds.), *Pattern recognition - Applications and methods* (Vol. 204, pp. 1–11). Berlin, Heidelberg: Springer Berlin Heidelberg.

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2013). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*, 90–107. doi:10.1177/0018720812450911

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2013). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*, 125–137. doi:10.1177/0018720812448475

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, *63*, 81–97. doi:10.1037/h0043158

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2012). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making*, *7*, 211–226. doi:10.1177/1555343412459192

Morrow, D., Lee, A., & Rodvold, M. (1993). Analysis of problems in routine controller-pilot communication. *The International Journal of Aviation Psychology*, *3*, 285–302. doi: 10.1207/s15327108ijap0304_3

Morrow, D., Rodvold, M., & Lee, A. (1994). Nonroutine transactions in controller- pilot communication. *Discourse Processes*, *17*, 235–258. doi:10.1080/01638539409544868

Müller, S., Abernethy, B., & Farrow, D. (2006). How do world-class cricket batsmen anticipate a bowler's intention? *The Quarterly Journal of Experimental Psychology*, *59*, 2162–2186. doi:10.1080/02643290600576595

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1017–1030. doi:10.1037/0096-1523.34.4.1017

O'Connor, J. D., & Arnold, G. F. (1961). *Intonation of colloquial English*. London, UK: Longman.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 3–20). Norwood, NJ: Ablex Publishing Corporation.

Paas, F. G. W. C., & Merriënboer, J. J. G. Van. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122–133. doi:10.1037/0022-0663.86.1.122

Prinzo, O. V., & Britton, T. W. (1993). *ATC/pilot voice communications: A survey of the literature* (Report No. DOT/FAA/AM-93/20). Washington, DC: Office of Aviation Medicine.

Prinzo, O. V., Hendrix, A. M., & Hendrix, R. (2006). *The Outcome of ATC message complexity on pilot readback performance* (Report No. DOT/FAA/AM-06/25). Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

Raghunathan, K. (2012). Checklists, safety, my culture and me. *BMJ Quality & Safety*, *21*, 617–620. doi:10.1136/bmjqs-2011-000608

Schneider, V. I., Healy, A. F., & Barshi, I. (2004). Effects of instruction modality and readback on accuracy in following navigation commands. *Journal of Experimental Psychology: Applied*, *10*, 245–257. doi:10.1037/1076-898X.10.4.245

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise difference in attentional strategies related to pilot decision making. *Human Factors*, *50*, 864–878. doi:10.1518/001872008X374974.

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*, 253–263. doi:10.1016/S0377-2217(01)00113-8

Skantze, G. (2005). Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication*, *45*, 325–341. doi:10.1016/j.specom.2004.11.005

Smeeton, N. J., Williams, A. M., Hodges, N. J., & Ward, P. (2005). The relative effectiveness of various instructional approaches in developing anticipation skill. *Journal of Experimental Psychology: Applied*, *11*, 98–110. doi:10.1037/1076-898X.11.2.98

Stokes, A. F., Kemper, K. L., & Marsh, R. (1992). *Time-stressed flight decision making: A study of expert and novice aviators* (Report No. ARL-93-1/INEL-93-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Strauch, B. (2010). Can cultural differences lead to accidents? Team cultural differences and sociotechnical system operations. *Human Factors*, *52*, 246–263. doi:10.1177/0018720810362238

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi:10.1016/0364-0213(88)90023-7

Weigand, E. (1999). Misunderstanding: The standard case. *Journal of Pragmatics*, *31*, 763–785. doi:10.1016/S0378-2166(98)00068-X

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd Ed.). Upper Saddle River, New Jersey: Prentice Hall.

Wiggins, M. W. (2006). Cue-based processing and human performance. In W. Karwowski (Ed.), *Encyclopedia of Ergonomics and Human Factors* (2nd ed., pp. 641–645). London, UK: Taylor and Francis.

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science, 65*, 118–124. doi:10.1016/j.ssci.2014.01.006

Wiggins, M. W., Harris, J., Loveday, T., & O'Hare, D. (2010). EXPERTise (Software Package). Sydney, Australia: Macquarie University.

Wiggins, M. W., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied, 1*, 305–320. doi:10.1037/1076-898X.1.4.305

Wiggins, M. W., & O'Hare, D. (2003). Weatherwise: Evaluation of a cue-based training approach for the recognition of deteriorating weather conditions during flight. *Human Factors, 45*, 337–345. doi:10.1518/hfes.45.2.337.27246

Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied, 8*, 259–270. doi:10.1037/1076-898X.8.4.259

Paper 4 Appendix

Semantic Content of Communicative Exchanges

Communicative Exchange	Instruction	Instruction Response Type	
		Full Readback	Partial Readback
1	Hi Steve, Can you go to Rangipo Station and check the status of the bypass valves on the two generators and give me a call.	Thanks Pete. I'm on my way to Rangipo right now and will check the status of the bypass valves and give you a call.	Thanks Pete. I'll do that now.
2	Hey Steve, we've just had a tripping down at Tekapo B station. Can you do a manual reset on Unit 2?	Thanks Pete. I'll go ahead and reset Unit 2 trips.	Thanks Pete. I'll go ahead with that.

Discussion of the Research Presented

Research Outcomes

The impetus behind this program of research was the long-term goal of improving the recognition of concealed nonunderstandings that potentially compromise the reliability of the readback/hearback protocol used in many operational environments. Communication, including that which takes place during a readback/hearback exchange, is considered a joint activity, where both participants are expected to collaborate to ensure mutual understanding, or *common ground* (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986; Cohen & Levesque, 1994). When a receiver conceals a nonunderstanding in the context of a readback/hearback exchange, the sender becomes responsible for recognising and repairing the issue to ensure common ground (Schegloff, Jefferson, & Sacks, 1977). To recognise a nonunderstanding, a sender must rely solely on the semantic and prosodic cues contained within a receiver's readback response, as the readback/hearback protocol is used exclusively during radio communication (Nygaard & Queen, 2008).

The first goal of the current thesis was to investigate whether particular prosodic and semantic features contained within a receiver's readback response were utilised by operators from various technical environments to gauge the extent to which a receiver was perceived to have understood an instruction. Empirical support for this proposition has implications for the training of technical operators to detect nonunderstandings during readback/hearback exchanges in a variety of industrial settings, thereby reducing the potential for communication error.

An allied aim of this thesis was to examine whether there was any relationship between visual cue utilisation (during various domain-specific tasks requiring the application of cues) and auditory cue utilisation (during assessments of nonunderstandings). According to the Recognition Primed Decision (RPD) model (Klein, 1993, 1997), experienced operators, compared to their less experienced counterparts, are expected to possess a superior ability to

acquire and utilise cues in the environment during situation assessment (e.g., Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Wallis & Horswill, 2007; Wiggins & Bollwerk, 2006; Wiggins & O'Hare, 1995). An empirical demonstration that operators with a greater capacity for visual cue utilisation, *also* possess a greater capacity to utilise auditory cues during assessments of nonunderstandings, would offer support for the notion that cue utilisation is a generalised capability that underpins expert performance during situation assessment. To address these two goals, four research papers were presented. Each paper assessed a targeted research question/s that is summarised in Figure 1.

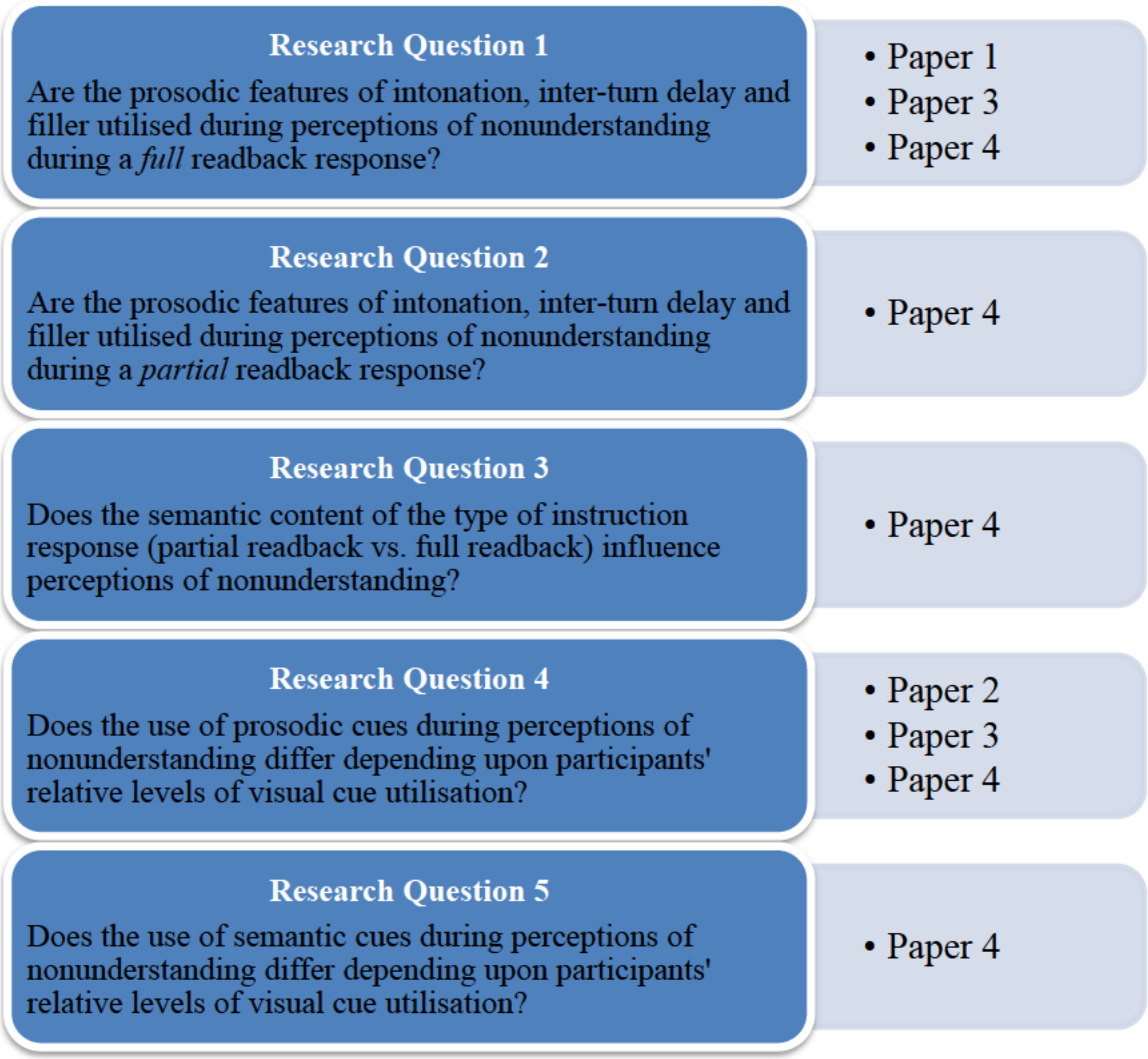


Figure 1. Diagrammatic representation of research questions and associated papers.

Research Question 1: *Are the prosodic features of intonation, inter-turn delay and filler, utilised during perceptions of nonunderstanding during a full readback response?*

The first research question examined whether the prosodic features, previously found to be useful cues during perceptions of uncertainty (Brennan & Williams, 1995), are also utilised by operators during perceptions of nonunderstanding. It was hypothesised that there would be a relationship between those features utilised during perceptions of uncertainty and nonunderstandings due to the potential relationship between the two constructs. Uncertain information is considered incomplete, ambiguous, erroneous and/or imprecise (Woods, 1988). It could be argued that the receipt of uncertain information precipitates instances of nonunderstandings. Therefore, the first research question investigated whether the prosodic features of intonation, inter-turn delay and filler contained within a full readback response, are utilised by technical operators to assess the extent to which a nonunderstanding had occurred.

This research question was examined in Papers 1, 3 and 4. In the study reported in Paper 1, electricity transmission network operators from Australia and New Zealand were asked to complete the *Auditory Readback Task*. This task comprised a series of domain-specific communicative exchanges that were designed to emulate the readback/hearback protocol. The readback portion of each exchange was manipulated to vary systematically the final intonation at the end of the readback response, the duration of the inter-turn delay between the offset of the sender's instruction and the onset of the receiver's readback, and whether or not the receiver preceded the readback with a filler. Operators were asked to listen to each exchange and subsequently rate the extent to which the receiver (i.e., the speaker of the readback) was perceived to have understood the instruction.

The outcomes of Paper 1 suggested that both intonation and filler represented important prosodic cues that were used by electricity transmission network operators during perceptions of nonunderstandings of full readback responses. Specifically, operators ('listeners') perceived that receivers ('speakers') had a lower level of understanding of the instruction when the full readback ended with rising, rather than falling, intonation. When full readbacks

were preceded by a filler, operators also perceived that the receiver had a lower level of understanding of the instruction, compared to readbacks that were preceded by an unfilled inter-turn delay of the same duration. Unlike Brennan and Williams' (1995) investigation, there was no association found in Paper 1 between perceptions of nonunderstanding and the duration of inter-turn delays. Overall, the outcomes of Paper 1 were largely consistent with Brennan and Williams' findings. Specifically, both intonation and filler were identified as important prosodic cues used by electricity transmission network operators during assessments of nonunderstanding.

Paper 3 also assessed the utility of intonation, inter-turn delay and filler during perceptions of nonunderstanding of full readback responses, but was distinct from Paper 1 in two ways. Firstly, Paper 3 was situated within the context of electrical rail control, rather than electricity transmission control. Secondly, Paper 3 incorporated an extended version of the Auditory Readback Task, which was designed to measure prosodic cue use in a more ecologically-valid manner. Despite these differences, the outcomes of Paper 3 replicated those in Paper 1. Namely, both intonation and filler were identified as prosodic cues that were used by operators during assessments of nonunderstandings of full readback responses.

Incorporating the extended version of the Auditory Readback Task in the domain of hydroelectric power generation, Paper 4 also included an examination of the first research question. Consistent with the outcomes of Papers 1 and 3, Paper 4 revealed that 'filler' was an important prosodic cue used by power generation operators during assessments of nonunderstanding. Specifically, when operators heard a full readback that was preceded by a filler, the receiver was rated as having a lower level of understanding of the instruction compared to a full readback that was not preceded by a filler. Similarly, the duration of the inter-turn delay between speakers was not salient to power generation operators during perceptions of nonunderstanding. Contrary to the results in Papers 1 and 3, however, intonation was not identified as an important prosodic cue in gauging a receiver's level of

understanding.

Overall, there was consistent support for the notion that ‘filler’ is an important prosodic cue used during perceptions of nonunderstandings across a variety of domains, namely electricity transmission control, electrical rail control, and hydroelectric power generation. While inter-turn delay has been shown to be important during perceptions of uncertainty (Brennan & Williams, 1995), the results of this program of research showed consistent support for the notion that ‘duration of inter-turn delay’ was not associated with perceptions of nonunderstanding within these domains. Finally, ‘intonation’ was not consistently identified as an important prosodic cue used to assess the extent to which a nonunderstanding had occurred during a full readback response.

Research Question 2: *Are the prosodic features of intonation, inter-turn delay and filler utilised during perceptions of nonunderstanding during a partial readback response?*

The first research question examined the utility of intonation, inter-turn delay and filler as prosodic cues of nonunderstanding during a *full* readback response. However, it is clear that some operational environments do not require a full readback during inter-operator communication (Alderson, 2009; Barshi & Farris, 2013; Cushing, 1994; Morrow, Rodvold, & Lee, 1994). Therefore, the second research question examined whether intonation, inter-turn delay and/or filler aided the recognition of nonunderstandings when embodied within a *partial, rather than a full* readback.

This research question was examined exclusively in Paper 4, in the context of hydroelectric power generation. Consistent with the results pertaining to a full readback, a partial readback response that ended with rising, rather than falling intonation was associated with a perceived lower level of understanding of the instruction. Similarly, a partial readback preceded by a filler was associated with a perceived lower level of understanding. Finally, no relationship was evident between the duration of inter-turn delay and perceived levels of understanding.

Research Question 3*: Does the semantic content of the type of instruction response (partial readback vs. full readback) influence perceptions of nonunderstanding?*

Although there are many different semantic features that could potentially influence perceptions of nonunderstanding during readbacks (such as message length, message complexity and the complexity of instructed tasks), the third research question investigated whether the receipt of a full readback, rather than a partial readback, was associated with perceptions of nonunderstanding. Consistent with expectations, Paper 4 showed that a greater level of understanding was assumed where full readbacks (i.e., readbacks that contained all semantic information evident in the instructions) were provided in comparison to partial readbacks (i.e., readbacks that contained only an acknowledgement that the instructions would be executed). In the context of this thesis, this outcome is significant insofar as an accurate readback represents, in and of itself, a relative indicator of understanding on the part of the receiver. For those environments where a response can be rehearsed, an accurate readback may potentially conceal a lack of understanding.

Research Question 4: *Does the use of prosodic cues during perceptions of nonunderstanding differ depending upon participants' relative levels of visual cue utilisation?*

There is reasonably consistent empirical evidence to suggest that the utilisation of visual cues distinguishes expert from non-expert operators during situation assessment (Cellier, Eyrolle, & Marine, 1997; Lipshitz, Klein, Orasanu, & Salas, 2001; Müller, Abernethy, & Farrow, 2006; Schriver, Morrow, Wickens, & Talleur, 2008). However, it remains unclear whether experts and non-experts differ in their utilisation of auditory cues. This association between visual and auditory cue utilisation was examined in Papers 2, 3 and 4. In the case of visual cue utilisation, participants were classified into two typologies based on performance on the EXPERT intensive skills evaluation, *EXPERTise 1.0*, situational judgement test (Wiggins, Harris, Loveday, & O'Hare, 2010). Operators who demonstrated relatively greater levels of visual cue utilisation were generally faster and more accurate in their retrieval of visual cues, were more discriminating between visual cues, and were less sequential in their retrieval of cue-based information. Operators with lesser levels of visual cue utilisation were

relatively slower and less accurate in their retrieval of visual cues, were less discriminating between visual cues, and were more sequential in their retrieval of cue-based information.

In comparison to participants with lesser levels of visual cue utilisation, it was hypothesised that participants with greater levels of visual cue utilisation would perceive a greater difference in levels of understanding on the part of the receiver between readback responses ending in rising intonation compared to readback responses ending in falling intonation; readback responses preceded by a shorter compared to a longer inter-turn delay, and; readback responses preceded by a filler compared to an unfilled inter-turn delay of the same duration.

The outcomes of Paper 2 failed to support the hypothesised relationship between visual cue utilisation and the three prosodic cues (which were used) during perceptions of nonunderstanding. Similarly, there was no relationship evident between visual and prosodic cue utilisation in Paper 3, despite the application of a more refined test of prosodic cue utilisation. The outcomes of Paper 4 were largely consistent with the results emerging from Papers 2 and 3 insofar as there was no relationship between visual cue utilisation performance and the use of either ‘intonation’ or ‘inter-turn delay’ during assessments of nonunderstanding.

Although the outcomes of the current research program generally failed to identify an association between visual and auditory cue utilisation, a relationship was evident when the prosodic cue comprised the presence or absence of a ‘filler’ in the context of hydroelectric power generation (Paper 4). Nevertheless, the combination of results suggests that, rather than being related to domain-specific visual cue utilisation, prosodic cues such as intonation and inter-turn delay are likely to represent a general skill associated with perceptions of non-understanding.

Research Question 5: *Does the use of semantic cues during perceptions of nonunderstanding differ depending upon participants' relative levels of visual cue utilisation?*

Although there was only limited support for a relationship between visual and prosodic cue utilisation during perceptions of nonunderstanding, a relationship was evident between visual cue utilisation and the interpretation of semantic cues (Paper 4). Specifically, in comparison to participants with a lesser level of visual cue utilisation, participants with a greater level of visual cue utilisation associated a full readback with a greater level of understanding on the part of the receiver, in comparison to a partial readback. This outcome reflects the complexity of prosodic and semantic cues in assessments of nonunderstanding, and suggests that the utilisation of cues, as a generalised construct, informs situation assessment but that their application is dependent upon the particular situation being confronted. A summary of the main research outcomes in relation to the proposed research questions is presented in Table 1.

Table 1

Summary of Key Research Outcomes

Research Question			Key Research Outcomes
Number	Content	Papers	
1	Are the prosodic features of intonation, inter-turn delay and filler utilised during perceptions of nonunderstanding during a *full* readback response?	Paper 1 Paper 3 Paper 4	When preceding a full readback response, 'filler' was consistently identified as an important prosodic feature during perceptions of nonunderstanding in the domains of electricity transmission control (Paper 1), electrical rail control (Paper 3), and hydroelectric power generation (Paper 4). Final 'intonation' at the end of a full readback response was identified as an important prosodic feature used to assess the extent to which a nonunderstanding had occurred in the domains of electricity transmission control (Paper 2) and electrical rail control (Paper 3). 'Inter-turn delay' between an instruction and full readback response was not identified as an important prosodic feature used during perceptions of nonunderstanding in any of the technical domains examined (Papers 1, 3 and 4).
2	Are the prosodic features of intonation, inter-turn delay and filler utilised during perceptions of nonunderstanding during a *partial* readback response?	Paper 4	Incorporated within a partial readback response, 'intonation' and 'filler' were identified as important prosodic features used during perceptions of nonunderstanding within the domain of hydroelectric power generation (Paper 4). No relationship was evident for the 'inter-turn delay' preceding a partial readback response and perceived levels of nonunderstanding within hydroelectric power generation (Paper 4).
3	Does the semantic content of the type of instruction response (partial vs. full readback) influence perceptions of nonunderstanding?	Paper 4	The semantic content of an instruction response influenced perceptions of nonunderstanding within the domain of hydroelectric power generation (Paper 4).
4	Does the use of prosodic cues during perceptions of nonunderstanding differ depending upon participants' relative levels of visual cue utilisation?	Paper 2 Paper 3 Paper 4	No relationship was found between levels of visual cue utilisation and the utilisation of prosodic cues during perceptions of nonunderstanding within electricity transmission control (Paper 2) and electrical rail control (Paper 3). A relationship was found between visual cue utilisation and the use of a 'filler' (but not 'intonation' and 'inter-turn delay') during perceptions of nonunderstanding within hydroelectric power generation (Paper 4).
5	Does the use of semantic cues during perceptions of nonunderstanding differ depending upon participants' relative levels of visual cue utilisation?	Paper 4	A relationship was evident between visual cue utilisation and the interpretation of semantic cues within the domain of hydroelectric power generation (Paper 4).

Key Findings

The four empirical papers that comprise this thesis were designed to examine the utility of specific prosodic and semantic features during assessments of nonunderstanding of readback responses. An allied aim was to investigate the relationship between visual cue utilisation during task-related activities and auditory cue utilisation during assessments of nonunderstanding. The key findings of the research presented were:

(a) 'Filler' was identified as an important prosodic cue utilised by operators during assessments of nonunderstandings of *full* readback responses in a range of applied environments, including electricity transmission control, electrical rail control and hydroelectric power generation. When preceded by a partial readback response, filler was also utilised during perceptions of nonunderstandings within the domain of hydroelectric power generation.

(b) There was reasonable support for the use of 'intonation' as an important prosodic cue utilised by operators during assessments of nonunderstandings of *full* readback responses within electricity transmission control and electrical rail control. Intonation was also utilised by hydroelectric power generation controllers during perceptions of nonunderstandings of *partial* readback responses.

(c) The semantic information associated with a full readback, compared to a partial readback, was salient for hydroelectric power generation controllers during perceptions of nonunderstandings.

(d) Visual cue utilisation performance was associated with the use of a 'filler' during perceptions of nonunderstandings, but only in hydroelectric power generation.

(e) Visual cue utilisation was associated with the use of semantic content during perceptions of nonunderstanding in hydroelectric power generation.

Prosodic Cues and the Recognition of Nonunderstandings During Readback Responses

Theoretical implications.

Models of dialogue structure, such as the collaborative model (Clark, 1994; Clark & Brennan, 1991; Clark & Schaefer, 1987, 1989; Clark & Wilkes-Gibbs, 1986), place joint responsibility on communicative partners to ensure that the information exchanged has been heard and understood before the conversation proceeds. A communicative exchange involves both the presentation of content (where a sender will present information for the receiver to understand) and the grounding of that content (where both the sender and receiver work together to establish that the receiver has understood the sender's intended message; Clark & Brennan, 1991).

It is during this second stage of a communicative exchange where a receiver is expected to display evidence that the sender's message has been understood, and such evidence must be accepted subsequently as valid by the sender before communication should proceed (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1990). When this has taken place, the utterance from the sender is referred to as *common ground* between the two communicative participants (Clark & Schaefer, 1987). Within the context of a readback/hearback exchange, the sender's instruction to the receiver represents the first stage of the communicative exchange (i.e., presentation of content). The sender and receiver must then collaborate to ensure that the instruction is common ground between them.

Within high-risk, high consequence environments, it is expected that, when a receiver has knowingly failed to understand a sender's instruction, the receiver will attempt to repair the communication error (Barshi & Farris, 2013; International Civil Aviation Organization, 2007; Monan, 1986). This type of repair might involve the receiver seeking repetition, clarification or elaboration of the information that was ambiguous or absent from a sender, such that common ground can be achieved (Bazzanella & Damiano, 1999). Within this readback context however, a receiver can deceptively conceal the fact that an instruction has

not been understood (i.e., concealed nonunderstanding) by failing to request clarification, and continuing the exchange with a full readback.

To detect such instances of concealed nonunderstandings, it becomes the responsibility of the sender to skilfully detect the miscommunication and initiate a repair (Schegloff, 2000). As the readback/hearback protocol is used exclusively during radio/telephone communicative exchanges, a sender is required to attend to the prosodic and semantic cues contained within a receiver's readback to gauge the extent to which an instruction has been understood.

The outcomes of the current thesis suggest that, during instances when receivers provide no explicit evidence of a nonunderstanding (such as seeking clarification of the instruction from the sender), senders are able to attend to specific prosodic and semantic cues of the readback response to aid the recognition of nonunderstandings. Specifically, there was general support that intonation and filler are prosodic cues used during assessments of nonunderstanding during *both* full and partial readback responses.

It is important to address the discrepancy in the findings regarding the importance of intonation during perceptions of nonunderstanding of full readback responses. Within the two domains of electrical transmission control (Paper 1) and electrical rail control (Paper3), it was evident that intonation was used as an indication of a receiver's level of understanding. However, within in the domain of hydroelectric power generation (Paper 4), intonation was not rated as an important cue during perceptions of nonunderstandings of full readback responses.

This discrepancy could be attributed to one of three explanations. Firstly, it could be the case that the prosodic cues that are used to perceive nonunderstandings during readback exchanges are not universally applicable within all high-risk, high-consequence environments. While intonation is useful within domains such as electricity transmission control and electrical rail control, operators within hydroelectric power generation may not rely on this feature to assess instances of nonunderstanding.

A second explanation to account for this discrepancy could be the possible influence of the different variations of English that were spoken by the participants in comparison to the dialect that was used to generate the voice stimuli (Australian English). The research described in Papers 1 and 3 included participants who were largely from Australasia (Australia and New Zealand). In contrast, those participants who were involved in the research described in Paper 4 were largely from the United States (US; 37.5%). Differences between the dialects of Australasian English and American English may have influenced how the prosodic features were perceived during assessments of nonunderstanding.

One distinction between Australasian and American English is the purported difference in the function of rising intonation at the end of declarative utterances. Amongst many English speakers, including those from the US, rising intonation at the end of a declarative utterance is used to transform a statement into a question (Guy, Horvath, Vonwiller, Daisley, & Rogers, 2011; Guy & Vonwiller, 1984). For example, when the declarative statement, *Peter is coming to the movies tonight,* ends with a rising intonation, it adopts the meaning of a question.

While rising intonation is used to convert statements into questions amongst speakers of Australian English and New Zealand English, there is a widely-held belief that rising intonation also serves an additional function amongst these groups (Fletcher & Harrington, 2001). In particular, the *Australian Questioning Intonation* (AQI; Bryant, 1980, as cited in Guy & Vonwiller, 1984) appears to be used by Australasian speakers to elicit verification that the listener has comprehended the utterance during the grounding process (Guy et al., 2011)[1]. Therefore, while the use of rising intonation at the end of declarative statements in many US variants of English serves to transform statements into questions, thereby demonstrating hesitation and/or the uncertainty of the speaker (Ladd, 1980; Lakoff, 1973), amongst Australasian speakers, it has the additional function of obtaining verification that the listener

[1]While the AQI is not exclusively used by Australasian speakers, it is believed to be most commonly used by these speakers compared to other varieties of English speakers (Guy et al., 2011).

has understood what the speaker has said.

However, this difference in the use of rising intonation amongst Australasian and American-English speakers would not necessarily account for the findings that the former, rather than the latter, rated intonation as important during perceptions of nonunderstandings of full readback responses. Arguably, Australasian listeners are more attuned to hearing a rising intonation at the end of declarative clauses, where they do not automatically ascribe uncertainty on behalf of speakers. Therefore, American English speakers would be more likely to attribute rising intonation at the end of a declarative clause (e.g., a readback response), as signalling uncertainty or nonunderstanding on behalf of the speaker (i.e., the receiver in a readback context). However, the outcomes of the current program of research suggests the reverse, where it was the Australasian speakers who associated rising intonation with greater levels of nonunderstanding in comparison to falling intonation. This effect was not evident for a group of participants who were largely American-English speaking.

If there was a potential influence of the different English dialects, it is more likely to be the case that the American-English participants may have experienced difficulties in understanding the communicative exchanges, which were spoken by two Australian-English males. These participants may have expended more effort in attempting to understand the semantic content of the exchanges, and therefore, the prosodic features of the readback may have become less salient. Alternatively, as the Australian accent was presumably foreign to the American-English participants, they may have been less confident in drawing conclusions on the basis of intonation.

The most likely explanation of the discrepancy in findings regarding of the importance of intonation during perceptions of nonunderstanding was the difference in methodology used in Papers 1 and 3, compared to the methodology employed in Paper 4. The Auditory Readback Task that was utilised in the research described in Papers 1 and 3, required participants to rate understandings for full readback responses in isolation. However,

participants in the research described in Paper 4 were asked to assess nonunderstandings of *both* full and partial readback responses. Consequently, it may be the case that intonation is only used to gauge a receiver's level of understanding when the full readback is unavailable (i.e., during partial readback responses). This suggests a hierarchy in which semantic cues are used initially, and prosodic cues are used subsequently during situation assessment.

Although the importance of intonation and filler during perceptions of nonunderstanding was evident throughout the current thesis, operators never associated the duration of inter-turn delay with perceptions of nonunderstanding during both full and partial readback responses. This was contrary to the findings of Brennan and Williams' (1995) investigation into prosodic cues and uncertainty, where a longer inter-turn delay was related to greater perceptions of uncertainty on the part of the speaker. The null findings within the current thesis regarding the influence of inter-turn delay on perceptions of nonunderstanding may be attributed to one of two explanations.

It may have been the case that the durations of inter-turn delay between speakers investigated in the current thesis (600 ms and 1200 ms[2]) were not sufficient to signal an issue in communication. In their investigation of perceptions of uncertainty, Brennan and Williams (1995) compared the effects of one second and five-second inter-turn delays. In contrast, the current thesis opted to compare the effect of durations of 600 ms and 1200 ms. The decision to adopt these durations was due to: (1) consultations with a subject matter expert within electricity transmission control that revealed that a delay of five seconds between speakers would be perceived by operators as unnaturally long and, therefore, any experimental effect created using longer delays was unlikely to be useful in practical settings, and; (2) previous research which indicated that an inter-turn delay of more than one second was sufficient to signal problematic communication (Jefferson, 1989; Roberts, Francis, & Morgan, 2006). Nevertheless, it may be the case that an inter-turn delay must exceed 1200 ms to be perceived

[2] The study described in Paper 1 was an exception to this. It investigated the utility of inter-turn delay durations of -300 ms (an overlap in speakers of 300 ms), 300 ms, 600 ms and 1200 ms. As null results for inter-turn delay were evident in Paper 1, subsequent papers incorporated the durations of 600 ms and 1200 ms to reduce the complexity of the design as well as allowing the prosodic feature of 'filler' to be incorporated in each level of the variable inter-turn delay.

as indicative of a nonunderstanding within technical environments.

Alternatively, it is possible that an inter-turn delay between speakers, irrespective of duration, is not a reliable cue to gauge receivers' level of understanding. In contrast to the context in which perceptions of uncertainty were investigated (i.e., a trivia 'game' setting), a long delay between speakers within technical environments could feasibly have been caused by a variety of operational events that are not necessarily an indicator of the cognitive state of the receiver. Consequently, an extended inter-turn delay may not reliably cue an operator to a speakers' state of mind.

Practical implications.

The outcomes of this thesis demonstrate that both intonation and filler are salient prosodic cues that appear to be used by operators during perceptions of nonunderstanding in both partial and full readback responses. This outcome has practical implications for the design of training interventions within high-risk, high-consequence environments. Although superior situation assessment skills are typically acquired following extensive experience and practice in the domain (Ericsson & Charness, 1994), the process of cue acquisition can be accelerated through deliberate practice, such as occurs in cue-based training (Ericsson, 2004).

Anderson's (1982) Adaptive Control of Thought – Rational (ACT*R) framework suggests that the development of cognitive skills is a process of strengthening production rules through experience. This model represents a theoretical basis for the utility of cue-based training interventions to accelerate learning and the development of domain-specific production rules (Smeeton, Williams, Hodges, & Ward, 2005). Cue-based training methods generally involve highlighting the critical cues to novices through explicit instruction, and allowing them to integrate and apply this knowledge during relevant situation assessment tasks (Wiggins & O'Hare, 2003b).

Several studies have investigated the practical utility of such cue-based training interventions for enhancing situation assessment skills in various domains, including sport

(Abernethy, Wood, & Parks, 1999; Smeeton et al., 2005; Williams, Ward, Knowles, & Smeeton, 2002), mining (Blignaut, 1979a), organisational settings (Frederiksen, Kehoe, & Wood, 2011; Gary & Wood, 2011), aviation (Wiggins & O'Hare, 2003b) and general medicine (Litchfield, Ball, Donovan, Manning, & Crawford, 2010). For example, Williams et al. (2002) investigated whether cue-based training would improve the ability of a novice tennis player to anticipate the direction of an opponent's tennis stroke.

The cue-based training implemented by Williams et al. (2002) involved both a laboratory and field test phase. The laboratory test phase focused on teaching participants to recognise the critical cues involved in anticipating stroke direction. This involved exposing participants to a series of film clips of a regional level tennis player performing forehand and backhand shots, where the critical features indicating stroke direction ('event') were highlighted to participants through the use of freeze frame and slow motion video playback. Following the cue instruction, participants were then able to practice responding to comparable film clips, and were provided with detailed performance feedback.

The second part of the training intervention took place on-court, where participants were required to react to a club-level male player's forehand and backhand strokes as akin to a match situation. The participants' actions were captured on digital video camera. According to Anderson's (1993) ACT-R model, this type of training is hypothesized to accelerate the learning of production rules. This training intervention allowed participants to practice relevant production rules, such that linkages between feature-event associations (e.g. IF the players hips and shoulders do not rotate as much relative to the position in other shots [feature], the player is going to hit a drop shot [event]) and appropriate actions (THEN run to the front of the net; Smeeton et al., 2005) were established. Indeed, the results indicated that those who received the laboratory cue-based training improved their anticipatory performance in the field compared to a control and placebo group, who did not receive the laboratory training.

Consistent with this approach, there is the potential for the prosodic cues identified in the current research to be incorporated into a cue-based training program. Similar to the approach described in Williams et al., the proposed training program might involve exposing participants to a series of auditory readback/hearback communicative exchanges between technical operators. The important auditory cues indicative of a nonunderstanding on the part of the receiver would be highlighted to participants through the use of pause and playback facilities. Following exposure, participants would be asked to listen to similar exchanges and indicate the extent to which they perceived that a nonunderstanding had taken place and what action, if any, would be taken as a result. Participants would be provided with feedback as to whether the appropriate auditory cues were used to judge understanding and any questions would be answered. This proposed cue-based training program would aim to strengthen the relationship between prosodic features and perceptions of nonunderstanding, thereby potentially increasing the recognition of communication-related errors in technical environments.

Semantic Cues and the Recognition of Nonunderstandings During Readback Responses

Theoretical and practical implications.

The outcomes of this program of research indicate that a receiver who responded to an instruction with a full readback, rather than a partial readback, was perceived by operators as possessing greater levels of understanding. This suggests that the semantic cue that is embodied in a full readback provides a ‘listener’ with evidence of understanding on the part of a ‘speaker’ during the grounding process of communication. However, in some situations, this may not be a reliable cue, since receivers can repeat an instruction correctly, but not necessarily understand the meaning of the instruction. What might be necessary is a training approach where operators can consider *both* an accurate readback and the prosodic cues when assessing issues of understanding, particularly in high risk, high consequence situations.

Typologies and Prosodic and Semantic Cue Utilisation

The RPD model (Klein, 1993, 1997) and the Lens Model (Brunswik, 1955) provide an explanation as to how high-performers are able to make rapid, yet reasonable responses within high-risk environments without the need for in-depth analysis. Expert performance is thought to be facilitated through the extraction of cues in the operational environment, where cues are features of the environment that generate meaning to operators as a result of their association in memory with specific events or objects (Klein, 1989; Wiggins, 2012). The recognition of a familiar pattern of cues within a situation is expected to activate an action script in memory, thereby priming an appropriate response based on past experiences (Klein, 1989, 2003).

Although empirical research generally supports the proposition that experienced operators are more skilled than their less experienced counterparts in their utilisation of visual cues during situation assessment (Blignaut, 1979a; Coderre, Mandin, Harasym, & Fick, 2003; Wiggins & O'Hare, 2003a; Williams, Hodges, North, & Barton, 2006), the outcomes of the current program of research were mixed in relation to the use of auditory cues during assessments of nonunderstanding.

Three experiments were designed to examine whether there was a relationship between visual cue utilisation and the use of prosodic (Papers 2, 3 and 4) and semantic (Paper 4) cues during assessments of nonunderstanding. Although there was no relationship between visual and prosodic cue utilisation in all but one of the experiments, a relationship was evident between visual cue utilisation and the use of semantic cues during assessments of nonunderstanding. Since cues are largely domain-specific, it may be the case that the cues were or were not triggered, depending upon their utility within a particular context. Moreover, the scenarios themselves may not have embodied the ecological validity necessary to trigger the application of a cue. Arguably, the extended Auditory Readback Task described in Paper 4 best encapsulated the range of responses that operators might experience during actual

operational communicative exchanges.

Importantly, the relationship between prosodic cues and perceptions of understanding were examined in isolation during the experiments, rather than being embedded within a broader communicative context. While previous research has suggested that isolated cues can be used to assess the system state (Ebbesen & Konečni, 1975), more often, it will be the recognition of a pattern of familiar cues that generates the recognition of a 'good enough' response within a situation (Schriver et al., 2008).

An alternative explanation is that, unlike visual cue utilisation, prosodic cue utilisation during assessments of nonunderstanding is a generalised skill unrelated to a specific domain. For example, a telephone exchange between two individuals represents a similar context to the readback/hearback protocol, where information is exchanged for mutual understanding without the aid of visual information. As intonation, filler and inter-turn delay are prosodic cues that permeate many different exchanges in everyday life, it is possible that many participants, irrespective of their domain-related experience, would have had opportunities to develop cue associations between intonation and filler (features) and nonunderstandings (event). Therefore, irrespective of their relative visual cue utilisation performance, participants may have acquired the skills necessary to detect nonunderstandings simply through everyday interactions.

Limitations and Future Directions

This program of research was the first to demonstrate the application of specific auditory cues during assessments of nonunderstanding in applied industrial settings. However, it is important to note that the approach in this program of research was experimental and cross-sectional, which may have impacted the nature of the results acquired. For example, it remains unclear whether prosodic cues are actually employed in the operational context to signal nonunderstanding. This is a particularly important consideration given the conclusion that the application of prosodic cues is likely to constitute a generalised capability that occurs

irrespective of the operational milieu.

Experimental artefacts were also created through the subjective nature of assessments of nonunderstanding, together with the seven-point scale along which participants were asked to rate levels of nonunderstanding. Evidence to support the potential restriction of outcomes can be drawn from the fact that responses, particularly in terms of intonation and filler, were constrained to upper end of the 7-point scale of understanding. This pattern is consistent with Brennan and Williams' (1995) investigation into perceptions of uncertainty, where the ratings ascribed were constrained to the upper half of the 7-point scale.

Another potential artefact was created through the requirement for participants to listen to exchanges without being involved in the exchange itself. This process is distinct from actual operational communication where the operator is part of the exchange (i.e., the 'sender'), and not a third-party listener to the exchange. Keysar and Henly (2002) have shown that, compared to a third-party listener, the cognitive demands inherent in speaking leads speakers to overestimate their communication effectiveness, suggesting that 'real life' senders may be less likely to perceive a communication error should this occur within the operational context.

To simulate the cognitive demands associated with involvement in a communicative exchange, the experiments described in Papers 3 and 4 incorporated a secondary task to be completed concurrently while listening to the communicative exchanges. Although future research is needed to established whether perceptions of nonunderstandings differ between operators involved in an exchange, and third-party listeners, the outcomes of the present program of research demonstrates that, under the cognitive load imposed by a concurrent task, intonation and filler are important prosodic cues in assessments of nonunderstanding.

Nonunderstandings, especially those that are concealed, represent a very complex issue when considered within the context of operational environments. While the outcomes of the current program of research were significant insofar as it highlighted the auditory cues that

could potentially improve the recognition of nonunderstandings in the context of readback responses, advancements in future technology may potentially change the nature of the voice-only radiotelephone communication processes inherent within many technical environments. For example, NASA and the Federal Aviation Administration are considering the implementation of a “data link” system in lieu of the current readback/hearback protocol (Helleberg & Wickens, 2003; Schneider, Healy, & Barshi, 2004). This type of system will enable the electronic transmission of text-based information from one operator (e.g., the ground-based ATC) to another operator (e.g., the airborne pilot) through a visual, text-based display (Kerns, 1991). Instead of a verbal readback as required in the current readback system, a key press would be used by the receiving operator to acknowledge the receipt and comprehension of the instruction. This proposed system is expected to protect against problems inherent within the readback/hearback process, such as overloading working memory capacity and the comprehension of voice communication (Helleberg & Wickens, 2003).

The implementation of such a data link system would inevitably remove the ability for senders to rely on the auditory cues of a readback response to gauge levels of understanding on the part of the receiver. As a result, it is important for future research to address the other intricacies involved in concealed nonunderstandings to assist in the development of future preventative strategies. For example, it is important to ascertain the actual prevalence of nonunderstandings within operational environments and specifically, the proportion of these incidents that develop into accidents. It is also important to gain a greater understanding of the cognitive processes that precipitate concealed nonunderstandings. An empirical consideration of these research questions will facilitate a broader understanding of the organisational and cultural factors that influence the frequency of concealed nonunderstandings and the monetary and societal costs associated with such communicative errors.

While the current research program addressed the issue of improving the recognition of concealed nonunderstandings within a range of operational environments, it is important to consider the highly experimental nature of the research in this instance. Future investigations should employ a more systemic analysis and consider the interrelationship between the social and technical components of the readback/hearback environment that facilitates successful (or unsuccessful) communication (Walker, Stanton, Salmon, & Jenkins, 2008). There are examples in the literature that explore the interaction between the social and technical factors of a system in an effort to understand how such a combination of factors impacts workplace performance (Chisholm & Ziegenfuss, 1986).

Within the context of coal mining, for example, Trist and Bamforth (1951) considered the potential influence of social factors that led to a reduction in productivity and increased absenteeism after the implementation of a new form of production technology. Using a largely qualitative approach, the potential interactions between social and technical factors were explored extensively through discussions and interviews with key informants within the work industry over a two-year period. Amongst other factors, it was evident that the differences in both the social organisation of work groups and the complexity of job tasks between the old and new forms of production technology were contributors to the productivity decrease. In a more recent study, Jenkins, Stanton, Salmon, Walker and Young (2008) explored the social and organisational constraints inherent in the design of a software system used in mission planning within military aviation. Through a series of interviews with subject matter experts regarding the design of the system and its use in a practical setting, the researchers were able to identify potential strategies to improve system performance.

A similar, qualitative approach could highlight the organisational and cultural factors involved in operator decisions to conceal nonunderstandings (such as a preference to save face when confronted with an ambiguous operational instruction). Participation from operators with different operational and cultural backgrounds and different experiences of

concealing or detecting nonunderstandings within readback responses could assist in identifying operational and cultural processes that potentially hinder successful communication.

Overall, the outcomes of the current thesis identified specific auditory cues that can potentially aid the recognition of nonunderstandings during radio/telephone exchanges. Future empirical investigations should adopt a more qualitative approach to gain a broader understanding of this complex issue. The outcomes of such research could corroborate and extend the outcomes of the current thesis, and have significant implications for the safety and reliability of communicative protocols employed within high-risk, high-consequence environments.

Research Strengths

The experimental approach taken in the current program of research had several strengths. Firstly, each study comprised experienced professionals across different domains to replicate the key findings. By recruiting participants from different organisations, domains and countries, the current program of research highlights the utility of particular auditory cues to aid the recognition of nonunderstandings across a variety of organisational and technical cultures.

Secondly, the current thesis utilised ecologically appropriate stimuli developed from cognitive interviews with subject matter experts to discern distinct levels of operator performance in each domain. It was critical to use naturalistic stimuli that accurately reflected the domain of interest, as previous studies have been unable to elicit distinct levels of performance due to their inability to replicate real-world conditions (Kahneman & Klein, 2009; Norman, Coblentz, Brooks, & Babcook, 1992). This is arguably because non-domain specific tasks do not facilitate the use of domain-specific cue-associations, and as a result, experienced operators may perform at a level akin to that of a novice (Brunswik, 1955).

With reference to the auditory stimuli, assessments of nonunderstanding were based on

communicative exchanges that embodied a degree of ecological validity in their structural and semantic characteristics. Further, the domains of interest were selected on the basis that task-relevant visual cues were presented to operators via standardised machine interfaces during normal operations. This allowed the current thesis to incorporate naturalistic stimuli, which was developed and refined in conjunction with subject matter experts in each domain, while still maintaining experimental control. Consequently, the differentiation of participants on the basis of their performance across the naturalistic tasks can be interpreted with a degree of confidence, where meaningful inferences can be drawn with regards to operator performance.

Finally, each study incorporated a robust design and demonstrated experimental control of potential confounds, such as the vocal characteristics of the male speaker who played the role of 'receiver' in the communicative exchanges, and the imposition of additional cognitive load during the Auditory Readback Task. Overall, the theoretical and practical outcomes of the current thesis can be interpreted with some confidence due to the robust design of each study, including the recruitment of experienced operators from various high-risk, high-consequence environments.

General Conclusion

The current thesis incorporated four papers that investigated the utility of specific auditory cues to assess nonunderstandings within readback responses across various operational domains. Unlike prior studies in which skilled cue utilisation has been assessed on the basis of visual cues in isolation, the present research also investigated whether a relationship existed between auditory and visual cue utilisation.

The use of a 'filler' was identified consistently as an important prosodic cue in the detection of nonunderstandings during full and partial readback responses across several operational domains. While similar effects were evident for intonation, the relationship between intonation and nonunderstandings was not evident across all domains. Finally, semantic cues proved useful to operators when assessing levels of understanding on the part

of receivers. Overall, the current thesis highlighted the utility of intonation and filler (prosodic cues), together with a full readback (semantic cue) during assessments of nonunderstanding within operational readback/hearback exchanges.

Less consistent results were evident for the relationship between visual cue utilisation and auditory cue utilisation, particularly in electricity transmission control and electrical rail control. However, in the context of hydroelectric power generation, a relationship was evident between visual and auditory cue utilisation. In combination, the results suggest that, unlike visual cue utilisation, the use of prosodic cues is a generalised skill that occurs independent of organisational context.

Complete Reference List

Abernethy, B. (1990). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sports Sciences*, *8*, 17–34. doi:10.1080/02640419008732128

Abernethy, B., Wood, J. M., & Parks, S. (1999). Can the anticipatory skills of experts be learned by novices? *Research Quarterly for Exercise and Sport*, *70*, 313-318. doi:10.1080/02701367.1999.10608050

Air Accidents Investigation Branch. (2008). *Boeing 737-500, SP-LKA* (Report No. EW/C2007/06/02). Retrieved from http://www.aaib.gov.uk/publications/bulletins/june_2008/boeing_737_500__sp_lka.cfm

Alderson, J. C. (2009). Air safety, language assessment policy, and policy implementation: The case of aviation English. *Annual Review of Applied Linguistics*, *29*, 168–187. doi:10.1017/S0267190509090138

Allwood, J., & Abelar, Y. (1984). Lack of understanding, misunderstanding and language acquisition. *Proceedings of the AILA-Conference, Brussels*, 1–19. Retreived from http://www.sskkii.gu.se/jens/publications/bfiles/B14.pdf

Anderson, J. R. (1982). Acquisition of a cognitive skill. *Psychological Review*, *89*, 369–406. doi:10.1037/0033-295X.89.4.369

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261–295. doi:10.1016/S0022-5371(83)90201-3

Anderson, J. R. (1987). Skill acqusition: Compilation of weak-method problem solutions. *Psychological Review*, *94*, 192–210. doi:10.1037/0033-295X.94.2.192

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.

Apple Inc. (2014). GarageBand (6.0.4) [Digital Audio Workstation and Music Sequencer]. Retrieved from https://itunes.apple.com/au/app/garageband/id682658836?mt=12&ls=1

Audacity (1.2.5) [Free Audio Editor and Recorder Software]. Retrieved from http://audacity.sourceforge.net/

Australian Transport Safety Bureau. (2007). *Radiotelephony readback compliance and its relationship to surface movement control frequency congestion* (Investigation Report No. B2006/0053). Retrieved from http://www.atsb.gov.au/media/29962/b20060053.pdf

Auton, J. C., Wiggins, M. W., & Loveday, T. (2012). Prosodic cues that signal non-understandings to power control operators during radio communication. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*, 483–487. doi:10.1177/1071181312561049

Auton, J. C., Wiggins, M. W., & Searle, B. J. (2014). The relationship between visual cue utilisation and prosodic cue utilisation during assessments of nonunderstanding in readback responses. *Unpublished Manuscript, Department of Psychology, Macquarie University, Sydney, Australia.*

Auton, J. C., Wiggins, M. W., Searle, B. J., Loveday, T., & Xu Rattanasone, N. (2013). Prosodic cues used during perceptions of nonunderstandings in radio communication. *Journal of Communication*, *63*, 600–616. doi:10.1111/jcom.12033

Bachorowski, J.-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science*, *8*, 53–57. doi:10.1111/1467-8721.00013

Bachorowski, J.-A., & Owren, M. J. (2006). Sounds of emotion: Production and perception of affect-related vocal acoustics. *Annals of the New York Academy of Sciences*, *1000*, 244–265. doi:10.1196/annals.1280.012

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, *106*, 410–433. doi:10.1037/0033-2909.106.3.410

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636. doi:10.1037/0022-3514.70.3.614

Barshi, I., & Farris, C. (2013). *Misunderstandings in ATC Communication: Lanugage, cognition, and experimental methodolgy*. Burlington, USA: Ashgate Publishing Group.

Baseheart, J. R. (1975). Nonverbal communication in the dentist-patient relationship. *Journal of Prosthetic Dentistry*, *34*, 4–10. doi:10.1016/0022-3913(75)90126-2

Bazzanella, C., & Damiano, R. (1999). The interactional handling of misunderstanding in everyday conversations. *Journal of Pragmatics*, *31*, 817–836. doi:10.1016/S0378-2166(98)00058-7

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309. doi:10.1017/S095267570000066X

Berger, C. R. (2011). From explanation to application. *Journal of Applied Communication Research, 39,* 214-222. doi:10.1080/00909882.2011.556141

Billings, C. E., & Cheaney, E. S. (1981). The information transfer problem: Summary and comments. In C. E. Billings & E. S. Cheaney (Eds.), *Information transfer problems in the Aviation system* (Report No. NASA TP-1875). Washington, DC: National Aeronautics and Space Administration.

Bisantz, A. M., & Pritchett, A. R. (2003). Measuring the fit between human judgments and automated alerting algorithms: A study of collision detection. *Human Factors*, *45*, 266–280. doi:10.1518/hfes.45.2.266.27243

Blignaut, C. J. H. (1979a). The perception of hazard: I. Hazard analysis and the contribution of visual search to hazard perception. *Ergonomics*, *22*, 991–999. doi:10.1080/00140137908924673

Blignaut, C. J. H. (1979b). The perception of hazard: II. The contribution of signal detection to hazard perception. *Ergonomics, 22,* 1177-1183. doi:10.1080/00140137908924692

Bolinger, D. (1972). *Intonation*. Great Britain: Hazell Watson & Viney Ltd.

Bolinger, D. (1985). *Intonation and its parts*. London, UK: Edward Arnold.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383–398. doi:10.1006/jmla.1995.1017

Brenner, M., Shipp, T., Doherty, E. T., & Morrisey, P. (1985). Voice measures of psychological stress - Laboratory and field data. In I. R. Titze & R. C. Scherer (Eds.), *Vocal fold physiology, biomechanics, acoustics, and phonatory control* (pp. 239–24). Denver: The Denver Center for the Performing Arts.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217. doi:10.1037/h0047470

Calderwood, R., Klein, G. A., & Crandall, B. W. (1988). Time pressure, skill, and move quality in chess. *The American Journal of Psychology*, *101*, 481–493. doi:10.2307/1423226

Cardosi, K. M. (1993). *An analysis of en route controller-pilot voice communications* (Report No. DOT/FAA/RD-93/11). Washington, DC: Research and Development Administration Transportation Systems Center.

Cardosi, K., Brett, B., & Han, S. (1996). *An analysis of TRACON (Terminal Radar Approach Control) controller-pilot voice communications* (DOT Report No. DOT/FAA/AR-96/66). Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Cellier, J. M., Eyrolle, H., & Mariné, C. (1997). Expertise in dynamic environments. *Ergonomics*, *40*, 28–50. doi:10.1080/001401397188350

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81. doi:10.1016/0010-0285(73)90004-2

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7–75). Hillsdale, NJ: Erlbaum.

Chiang, H.-Y., & Pepper, G. A. (2006). Barriers to nurses' reporting of medication administration errors in Taiwan. *Journal of Nursing Scholarship, 38,* 392–399. doi:10.1111/j.1547-5069.2006.00133.x

Chisholm, R. F., & Ziegenfuss, J. T. (1986). A review of applications of the sociotechnical systems approach to health care organizations. *The Journal of Applied Behavioral Science, 22,* 315–327. doi:10.1177/002188638602200310

Clark, H. H. (1994). Managing problems in speaking. *Speech Communication, 15,* 243–250. doi:10.1016/0167-6393(94)90075-2

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84,* 73–111. doi:10.1016/S0010-0277(02)00017-3

Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes, 2,* 19–41. doi:10.1080/01690968708406350

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13,* 259–294. doi:10.1207/s15516709cog1302_7

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39. doi:10.1016/0010-0277(86)90010-7

Clark, H. H., & Wilkes-Gibbs, D. (1990). Referring as a collaborative process. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 463–493). Cambridge, MA: MIT Press.

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education, 37,* 695–703. doi:10.1046/j.1365-2923.2003.01577.x

Cohen, P. R., & Levesque, H. J. (1994). Preliminaries to a collaborative model of dialogue. *Speech Communication, 15*, 265–274. doi:10.1016/0167-6393(94)90077-9

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine, 84*, 1022–1028. doi:10.1097/ACM.0b013e3181ace703

Cruttenden, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics, 17*, 77–91. doi:10.1017/S0022226700006782

Cruttenden, A. (1986). *Intonation*. Cambridge, UK: Cambridge University Press.

Crystal, D. (1969). *Prosodic systems and intonation in English.* Cambridge: Cambridge University Press.

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

Cushing, S. (1995). Pilot-Air traffic Control Communications: It's not (only) what you say, it's how you say it. *Flight Safety Digest, 14*(7), 1–10.

Dascal, M., & Berenstein, I. (1987). Two modes of understanding: Comprehending and grasping. *Language and Communication, 7*, 139–151. doi:10.1016/0271-5309(87)90004-8

De Groot, A. D. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

Denes, P., & Milton-William, J. (1962). Further studies in intonation. *Language and Speech, 5*, 1–14. doi:10.1177/002383096200500101

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74-112. doi:10.1037/0033-2909.129.1.74

Dromey, C., Silveira, J., & Sandor, P. (2005). Recognition of affective prosody by speakers of English as a first or foreign language. *Speech Communication, 47*, 351–359. doi:10.1016/j.specom.2004.09.010

Drummond, K., & Hopper, R. (1991). Misunderstanding and its remedies: Telephone miscommunication. In N. Coupland, H. Giles, & J. Wiemann (Eds.), *"Miscomunication" and problematic talk* (pp. 301–314). Newbury Park, CA: Sage.

Ebbesen, E. B., & Konečni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology*, *32*, 805–821. doi:10.1037/0022-3514.32.5.805

Eccles, D. W., Walsh, S. E., & Ingledew, D. K. (2002). The use of heuristics during route planning by expert and novice orienteers. *Journal of Sports Sciences*, *20*, 327–337.

Edelsky, C. (1979). Question intonation and sex roles. *Language in Society*, *8*, 15–32. doi:10.1017/S0047404500005923

Elkerton, J., & Williges, R. C. (1985). A performance profile methodology for implementing assistance and instruction in computer-based tasks. *International Journal of Man-Machine Studies*, *23*, 135–151. doi:10.1016/S0020-7373(85)80029-8

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32–64. doi:10.1518/001872095779049543

Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 269–284). Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine*, *79*, S70–S81. doi:10.1097/00001888-200410001-00022

Ericsson, K. A., & Charness, N. (1994). Expert performance: It's structure and acquisition. *American Psychologist*, *49*, 725–747. doi:10.1037/0003-066X.49.8.725

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory, *102*(2), 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406. doi:10.1037/0033-295X.100.3.363

Ettenson, R., Shanteau, J., & Krogstad, J. (1987). Expert judgment: Is more information better? *Psychological Reports*, *60*, 227–238. doi:10.2466/pr0.1987.60.1.227

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–78. doi:10.1146/annurev.psych.59.103006.093629

Fay, N., Page, A. C., & Serfaty, C. (2010). Listeners influence speakers' perceived communication effectiveness. *Journal of Experimental Social Psychology*, *46*, 689–692. doi:10.1016/j.jesp.2010.02.012

Federal Aviation Administration. (2014). *Aeronautical information manual*. Washington, DC: Government Printing Office. Retrieved from http://www.faa.gov/air_traffic/publications/atpubs/aim/index.htm

Fegyveresi, A. E. (1997). *Vocal cues and pilot/ATC communication*. Paper presented at the Ninth International Symposium on Aviation Psychology, Columbus, OH. Abstract retreived from http://md1.csa.com/partners/viewrecord.php?requester=gs&collection=TRD&recid=A9828885AH&q=FEGYVERESI+1997&uid=1000599&setcookie=yes

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 327–358. doi:10.1037/h0061470

Fletcher, J., & Harrington, F. (2001). High-rising terminals and fall-rise tunes in Australian English. *Phonetica*, *58*, 215–229. doi:10.1159/000046176

Frederiksen, C., Kehoe, E. J., & Wood, R. (2011). Effects of instructional aids on the acquisition of dynamic decision-making skills. *Learning and Instruction*, *21*, 601–613. doi:10.1016/j.learninstruc.2011.01.002

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*, 412–429. doi:10.1037/0033-2909.97.3.412

Gary, M. S., & Wood, R. E. (2011). Mental models, decision rules and performance heterogeneity. *Strategic Management Journal*, *594*, 569–594. doi:10.1002/smj.899

Garzone, G., Catino, M., Gobo, G., Bait, M., Catenaccio, P., Degano, C., & Rozzi, S. (2010). Towards an integrated model for the understanding of communication failures in aviation accidents: Tenuous identities under pressure. In G. Garzone & J. Archibald (Eds.), *Discouse, identities and roles in specialized communication* (pp. 209–244). Bern, Switzerland: International Academic Publishers.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gobl, C., & Chasaide, A. N. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*, 189–212. doi:10.1016/S0167-6393(02)00082-1

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language and Speech*, *4*, 18–26. doi:10.1177/002383096100400102

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*, 519–532. doi:10.1177/001872088402600504

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*, 42–54. doi:10.1037/1076-898X.10.1.42

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*, 42–54. doi:10.1037/1076-898X.10.1.42

Grayson, R. L., & Billings, C. E. (1981). Information transfer between air traffic control and aircraft: Communication problems in flight and operations. In C. E. Billings & E. S. Cheaney (Eds.), *Information Transfer Problems in the Aviation System* (Report No. NASA TP-1875) (pp. 47–62). Washington, DC: National Aeronautics and Space Administration.

Greenberg, C. C., Regenbogen, S. E., Studdert, D. M., Lipsitz, S. R., Rogers, S. O., Zinner, M. J., & Gawande, A. A. (2007). Patterns of communication breakdowns resulting in injury to surgical patients. *Journal of the American College of Surgeons*, *204*, 533–540. doi:10.1016/j.jamcollsurg.2007.01.010

Greitzer, F. L., Podmore, R., Robinson, M., & Ey, P. (2010). Naturalistic decision making for power system operators. *International Journal of Human-Computer Interaction*, *26*, 278–291. doi:10.1080/10447310903499070

Griffin, G. R., & Williams, C. E. (1987). The effects of different levels of task complexity on three vocal measures. *Aviation, Space, and Environmental Medicine*, *58*, 1165–1170.

Grimshaw, A. D. (1980). Mishearings, misunderstandings and other nonsuccesses in talk: A plea for redress of speaker-oriented bias. *Sociological Inquiry*, *50*, 31–74. doi:10.1111/j.1475-682X.1980.tb00016.x

Guy, G. R., & Vonwiller, J. (1984). The meaning of an intonation in Australian English. *Australian Journal of Linguistics*, *4*, 1–17. doi:10.1080/07268608408599317

Guy, G. R., Horvath, B., Vonwiller, J., Daisley, E., & Rogers, I. (2011). An intonational change in progress in Australian English. *Language in Society*, *15*, 23–51. doi:10.1017/S0047404500011635

Hall, J. A. (1984). *Nonverbal sex difference: Communication accuracy and expressive style.* Baltimore: John Hopkins University.

Hanna, D., Griswold, P., Leape, L. L., & Bates, D. W. (2005). Communicating critical test results: Safe practice recommendations. *Joint Comission Journal on Quality and Patient Safety*, *31*(2), 68–80.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208–216. doi:10.1037/h0022263

Helleberg, J. R., & Wickens, C. D. (2003). Effects of data-link modality and display redundancy on pilot performance: An attentional perspective. *The International Journal of Aviation Psychology*, *13*, 189–210. doi:10.1207/S15327108IJAP1303_01

Hirst, G., McRoy, S., Heeman, P., Edmonds, P., & Horton, D. (1994). Repairing conversational misunderstandings and non-understandings. *Speech Communication*, *15*, 213–229. doi:10.1016/0167-6393(94)90073-6

Hoc, J. M., Amalberti, R., & Boreham, N. (1995). Human operator expertise in diagnosis, decision-making and time management. In E. Hollnagel, P. C. Cacciabue, & J. M. Hoc (Eds.), *Expertise and Technology: Cognition and Human-Computer Cooperation*. New Jersey: Lawrence Erlbaum Associates, Inc.

Hoffman, K. A., Aitken, L. M., & Duffield, C. (2009). A comparison of novice and expert nurses' cue collection during clinical decision-making: Verbal protocol analysis. *International Journal of Nursing Studies*, *46*, 1335-44. doi:10.1016/j.ijnurstu.2009.04.001

Howard, J. W. (2008). "Tower, am I cleared to land?": Problematic communication in Aviation discourse. *Human Communication Research*, *34*, 370-391. doi:10.1111/j.1468-2958.2008.00325.x

Hunter, M. (2004). The music is the message: Prosody in aviation discourse. In M. A. Turney (Ed.), *Tapping diverse talent in Aviation: Culture, gender, and diversity* (pp. 129–151). Hampshire, England: Ashgate Publishing Ltd.

Huttar, G. L. (1968). Relations between prosodic variables and emotions in normal American English utterances. *Journal of Speech and Hearing Research*, *11*, 481–488. doi:10.1044/jshr.1103.481

International Civil Aviation Organization. (2007). *Manual of Radiotelephony* (Report No. Doc 9432 AN/925). Retreived from http://www.aeroclubvercelli.it/download/Manual%20of%20RT%20DOC%209432%20-%204%20ed.%202007.pdf

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*, 146–154. doi:10.1016/j.learninstruc.2009.02.019

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An Interdisciplinary Perspective* (pp. 166–196). Clevedon, UK: Multilingual Matters.

Jenkins, D. P., Stanton, N. A, Salmon, P. M., Walker, G. H., & Young, M. S. (2008). Using cognitive work analysis to explore activity allocation within military domains. *Ergonomics*, *51*, 798–815. doi:10.1080/00140130801915246

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, *43*, 280–283. doi:10.1001/archpsyc.1986.01800030098011

Johnstone, T., & Scherer, K. R. (2000). Vocal Communication of Emotion. In M. Lewis & J. Haviland (Eds.), *The Handbook of Emotion* (pp. 220–235). New York, NY: Guilford.

Jones, R. K. (2003). Miscommunication between pilots and air traffic control. *Language Problems and Language Planning*, *27*, 233–248. doi:10.1075/lplp.27.3.03jon

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, *1*, 381–412. doi:10.1037//1528-3542.1.4.381

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814. doi:10.1037/0033-2909.129.5.770

Juslin, P. N., & Scherer, K. R. (2008). Vocal Expression of Emotion. In J. A. Harrigan, R. Rosenthal, & K. R. Scherer (Eds.), *The New Handbook of Methods in Nonverbal Behavior Research* (pp. 65–136). New York, NY: Oxford University Press.

Kaempf, G. L., & Klein, G. A. (1994). Aeronautical Decision Making: The next generation. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation Psychology in Practice* (pp. 223–254). Hants, England: Avebury Technical.

Kaempf, G. L., Klein, G. A., Thordsen, Marvin, L., & Wolf, S. (1996). Decision making in complex naval command-and-control environments. *Human Factors*, *38*, 220–231. doi:10.1518/001872096779047986

Kahneman, D. (2003). Maps of Bounded Rationality : Economicst Psychology for Behavioral. *The American Economic Review*, *93*, 1449–1475. doi:10.1257/000282803322655392

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267–294). Cambridge, UK: Cambridge University Press.

Kahneman, D., & Klein, G. A. (2009). Conditions for intuitive expertise: A failure to disagree. *The American Psychologist*, *64*, 515–526. doi:10.1037/a0016755

Kanki, B. G., & Smith, G. M. (2001). Training Aviation communication skills. In E. Salas, C. A. Bowers, & E. Edens (Eds.), *Improving teamwork in organizations* (pp. 31–54). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, *134*, 404–26. doi:10.1037/0033-2909.134.3.404

Kee, F., Jenkins, J., McIlwaine, S., Patterson, C., Harper, S., & Shields, M. (2003). Fast and frugal models of clinical judgment in novice and expert physicians. *Medical Decision Making*, *23*, 293–300. doi:10.1177/0272989X03256004

Kerns, K. (1991). Data-link communication between controllers and pilots: A review and synthesis of the simulation literature. *The International Journal of Aviation Psychology*, *1*, 181–204. doi:10.1207/s15327108ijap0103_1

Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, *13*, 207–212. doi:10.1111/1467-9280.00439

Kim, A., Park, J., Han, S., & Kim, H. (2010). Development of extended speech act coding scheme to observe communication characteristics of human operators of nuclear power plans under abnormal conditions. *Journal of Loss Prevention in the Process Industries*, *23*, 539-548. doi:10.1016/j.jlp.2010.04.005

Klein, G. A. (1989). Recognition-primed decisions (RPD). *Advances in Man-Machine Systems*, *5*, 47–92.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138–147). Norwood, NJ: Ablex.

Klein, G. A. (1997). The Recognition-Primed Decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 285–292). Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, G. A. (1999). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Klein, G. A. (2003). *Intuition at work.* New York: Currency Books.

Klein, G. A. (2008). Naturalistic Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*, 456–460. doi:10.1518/001872008X288385

Klein, G. A., & Calderwood, R. (1991). Decision models : Some lessons from the field. *IEEE Transactions on Systems, Man, and Cybernetics*, *21*, 1018–1026. doi:10.1109/21.120054

Klein, G. A., & Klinger, D. (1991). Naturalistic Decision Making. *Human Systems IAC Gateway*, *XI*, 16–19. doi:10.1518/001872008X288385

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. *Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting* (Vol. 1, pp. 576–580). Norwood, NJ: Ablex.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1988). *Rapid decision making on the fire ground* (ARI Technical Report 796). Yellow Springs, OH: U.S. Army Research Institute for the Behavioral and Social Sciences.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*, 462–472. doi:10.1109/21.31053

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2014). *Nonverbal communication in human interaction* (8th Ed.). Boston, MA: Monica Ekman.

Krahmer, E., & Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Language and Speech*, *48*, 29–53. doi:10.1177/00238309050480010201

Krieger, J. L. (2005). Shared mindfulness in cockpit crisis situations: An exploratory analysis. *Journal of Business Communication*, *42*, 135-167. doi:10.1177/0021943605274726

Ladd, D. R. (1980). *The structure of intonational meaning*. Bloomington: Indiana University Press.

Lakoff, R. (1973). Language and woman's place. *Language in Society*, *2*, 45–80. doi:10.1017/S0047404500000051

Lau, R. R., & Redlawsk, D. P. (2001). Advantages and disadvantages of cognitive heuristics in political decision making. *American Journal of Political Science*, *45*, 951–971. doi:10.2307/2669334

Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, *33*, 37-63. doi:10.2307/3588190

Levis, J. M. (2002). Reconsidering low-rising intonation in American English. *Applied Linguistics*, *23*, 56–82. doi:10.1093/applin/23.1.56

Lipshitz, R. (1993). Converging themes in the study of decision making in realistic settings. In G. A. Klein, J. M. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 103–137). Norwood, New Jersey: Ablex Publishing Corporation.

Lipshitz, R., Klein, G. A., Orasanu, J., & Salas, E. (2001). Focus article: Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*, 331–352. doi:10.1002/bdm.381

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *Journal of Experimental Psychology: Applied*, *16*, 251–262. doi:10.1037/a0020082

Loveday, T., Wiggins, M. W., & Searle, B. J. (2013). Cue utilization and broad indicators of workplace expertise. *Journal of Cognitive Engineering and Decision Making*, *8*, 98–113. doi:10.1177/1555343413497019

Loveday, T., Wiggins, M. W., Festa, M., Schell, D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In P. Latorre Carmona, J. S. Sánchez, & A. L. N. Fred (Eds.), *Pattern recognition - Applications and methods* (Vol. 204, pp. 1–11). Berlin, Heidelberg: Springer Berlin Heidelberg.

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2013). An objective approach to identifying diagnostic expertise among power system controllers. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55,* 90–107. doi:10.1177/0018720812450911

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2013). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55,* 125–137. doi:10.1177/0018720812448475

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review, 3,* 215–221. doi:10.3758/BF03212421

McGregor, J., & Palethorpe, S. (2008). High rising tune in Australian English: The communication function of L* and H* pitch accent onsets. *Australian Journal of Linguistics, 28,* 171-193. doi:10.1080/07268600802308766

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63,* 81–97. doi:10.1037/h0043158

Monan, W. P. (1986). *Human factors in aviation operators: The hearback problem.* (NASA Contractor Report 177398). Moffett Field, California: Ames Research Center.

Monnot, M., Nixon, S., Lovallo, W., & Ross, E. (2001). Altered emotional perception in alcoholics: Deficits in affective prosody comprehension. *Alcoholism: Clinical and Experimental Research, 25,* 362–369. doi:10.1111/j.1530-0277.2001.tb02222.x

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2012). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making, 7,* 211–226. doi:10.1177/1555343412459192

Morrison, R., & Wright, R. H. (1989, April). ATC Control and communications problems: An overview of recent ASRS data. In R. S. Jensen (Ed), *Fifth Symposium on Aviation Psychology* (Vol 2). Columbus, Ohio, U.S.

Morrow, D., Lee, A., & Rodvold, M. (1993). Analysis of problems in routine controller-pilot communication. *The International Journal of Aviation Psychology*, *3*, 285–302. doi:10.1207/s15327108ijap0304_3

Morrow, D., Rodvold, M., & Lee, A. (1994). Nonroutine transactions in controller- pilot communication. *Discourse Processes*, *17*, 235–258. doi:10.1080/01638539409544868

Müller, S., Abernethy, B., & Farrow, D. (2006). How do world-class cricket batsmen anticipate a bowler's intention? *The Quarterly Journal of Experimental Psychology*, *59*, 2162–2186. doi:10.1080/02643290600576595

Murray, S. R. (1999). FACE: Fear of loss of face and the five hazardous attitudes concept. *The International Journal of Aviation Psychology*, *9*, 403–411. doi:10.1207/s15327108ijap0904

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 553–557. doi:10.1037/0278-7393.14.3.553

Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2007). *Straight choices: The psychology of decision making*. New York, NY: Psychology Press.

Norman, G., Young, M., & Brooks, L. (2007). Non-analytical models of clinical reasoning: the role of experience. *Medical Education*, *41*(12), 1140–1145.

Norman, G. R., Coblentz, C. L., Brooks, L. R., & Babcook, C. J. (1992). Expertise in visual diagnosis: A review of the literature. *Academic Medicine*, *67*(10), S78–S83.

North, J. S., Williams, A. M., Hodges, N., Ward, P., & Ericsson, K. A. (2009). Perceiving patterns in dynamic action sequences: Investigating the processes underpinning stimulus

recognition and anticipation skill. *Applied Cognitive Psychology*, *23*, 878–894. doi:10.1002/acp.1581

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, *39*, 84–97. doi:10.1016/0749-5978(87)90046-X

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1017–1030. doi:10.1037/0096-1523.34.4.1017

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*, 127–146. doi:10.1111/j.1551-6709.2008.01007.x

O'Connor, J. D., & Arnold, G. F. (1961). *Intonation of colloquial English*. London, UK: Longman.

O'Hare, D., Wiggins, M., Williams, M., & Wong, W. (1998). Cognitive task analyses for decision centred design and training. *Ergonomics, 41,* 1698-1718. doi:10.1080/001401398186144

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 3–20). Norwood, NJ: Ablex Publishing Corporation.

Paas, F. G. W. C., & Merriënboer, J. J. G. Van. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122–133. doi:10.1037/0022-0663.86.1.122

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63–71. doi:10.1207/S15326985EP3801_8

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology). Retreived from http://hdl.handle.net/1721.1/16065

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Prabhakar, H., Cooper, J. B., Sabel, A., Weckbach, S., Mehler, P. S., & Stahel, P. F. (2012). Introducing standardized "readbacks" to improve patient safety in surgery: A prospective survey in 92 providers at a public safety-net hospital. *BMC Surgery*, *12*, 1–8. doi:10.1186/1471-2482-12-8

Prinzo, O. V., & Britton, T. W. (1993). ATC/pilot voice communications: A survey of the Literature (Report No. DOT/FAA/AM-93/20). Washington, DC: Office of Aviation Medicine.

Prinzo, O. V., Hendrix, A. M., & Hendrix, R. (2006). The outcome of ATC message complexity on pilot readback performance (Report No. DOT/FAA/AM-06/25). Oklahoma City, OK: FAA Civil Aerospace Medical Institute.

Raghunathan, K. (2012). Checklists, safety, my culture and me. *BMJ Quality & Safety*, *21*, 617–620. doi:10.1136/bmjqs-2011-000608

Randel, J. M., Pugh, H. L., & Reed, S. K. (1996). Differences in expert and novice situation awareness in naturalistic decision making. *International Journal of Human-Computer Studies*, *45*, 579–597. doi:10.1006/ijhc.1996.0068

Rasmussen, J. (1983). Skills, rules, knowledge: Signals, signs, and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*, 257–267. doi:10.1109/TSMC.1983.6313160

Rasmussen, J. (1993). Diagnostic Reasoning in Action. *IEEE Transactions on Systems Man and Cybernetics*, *23*, 981–992. doi:10.1109/21.247883

Rieger, C. L. (2003). Repetitions as self-repair strategies in English and German conversations. *Journal of Pragmatics*, *35*, 47–69. doi:10.1016/S0378-2166(01)00060-1

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of “trouble” in conversation. *Speech Communication*, *48*, 1079–1093. doi:10.1016/j.specom.2006.02.001

Schegloff, E. A. (2000). When “others” initiate repair. *Applied Linguistics*, *21*, 205–243. doi:10.1093/applin/21.2.205

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*, 361–382. doi:10.1353/lan.1977.0041

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*, 143–165. doi:10.1037/0033-2909.99.2.143

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227–256. doi:10.1016/S0167-6393(02)00084-5

Schlanger, B. B., Schlanger, P., & Gerstman, L. J. (1976). The perception of emotionally toned sentences by right hemisphere-damaged and aphasic subjects. *Brain and Language*, *3*, 396–403. doi:10.1016/0093-934X(76)90035-3

Schmidt, H. G., & Boshuizen, P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review*, *5*, 205–221. doi:10.1007/BF01323044

Schneider, V. I., Healy, A. F., & Barshi, I. (2004). Effects of instruction modality and readback on accuracy in following navigation commands. *Journal of Experimental Psychology: Applied*, *10*, 245–257. doi:10.1037/1076-898X.10.4.245

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise difference in attentional strategies related to pilot decision making. *Human Factors*, *50*, 864–878. doi:10.1518/001872008X374974.

Scialfa, C. T., Deschênes, M. C., Ference, J., Boone, J., Horswill, M. S., & Wetton, M. (2011). A hazard perception test for novice drivers. *Accident Analysis and Prevention*, *43*, 204–208. doi:10.1016/j.aap.2010.08.010

Searle, B. J., & Auton, J. C. (in press). The merits of measuring challenge and hindrance appraisals. *Anxiety, Stress, and Coping: An International Journal*. doi:10.1080/10615806.2014.931378

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*, 253–263. doi:10.1016/S0377-2217(01)00113-8

Shim, J., Carlton, L. G., Chow, J. W., & Chae, W.-S. (2005). The use of anticipatory visual cues by highly skilled tennis players. *Journal of Motor Behavior*, *37*, 164–175. doi:10.3200/JMBR.37.2.164-175

Shimojima, A., Katagiri, Y., Koiso, H., & Swerts, M. (2002). Informational and dialogue-coordinating functions of prosodic features of Japanese echoic responses. *Speech Communication*, *36*, 113-132. doi:10.1016/S0167-6393(01)00029-2

Simnett, R. (1996). The effect of informaiton selection, information processing and task complexity on predictive accuracy on auditors. *Accounting, Organizations & Society*, *21*, 699–719. doi:10.1016/0361-3682(96)00006-2

Simon, H. A. (1978). Rationality as a process and as product of thought. *The American Economic Review*, *68*(2), 1–16.

Simon, H. A. (1992). What is an “explanation” of behavior? *Psychological Science*, *3*, 150–161. doi:10.1111/j.1467-9280.1992.tb00017.x

Skantze, G. (2005). Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication*, *45*, 325–341. doi:10.1016/j.specom.2004.11.005

Slovic, P., Rorer, L. G., & Hoffman, P. J. (1971). Analyzing use of diagnostic signs. *Investigative Radiology, 6,* 18–26. doi:10.1097/00004424-197101000-00003

Smeeton, N. J., Williams, A. M., Hodges, N. J., & Ward, P. (2005). The relative effectiveness of various instructional approaches in developing anticipation skill. *Journal of Experimental Psychology: Applied, 11,* 98–110. doi:10.1037/1076-898X.11.2.98

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Jornal of Memory and Language, 32,* 25–38. doi:10.1006/jmla.1993.1002

Sniderman, P. M., Brody, R. A., & Tetlock, P. E. (1991). *Reasoning and choice: Explorations in political psychology*. Cambridge: Cambridge University Press.

Soskin, W. F. (1953). Symposium: Communication in the counselling situation. *American Psychologist, 8,* 272.

Speier, C. (2006). The influence of information presentation formats on complex task decision-making performance. *International Journal of Human-Computer Studies, 64,* 1115–1131. doi:10.1016/j.ijhcs.2006.06.007

Stokes, A. F., Kemper, K. L., & Marsh, R. (1992). Time-stressed flight decision making: A study of expert and novice aviators (Report No. ARL-93-1/INEL-93-1). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Stokes, A. F., Kemper, K., & Kite, K. (1997). Aeronautical decision making, cue recogniton, and expertise under time pressure. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 183–196). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Stolper, E., Van De Wiel, M., Van Royen, P., Van Bokhoven, M., Van Der Weijden, T., & Dinant, G. J. (2011). Gut feelings as a third track in general practitioners' diagnostic reasoning. *Journal of General Internal Medicine, 26,* 197–203. doi:10.1007/s11606-010-1524-5

Strauch, B. (2010). Can cultural differences lead to accidents? Team cultural differences and sociotechnical system operations. *Human Factors*, *52*, 246–263. doi:10.1177/0018720810362238

Streitenberger, K., Breen-Reid, K., & Harris, C. (2006). Handoffs in care - Can we make them safer? *Pediatric Clinics of North Americalinics of North America*, *53*, 1185–1195. doi:10.1016/j.pcl.2006.09.010

Sutter, C., & Ziefle, M. (2005). Interacting with notebook input devices: An analysis of motor performance and users' expertise. *Human Factors*, *47*, 169–187. doi:10.1518/0018720053653893

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi:10.1016/0364-0213(88)90023-7

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*, 81–94. doi:10.1016/j.jml.2005.02.003

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the Longwall Method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Human Relations*, *4*, 3–38. doi:10.1177/001872675100400101

Walker, G. H., Stanton, N. A., Salmon, P. M., & Jenkins, D. P. (2008). A review of sociotechnical systems theory: A classic concept for new command and control paradigms. *Theoretical Issues in Ergonomics Science*, *9*, 479–499. doi:10.1080/14639220701635470

Wallis, T. S. A., & Horswill, M. S. (2007). Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test. *Accident Analysis and Prevention*, *39*, 1177–1185. doi:10.1016/j.aap.2007.03.003

Weigand, E. (1999). Misunderstanding: The standard case. *Journal of Pragmatics*, *31*, 763–785. doi:10.1016/S0378-2166(98)00068-X

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors*, *45*, 104–114. doi:10.1518/hfes.45.1.104.27233

Wellens, A. R. (1993). Group situation awareness and distributed decision making: From military to civilian applications. In J. Castellan (Ed.), *Individual and group decision making: Current issues* (pp. 267–291). Hillsdae, NJ: Lawrence Erlbaum Associates.

Wever, R., Van Es, G., & Verbeek, M. (2006). *Air-Ground communication safety study: Causes and recommendations* (Report No. DAP/SAF 2006-09). The Neverlands: National Aerospace Laboratory.

Wickens, C. D. (2007). Aviation. In F. T. Durson, R. S. Nickerson, S. T. Dumais, S. Lewandowsky, & T. J. Perfect (Eds.), *Handbook of applied cognition* (2nd ed., pp. 361–389). Hoboken, NJ: John Wiley & Sons, Ltd.

Wickens, C. D., & Flach, J. M. (1988). Information processing. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in Aviation* (pp. 111–155). San Diego, CA: Academic Press.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall.

Wiggins, M. W. (2006). Cue-based processing and human performance. In W. Karwowski (Ed.), *Encyclopedia of Ergonomics and Human Factors* (2nd ed., pp. 641–645). London, UK: Taylor and Francis.

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science*, 1–10. doi:10.1080/1463922X.2012.724725

Wiggins, M. W. (2014). Differences in situation assessments and prospective diagnoses of simulated weather radar returns amongst experienced pilots. *International Journal of Industrial Ergonomics*, *44*, 18–23. doi:10.1016/j.ergon.2013.08.006

Wiggins, M. W., & Bollwerk, S. (2006). Heuristic-based information acqusition and decision making among pilots. *Human Factors*, *48*, 734–747. doi:10.1518/001872006779166325

Wiggins, M. W., & Henley, I. (1997). A computer-based analysis of expert and novice flight instructor preflight decision making. *The International Journal of Aviation Psychology*, *7*, 365–379. doi:10.1207/s15327108ijap0704_8

Wiggins, M. W., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied*, *1*, 305–320. doi:10.1037/1076-898X.1.4.305

Wiggins, M. W., & O'Hare, D. (2003a). Expert and novice pilot perceptions of static in-flight images of weather. *The International Journal of Aviation Psychology*, *13*, 173–187. doi:10.1207/S15327108IJAP1302_05

Wiggins, M. W., & O'Hare, D. (2003b). Weatherwise: Evaluation of a cue-based training approach for the recognition of deteriorating weather conditions during flight. *Human Factors*, *45*, 337–345. doi:10.1518/hfes.45.2.337.27246

Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science*, *65*, 118–124. doi:10.1016/j.ssci.2014.01.006

Wiggins, M. W., Brouwers, S., Davies, J., & Loveday, T. (2014). Trait-based cue utilization and initial skill acquisition: Implications for models of the progression to expertise. *Frontiers in Psychology*, *5*, 1–8. doi:10.3389/fpsyg.2014.00541

Wiggins, M. W., Harris, J., Loveday, T., & O'Hare, D. (2010). EXPERTise (Software Package). Sydney, Australia: Macquarie University.

Wigton, R. S. (1988). Use of linear models to analyze physicians' decisions. *Medical Decision Making*, *8*, 241–252. doi:10.1177/0272989X8800800404

Wigton, R. S., Hoellerich, V. L., & Patil, K. D. (1986). How physicians use clinical information in diagnosing pulmonary embolism: An application of conjoint analysis. *Medical Decision Making*, *6*, 2–11. doi:10.1177/0272989X8600600102

Wigton, R. S., Patil, K. D., & Hoellerich, V. L. (1986). The effect of feedback in learning clinical diagnosis. *Journal of Medical Education*, *61*, 816–822.

Williams, A. M., Hodges, N. J., North, J. S., & Barton, G. (2006). Perceiving patterns of play in dynamic sport tasks: Investigating the essential information underlying skilled performance. *Perception*, *35*, 317–332. doi:10.1068/p5310

Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied*, *8*, 259–270. doi:10.1037/1076-898X.8.4.259

Woods, D. D. (1988). Coping with complexity: The psychology of human behavior in complex systems. In L. P. Goodstein, H. B. Andersen, & S. E. Olsen (Eds.), *Tasks, errors, and mental models* (pp. 128–148). New York, NY: Taylor and Francis.

Young, J. S., Smith, R. L., Guerlain, S., & Nolley, B. (2007). How residents think and make medical decisions: Implications for education and patient safety. *The American Surgeon*, *73*, 548–554.

Zaefferer, D. (1977). Understanding misunderstanding: A proposal for an explanation of reading choices. *Journal of Pragmatics*, *1*, 329–346. doi:10.1016/0378-2166(77)90027-3

Zuckerman, M., Amidon, M. D., Bishop, S. E., & Pomerantz, S. D. (1982). Face and tone of voice in the communication of deception. *Journal of Personality and Social Psychology*, *43*, 347–357. doi:10.1037//0022-3514.43.2.347

Appendix A: Journal of Communication (2013) Publication

(N.B. This publication has been removed for copyright purposes)

Appendix B: Human Factors and Ergonomics Society Proceedings Publication (2012)

(N.B. This publication has been removed for copyright purposes)

Appendix C: Numeric Digit Span Recall Task and Manipulation Check Description

Numeric Digit Span Recall Task and Manipulation Check

The Numeric Digit Span Recall Task constituted a concurrent task that was completed as part of the Auditory Readback Task [Revised1] in the experiments described in Papers 3 and 4. This concurrent task was designed to address the outcomes of Paper 2 that established a null relationship between visual cue utilisation performance and prosodic cue use during assessments of nonunderstanding. It was hypothesised that the Auditory Readback Task, in the absence of a concurrent task, may have been unable to discriminate those participants who had pre-established prosodic cue associations in Long-Term Memory (LTM) and those who did not possess such associations.

From a theoretical perspective, cues are expected to represent relatively automatic associations within LTM between a feature (such as rising intonation) and an object/event (such as a nonunderstanding; Wiggins, 2012). Such associations can presumably be retrieved from memory, even when cognitive resources are limited (Sweller, 1988). Considering that the Auditory Readback Task described in Paper 2 was not cognitively demanding, it was possible that participants, who did not possess pre-established cue associations in memory, were able to consider at a conscious, analytical level, the significance of the prosodic features in relation to nonunderstandings with only limited reference to cue-based associations in memory.

To address this potential explanation of the null relationship established in the experiment reported in Paper 2, Papers 3 and 4 incorporated a concurrent task. Referred to as the Numeric Digit Span Recall Task, it was designed to impose a cognitive demand on participants, thereby restricting the cognitive resources available during assessments of nonunderstanding. The Numeric Digit Span Recall Task presented participants with a string of eight, single numeric digits that they were asked to remember (and subsequently report) while listening simultaneously to a communicative exchange (see Figure 1).

[1] To distinguish between the structure of the Auditory Readback Task described in Paper 2, and the structure of the Auditory Readback Task described in Papers 3 and 4 (which incorporated the concurrent Numeric Digit Span Recall Task), the term Auditory Readback Task[Revised] will be used to refer to the latter. The term Numeric Digit Span Recall Task will also be used where necessary to refer to that particular task within the Auditory Readback Task[Revised].

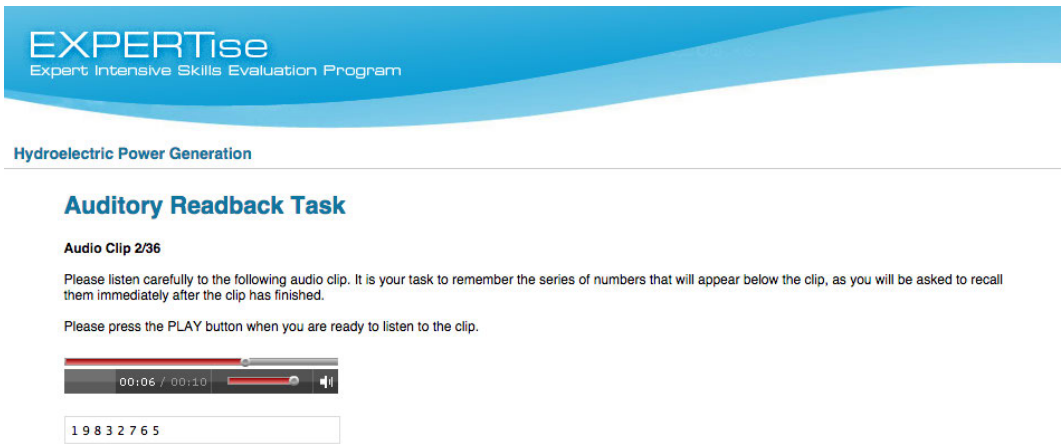


Figure 1. The visual presentation of the Auditory Readback Task[Revised] (incorporating the Numeric Digit Span Recall Task).

Each series of numeric digits included numerals from one to nine in a randomised order for each exchange, with no repetition of digits. Following each communicative exchange, the series of digits was removed, and participants were asked to enter the numeric digits into an open response box (see Figure 2). Participants were then asked to rate the extent to which they perceived the second speaker (i.e., the receiver) had understood the instruction (see Figure 3).

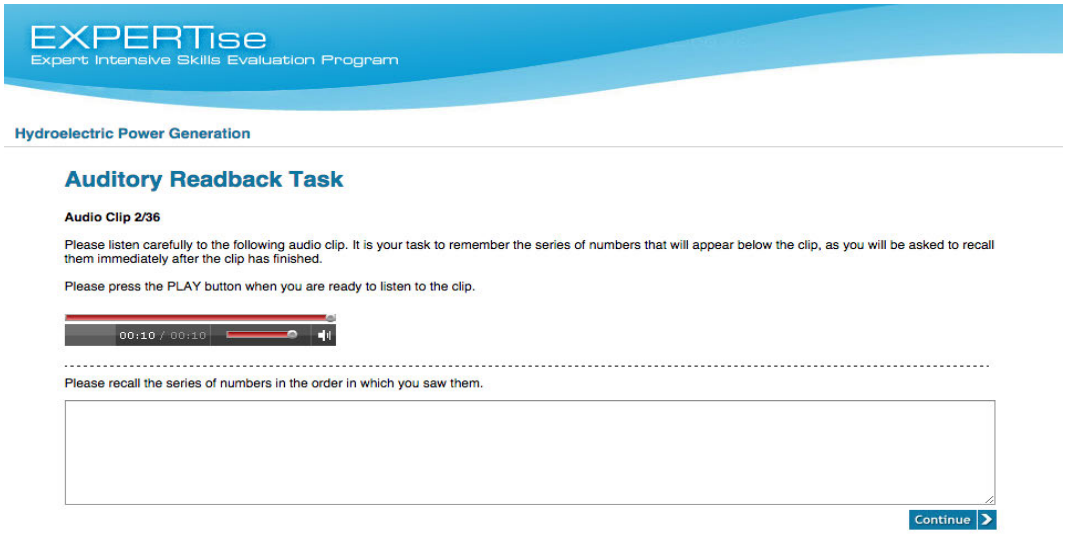


Figure 2. Participants were required to recall the series of eight numeric digits at the completion of the communicative exchange.

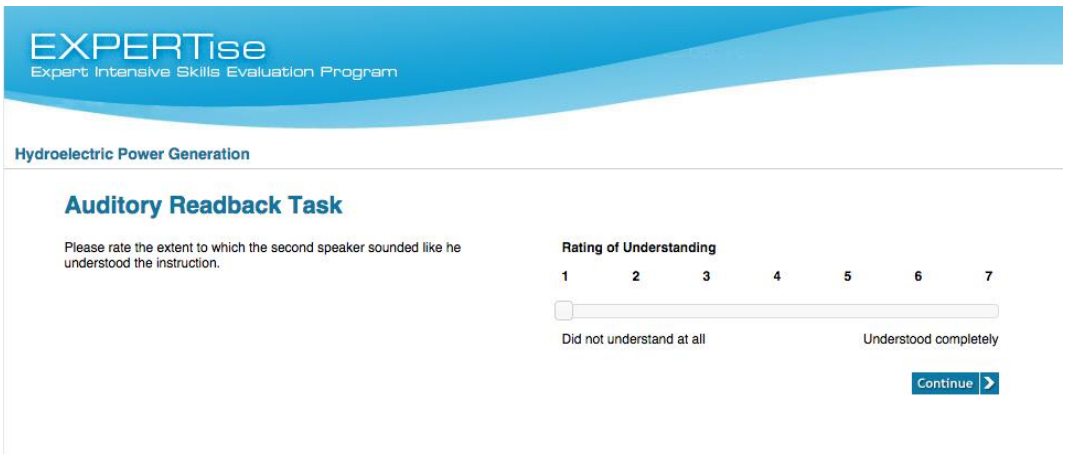


Figure 3. Participants were required to make ratings of nonunderstanding on a subsequent screen.

To ensure that participants would not attend exclusively to memorising the eight digits and thereby fail to attend to the communicative exchanges, they were asked to respond occasionally to a multiple-choice question relating to the semantic content of the communicative exchange, prior to rating the level of nonunderstanding. The communicative exchanges that required a response to a multiple-choice question within each study (described in Papers 2 and 3) were separate from those exchanges that were systematically manipulated to vary the prosodic cues contained within readback responses. These responses were not included in the analyses.

Overall, the number of communicative exchanges requiring a response within the Auditory Readback Task^{Revised} differed between the experiments described in Papers 3 and 4 due to the differences in the research questions. The participants described in Paper 3 were asked to listen and respond to 30 communicative exchanges presented in randomised order, of which two were practice items, 24 were communicative exchanges that varied systematically the prosodic cues contained within the readback responses, and four were communicative exchanges requiring a response to a multiple-choice question (see Table 1).

Table 1

Questions and Multiple-Choice Responses for the Auditory Readback Task^{Revised} in Paper 3

Questions	Multiple Choice Responses
What did the second speaker agree to do to the overhead line equipment?	*A)* Isolate and Earth *B)* Test and Earth (*correct*) *C)* Manually cut the power *D)* Evacuate the surrounding area
What was the second speaker asked to test and earth?	*A)* It wasn't clear *B)* Part 4 of the Form B *C)* The overheads on section 102/3 *D)* The overhead line equipment (*correct*)
The second speaker was authorized to test and earth the overhead line equipment within the limits of isolation stated where?	*A)* On the Form B (*correct*) *B)* On the Form C *C)* On the Form D *D)* Within section 102/3
What was the name of the person giving the instruction?	*A)* Mark *B)* Garry *C)* Greg *D)* It wasn't stated (*correct*)

Note. These questions were based on the following communicative exchange. *Instruction:* You may now test and apply earths at or within the limits of isolation stated within Part One B of the Form B.
Readback: I understand that I am authorized to test and earth the overhead line equipment at or within the limits of isolation stated on the Form B.

The participants described in Paper 4 were asked to listen and respond to 38 communicative exchanges presented in randomised order, of which two were practice items, 32 were communicative exchanges that varied systematically the prosodic and semantic cues contained within the readback responses, and four were communicative exchanges requiring a response to a multiple-choice question (see Table 2).

Table 2

Questions and Multiple Choice Responses for the Auditory Readback Task^{Revised} in Paper 4

Questions	Multiple Choice Responses
What did Steve agree to do to Unit 2?	*A)* Shut it down (*correct*) *B)* Reset it manually *C)* Turn it off *D)* Nothing
When did Steve agree to carry out the directive?	*A)* In ten minutes *B)* Tomorrow *C)* Right away (*correct*) *D)* After lunch
What salutation did Steve use to respond to Pete?	*A)* Sure Pete (*correct*) *B)* Got that Pete *C)* Thanks Pete *D)* Will do Pete
What was the name of the person giving the instruction?	*A)* Andy *B)* Steve *C)* John *D)* Pete (*correct*)

Note. These questions were based on the following communicative exchange. *Instruction:* Hey Steve, can you just shut down Unit 2? *Readback:* Sure Pete. I'll shut down Unit 2 now.

The Auditory Readback Task^{Revised} described in Papers 3 and 4, required participants to listen to auditory communicative exchanges concurrently with the Numeric Digit Span Recall Task. Considering the limits of working memory capacity, a manipulation check was undertaken to determine the optimal length of the numeric digit string for the Numeric Digit Span Recall Task (which was subsequently determined to be eight digits). The purpose of the manipulation check was to ascertain the optimal length of the numeric digit string that restricted cognitive resources available while listening to the auditory communicative exchanges, but not to such an extent that it prevented participants from attending to the exchanges altogether.

To obtain a valid indication of the difficulty in recalling a string of single digits in the manipulation check that could be extrapolated to participants described in Papers 3 and 4, it was important to maintain a degree of consistency between the structure of tasks involved in the manipulation check and the Auditory Readback Tasks^{Revised} completed during the

experiments. This consistency was achieved insofar as participants were asked to listen to domain-specific readback/hearback exchanges while attending to a series of single digits in both the manipulation check and Auditory Readback Tasks^{Revised}, but with two main distinctions: (1) Participants in the manipulation check were only required to respond to a portion of the domain-specific communicative exchanges described in the experiments in Papers 3 and 4 as the influence of auditory cues on assessments of nonunderstanding was not being measured, and; (2) the dependent variables differed between the manipulation check (measurements of task difficulty) and the Auditory Readback Tasks^{Revised} (assessments of nonunderstanding).

The manipulation check for the numeric digit span length from both domains described in Paper 3 (electrical rail control) and Paper 4 (hydroelectric power generation) were incorporated into a single manipulation check completed by one group of participants. The outcomes of this manipulation check are reported in the following sections.

Method

Participants

Thirteen undergraduate students from an Australian University participated in the manipulation check. All of the participants were native speakers of English and ranged in age from 18 to 55 years (M = 25.31, SD = 11.79). The sample was predominantly male (61.54%). Participants received course credit in exchange for completing the study.

Design

Although the manipulation check for the Numeric Digit Span Recall Task for the communicative stimuli described in Papers 3 and 4 were incorporated into a single manipulation check, the design and results are separated according to the different experiments in the event that case-domain differences were evident.

Paper 3.

The Numeric Digit Span Recall Task manipulation check for the communicative stimuli described in Paper 3 comprised a 4 x 4 (Communicative Exchange [1, 2, 3, 4] x Numeric Digit Span Length [0, 4, 6, 8]) design, resulting in 16 distinct scenarios. 'Communicative Exchange' referred to a selection of the electrical rail control operational exchanges used in Auditory Readback Task[Revised] described in Paper 3. 'Numeric Digit Span Length' referred to the number of single numeric digits that a participant was required to retain while listening to each communicative exchange (see Figure 4 for an example of how this was visually presented to participants). This ranged from 0 (i.e., participants were not required to remember any digits while listening to each exchange) to 8 (i.e., participants were required to remember a series of eight single digits as they were listening to each exchange).

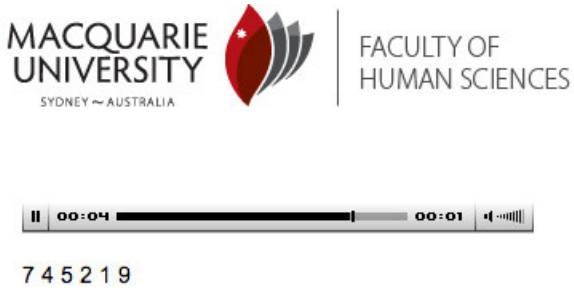


Figure 4. Screen capture of how the numeric digit span and auditory communicative exchange were concurrently presented to participants on screen during the manipulation check. This example displays a numeric digit span length of six.

Paper 4.

The Numeric Digit Span Recall Task manipulation check for the communicative stimuli described in Paper 4 comprised a 3 x 2 x 4 (Communicative Exchange [1, 2, 3] x Instruction Response Type [Readback, Partial Readback) x Numeric Digit Span Length [0, 4, 6, 8]) design, resulting in 24 distinct scenarios. 'Communicative Exchange' and 'Instruction Response Type' were identical to those described in the Auditory Readback Task within Paper 4. However, participants in the manipulation check were only asked to respond to a portion of the exchanges. 'Numeric Digit Span Length' referred to the number of single numeric digit that a participant was required to retain in memory, while listening to each

communicative exchange.

Participants in the manipulation check were asked to listen and respond to a total of 41 communicative exchange scenarios (one example exchange, 16 exchanges pertinent to Paper 3 and 24 exchanges to Paper 4). They were presented with the exchanges in randomised order. The dependent variables pertinent to both Papers 3 and 4 included subjective and objective measurements of task difficulty (described below) and accuracy in the multiple-choice question task. Each communicative exchange, in addition to the three dependent variable measures represented a 'scenario'.

Materials

The semantic content and the development of the communicative exchanges used within the manipulation check have been reported in Papers 3 and 4 and will not be reiterated here. *Task Difficulty* was used to assess the cognitive resources expended on the task (i.e., listening to a communicative exchange while remembering a series of numeric digits). Task difficulty was assessed using both a subjective and objective measurement. For the subjective measurement, participants indicated the extent to which they found the task difficult on a 7-point Likert scale from 1 (*very easy*) to 7 (*very difficult*) after each of the communicative exchanges. The use of this type of unidimensional rating scale was based on evidence that suggests individuals are aware of the amount of mental effort expended to complete a particular task (Gopher & Braune, 1984; Paas & Merriënboer, 1994; Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

For the objective measurement of task difficulty, participants were asked to report back the numeric digit span immediately after the auditory exchange was completed. The use of this type of measurement was based on the assumption that more difficult tasks would result in a greater frequency of errors compared to relatively easier tasks, as there are insufficient cognitive resources available to complete the task successfully.

To ensure that the difficulty of the Numeric Digit Span Recall Task manipulation check

would be consistent with the difficulty of the actual Numeric Digit Span Recall Task employed as part of the Auditory Readbacks Tasks[Revised] used in the experiments described in Papers 3 and 4, it was important that the tasks during the manipulation check maintained a consistent level of difficulty. As part of the Auditory Readback Tasks[Revised] described within Papers 3 and 4, participants completed two tasks concurrently; (1) they listened to each communicative exchange such that a rating of nonunderstanding could be made, and; (2) they attended to the string of eight numeric digits during each exchange to be recalled after the exchange was completed.

It was important that the manipulation check also involved a task whereby participants were required to attend to the communicative exchanges. Rather than asking participants to make ratings of nonunderstanding, participants in the manipulation check were asked to respond to a multiple-choice question relating to the semantic content of the communicative exchange (see Table 3).

Table 3

Examples of the Multiple-Choice Questions Used in the Manipulation Check

Example Communicative Exchange	Instruction	Readback	Multiple Choice Question	Multiple Choice Options
Paper 3	You may now test and apply earths at or within the limits of isolation stated within Part 1 (b) of the Form B.	I understand that I am authorized to test and earth the overhead line equipment at or within the limits of isolation stated on the Form B.	What did the second speaker agree to do to the overhead line equipment?	1. Test and earth 2. Isolate and earth 3. Manually cut the power 4. Evacuate the surrounding area
Paper 4	Hi Steve, Can you go to Rangipo Station and check the status of the bypass valves on the two generators and give me a call.	Thanks Pete. I'm on my way to Rangipo right now and will check the status of the bypass valves and give you a call.	After checking the status of the bypass valves, Steve agreed to do what?	1. Nothing 2. Go to lunch 3. Return back to Tekapo A station 4. Call Pete

Procedure

The participants were invited to take part in the study through an online university advertisement. They were directed to the domain address of a survey provider, and completed the study online in their own time. Participants first completed a selection of brief demographic items before listening and responding to the 41 scenarios (see Figure 4 for the visual presentation of the communicative exchange and numeric digit string). Each scenario was completed in the sequence specified in Figure 5.

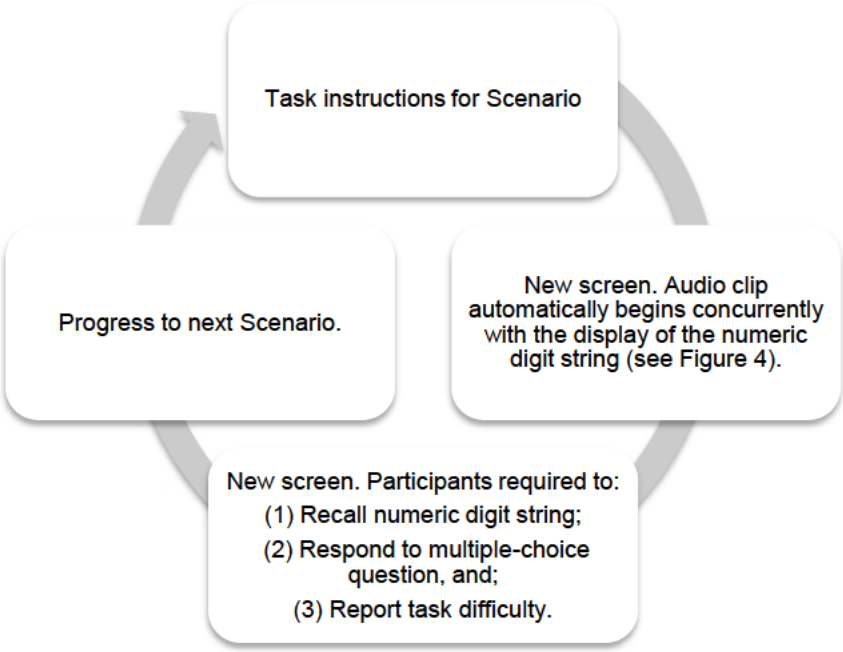


Figure 5. Sequence of tasks to be completed within each scenario.

Data Reduction

The data was reduced in the same manner for the scenarios pertinent to Study 3 and Study 4. For the subjective assessment of task difficulty, participants made ratings on a 7-point Likert scale from 1 (*very easy*) to 7 (*very difficult*) after listening to each exchange and recalling the single digits. The ratings for each of the Numeric Digit Span Length variable levels (0, 4, 6, 8) across the communicative exchanges were then aggregated into an average score of perceived difficulty for each variable level for each participant.

For the objective measure of task difficulty, the proportion of numbers accurately recalled in the correct position of the numeric digit string was recorded for each scenario. For

example, if the numeric digit span for the variable level [6] was [1 4 5 2 7 8] and the participant recalled the following numbers [1 4 9 2 6 8], the participant would have correctly identified four of the six numbers correctly (as nine was incorrectly recalled instead of five and six instead of seven). The four accuracy scores for each of the Numeric Digit Span Length variable levels (0, 4, 6, 8) were then aggregated into a mean proportion of digits recalled correctly for each variable level.

For the multiple-choice question included in each scenario, responses were recorded as either correct or incorrect. A mean accuracy of responses was created for each Numeric Digit Span Length variable level (0, 4, 6, 8). For the manipulation check pertinent to the communicative stimuli in Paper 3 for example, if a participant responded correctly to three of the four questions for the numeric digit span length of eight digits, they received a mean accuracy of 75% for this variable level.

Results

Prior to the separate analyses for scenarios pertinent to the experiments described in Paper 3 and Paper 4, the data were collapsed across all communicative exchanges for each Paper. The data for each Paper were analysed using a one-way repeated measures Analysis of Variance (ANOVA), with four levels of the variable Numeric Digit Span Length (0, 4, 6, 8) for the three dependent variables; (1) subjective measure of task difficulty; (2) objective measure of task difficulty, and; (3) mean accuracy in multiple choice questions.

Results for Scenarios Relevant to Paper 3

For the subjective measure of task difficulty, a statistically significant main effect was evident for Numeric Digit Span Length $F(3, 36) = 16.68$, $p < .001$, $\eta_p^2 = .58$. Post-hoc pairwise comparisons indicated that there was a greater level of perceived difficulty when participants were asked to recall eight digits ($M = 4.27$, $SD = 1.36$) compared to six digits ($M = 3.33$, $SD = 1.22$), $p = .002$, four digits ($M = 3.44$, $SD = 1.11$), $p = .009$, and during instances when no digits needed to be recalled ($M = 2.75$, $SD = 1.01$), $p < .001$. Therefore, participants

rated the task as most difficult when they were asked to recall a numeric digit span containing eight digits.

For the objective measure of task difficulty, a statistically significant main effect was evident for Numeric Digit Span Length, $F(2, 24) = 7.92$, $p = .002$, $\eta_p^2 = .40$. Post-hoc pairwise comparisons using a Bonferroni adjustment indicated that there was a significant reduction in proportion of digits recalled accurately when participants were asked to recall an eight digit string ($M = 79.09$, $SD = 20.43$) compared to a six ($M = 93.27$, $SD = 8.08$), $p = .047$, or four digit string ($M = 98.08$, $SD = 6.93$), $p = .027$.

For the accuracy of the multiple choice questions, no statistically significant main effect was evident for Numeric Digit Span Length, $F(3, 36) = 2.18$, $p = .107$, $\eta_p^2 = .15$. This indicates that participants were as accurate in their responses to the multiple choice questions, irrespective of the number of digits required to recall while listening to the communicative exchanges.

Results for Scenarios Relevant to Paper 4

For the subjective measure of task difficulty, a statistically significant main effect was evident for the Numeric Digit Span Length, $F(3, 36) = 14.33$, $p < .001$, $\eta_p^2 = .54$. Post-hoc pairwise comparisons indicated that there was a greater level of perceived difficulty when participants were asked to recall eight digits ($M = 3.98$, $SD = 1.59$) compared to six digits ($M = 3.31$, $SD = 1.30$), $p = .012$, four digits ($M = 3.15$, $SD = 1.20$), $p = .026$, and when no digits were required to be recalled ($M = 2.31$, $SD = 1.19$), $p < .001$. This indicates that participants rated the task as most difficult when they were asked to recall a numeric digit span containing eight digits.

For the objective measure of task difficulty, a statistically significant main effect was evident for Numeric Digit Span Length, $F(2, 24) = 4.29$, $p = .026$, $\eta_p^2 = .26$. Post-hoc pairwise comparisons revealed a statistically significant reduction in percentage of numbers recalled accurately when participants were asked to recall an eight digit numeric string ($M =$

79.81, *SD* = 19.71) compared to a four digit numeric string (*M* = 93.27, *SD* = 11.26), *p* = .034.[2]

For the accuracy of the multiple choice questions, no statistically significant main effect was evident for the Numeric Digit Span Length, $F(3, 36) = .16$, $p = .922$, $\eta_p^2 = .01$. This suggests that participants were as accurate with their responses to the multiple choice questions, irrespective of the number of digits required to be recalled while listening to the communicative exchanges.

Summary of Results

Overall, the results pertaining to the stimuli developed for the experiments described in Papers 3 and 4 were largely consistent. Across both groups of stimuli, recalling eight digits while listening to the communicative exchanges was significantly more difficult than six digits, four digits or no digits at all (see Figure 6).

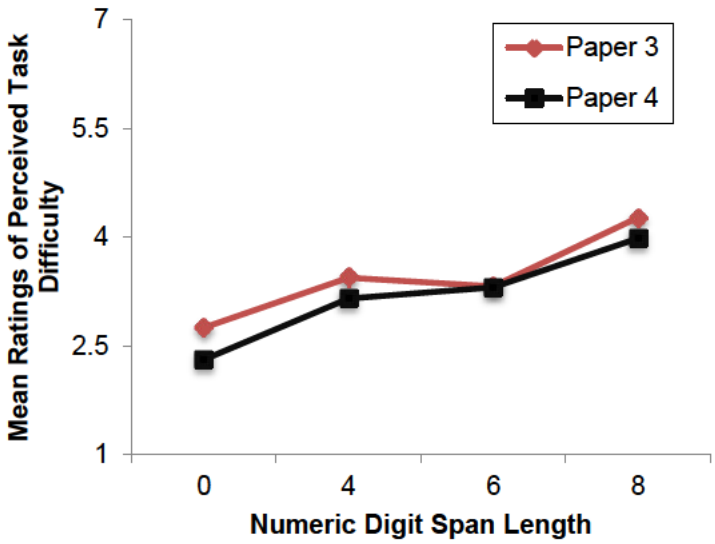


Figure 6. Mean ratings of perceived task difficulty as a function of the numeric digit span length across the stimuli pertinent to Paper 3 and Paper 4. Ratings of task difficulty were made on a scale from 1 (*very easy*) to 7 (*very difficult*).

Similar results were also obtained in relation to the objective measure of task difficulty. Specifically, participants were less accurate in recalling the numeric digit string when it contained eight, rather than six or four numeric digits (see Figure 7). Finally, across all

[2] The post-hoc pairwise comparison between an eight digit span (*M* = 79.81, *SD* = 19.71) and six digit span (*M* = 91.67, *SD* = 8.33) on mean proportion of numbers recalled accurately was not statistically significant, *p* = .072.

stimuli, accuracy in responding to the multiple-choice responses was not a function of the length of numeric digit span.

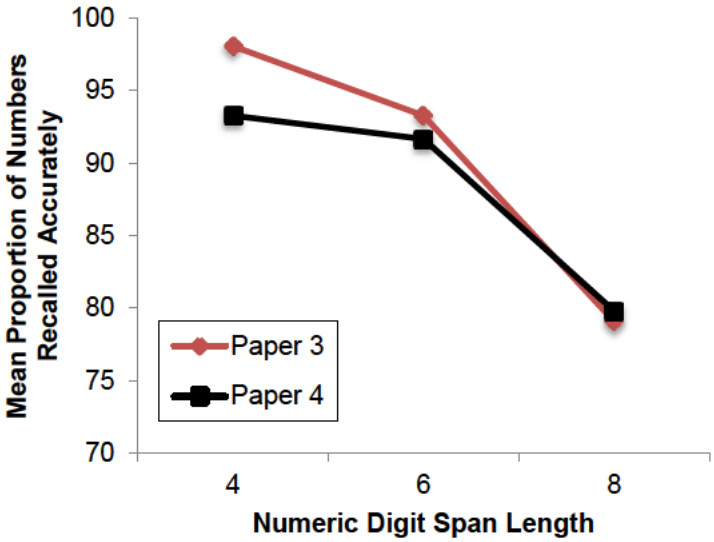


Figure 7. Mean proportion of numbers recalled accurately (out of 100) as a function of the numeric digit span length across the stimuli pertinent to Paper 3 and Paper 4.

Discussion

The objective of this manipulation check was to determine the ideal length of the numeric digit span to be used in the Numeric Digit Span Recall Task that would restrict participants' cognitive resources while listening to the auditory communicative exchanges. The results from the manipulation check indicated that, compared to the six and four-digit numeric spans, when participants were asked to recall the eight-digit numeric span while listening the communicative exchanges, the task was perceived to be relatively more difficult. Participants were also significantly less accurate in recalling the eight digit strings, compared to the six and four digit strings. Nevertheless, the number of digits that participants were asked to recall did not influence participants' accuracy in the multiple-choice questions.

In combination, these results suggest that a numeric digit span of eight digits is likely to be most appropriate for the Numeric Digit Span Recall Task. The results demonstrate that the task of recalling a string of eight digits while listening to an exchange would be sufficiently difficult, since it would restrict cognitive resources, but not completely deprive participants of the resources necessary to attend to the exchanges. As the results were consistent across the

stimuli specific to the domains described in Paper 3 and Paper 4, a span of eight numeric digits was included in the Numeric Digit Span Recall Task described in both papers.

Appendix C References

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*, 519–532. doi:10.1177/001872088402600504

Paas, F. G. W. C., & Merriënboer, J. J. G. Van. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*, 122–133. doi:10.1037/0022-0663.86.1.122

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63–71. doi:10.1207/S15326985EP3801_8

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi:10.1016/0364-0213(88)90023-7

Wiggins, M. W. (2012). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science*, 1–10. doi:10.1080/1463922X.2012.724725

Appendix D: Paper 4 EXPERTise 1.0 Stimuli

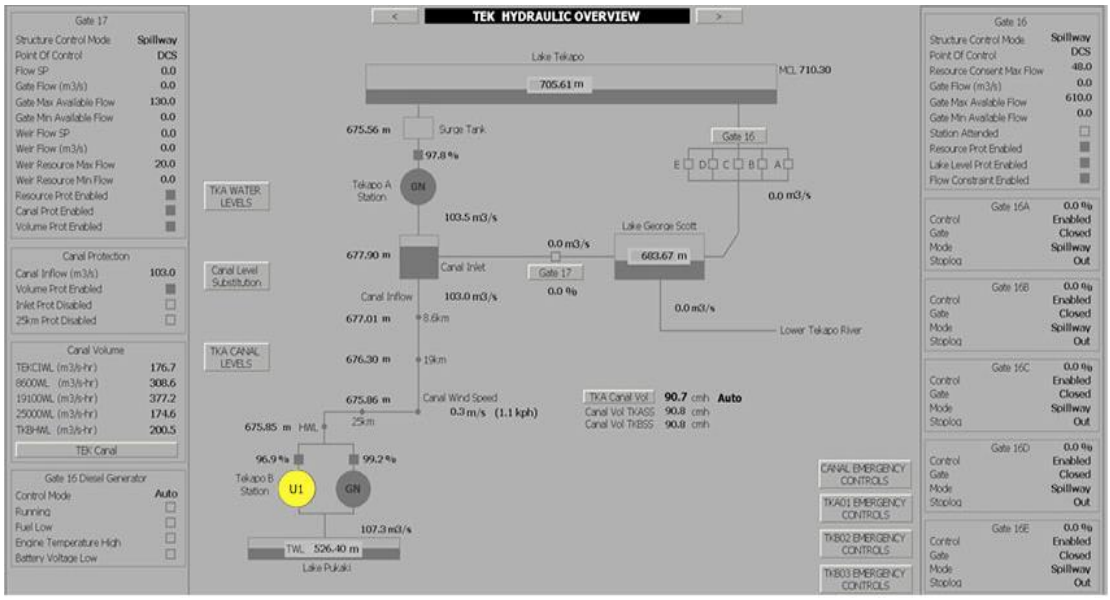
Paired Association Task Items

Sequential Pairs		Simultaneous Pairs	
Feature	Event	Feature	Event
Maximum Load	Set Point Change	Inflow	Outflow
Radial Gates	Derated	Load Change	Operator Alarm
Station Trip	Load Change	Turbine	Dispatcher
Canal	Public Safety	Fire	Column Internal
Operator Alarm	Increase Flow	Canal	Overtopping
Generator	Downstream Flow Low	Operator Alarm	Corrective Action
Pressure Tap	Emergency	Steady Flow	Auxiliary Generator
Gate Change	Increase Flow	Minimal Flow	Compliance Issue
Station Trip	Downstream Flow Low	Flow	Distributor Nozzle
Increase Generation	Maintain Flow	Generator	Turbine
Canal	Spillway	Yellow	Fault
Gate Change	Spillway	Unit Tripped	Lack of Power
Increase Generation	River Bed Level	Overtopping	Yellow
Turbine	Maintain Flow	Generating	Control Valve
Flow Rating	River Bed Level	Fault	Generator
Steady River Flow	Set Point Change	Lack of Power	Inflow
Shear Pin	Routine Surveillance	Quick Intervention	Canal
Generator	Bypass Valves	Corrective Action	Canal Volume
Turbine	Relief Valves	Outflow	Minimal Flow
Operator Alarm	Load Change	Station Tripped	Canal Volume
Maximum Load	Bypass Valves	Station Tripped	Unit Tripped
Transformer	Uncertain Status	Compliance Issue	Lack of Flow
Circuit Breaker	Dispatch Notification	Lack of Flow	Quick Intervention
Vent Fan	No Flow	Abnormal Situation	Penstock Valve
Flow Rating	Relief Valves	Dispatcher	Load Change
Steady River Flow	Public Safety	Maintenance Outage	Sluice Gates

Feature Discrimination Task (1)

Scenario

It is 1400 and you notice a generator trip at Tekapo B station. Please press Continue when you have determined the first step you would take to manage this situation.



Response Options

From the options available, determine the first step you would take to manage this situation. You may only select one response. If you need to review the scenario again, please click Back. Press Continue when you have made your selection.

- Contact dispatcher to inform of situation status
- Change the canal volume
- Manually alter other generator at Tekapo B to full load
- Shut the other machines down
- Diagnose cause of generator trip
- Contact maintenance operator to complete manual reset onsite
- Do nothing

Ratings of Feature Utility

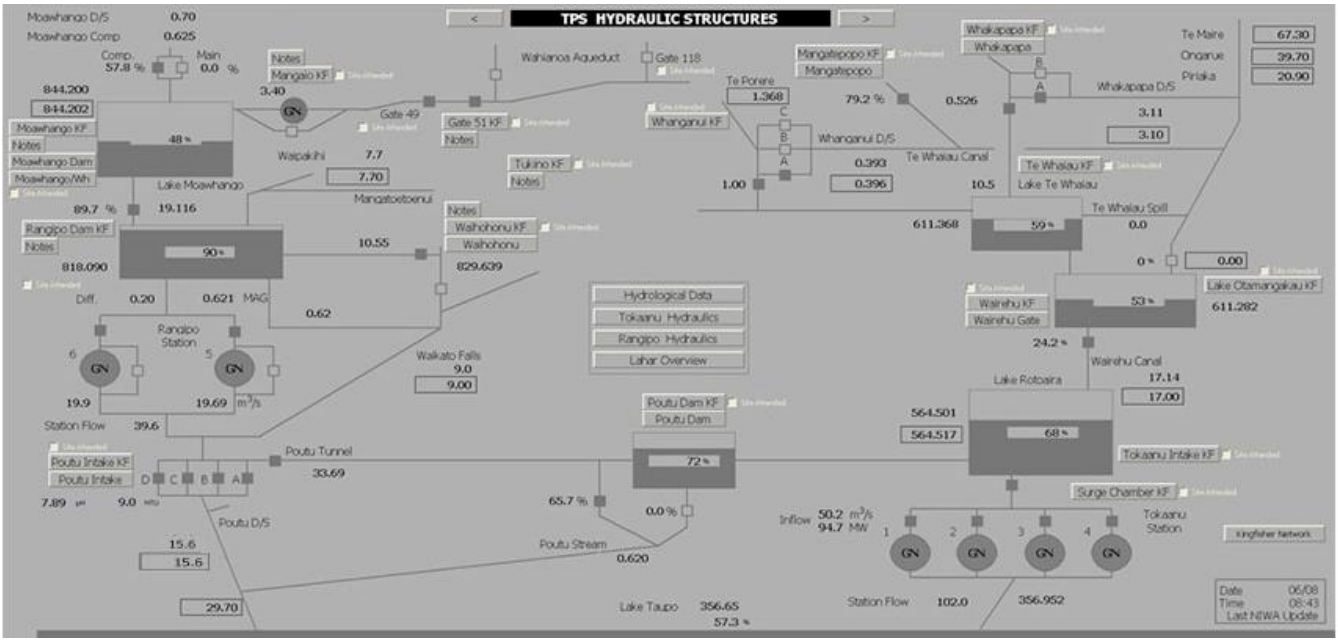
Rate the relevance of the following features in informing your response.

- Time of day
- Canal inflow
- Canal outflow
- Canal wind speed
- Generator megawatts
- Tekapo A flow
- Water level at Lake Tekapo
- Estimated time before canal overtops
- Canal volume

Feature Discrimination Task (2)

Scenario

It is 0410 on a Saturday morning. There has been a recent load change at Rangipo power station at 0400. Although not alerted by any alarm on the hydraulic overview screen, you notice that the flow at Tongariro River downstream Poutu (Poutu D/S) is lower than the accepted minimum (16 m3/s). Please press Continue when you have determined the first step you would take to manage this situation.



Response Options

From the options available, determine the first step you would take to manage this situation. You may only select one response. If you need to review the scenario again, please click Back. Press Continue when you have made your selection.

- Increase generation at Rangipo Station
- Do nothing
- Make gate changes at Poutu intake to increase flow
- Call maintenance operators onsite to fix structures

Ratings of Feature Utility

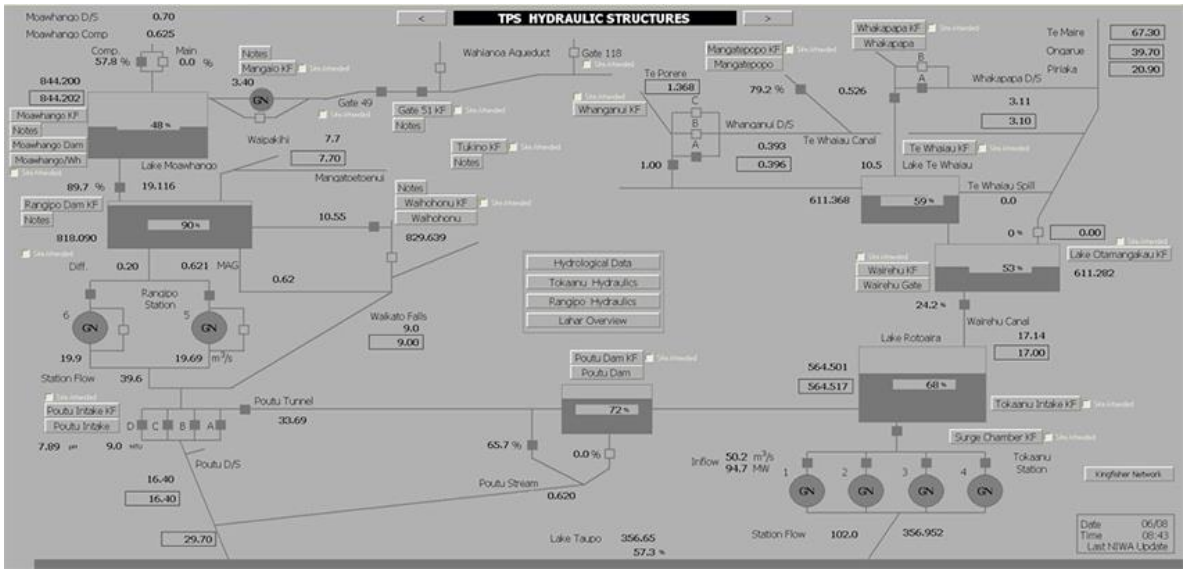
Rate the relevance of the following features in informing your response.

- Time of day
- Recent load changes at Rangipo Station
- Day of week
- Minimum flow accepted at Poutu Intake Downstream
- Rangipo Station down flow
- Recent changes made to flow ratings
- Water level at Lake Maowhango

Transition Task

Scenario

You have just received an alarm indicating a tripping at Rangipo station. There is a scheduled recreational release at Tongariro River below Poutu Intake later today. Below is a hydraulic overview of the Tongariro Power Scheme during normal operation (rather than during the incident described) for your reference. Use the background information in conjunction with the hydraulic overview to determine your response.



Information Categories

- Time of scheduled recreational release
- Poutu Intake gates
- Time of station trip alarm
- Rangipo Station down flow
- Weather
- Minimum flow accepted at Poutu Intake Downstream
- Duration of Poutu Intake Downstream flow in alarm state
- Current flow at Poutu Intake Downstream
- Generator bypass valves at Rangipo station
- Poutu Tunnel flow

- Day of week
- Cause of station trip
- River flow of previous two days at Tongariro downstream of Poutu Intake

Response Options

From the options available, please select the response that you would most likely implement first. You may only select one response. When you have made your selection, press Continue.

- Manually open bypass valves on generators at Rangipo Station
- Nothing - Reassess situation in 15 minutes
- Diagnose cause of station trip
- Open Poutu Intake gates (A & B)
- Trip Poutu Tunnel gate
- Close off Lake Moawhango

Appendix E: Human Factors and Ergonomics Society Proceedings

(Accepted for Publication, 2014)

(N.B. This publication has been removed for copyright purposes)

Appendix F: Cluster Analysis Procedure

Cluster Analysis Procedure and Cue Utilisation

Cluster analysis is a procedure used to cluster together cases based on responses to variables. In the current thesis, the k-means cluster analysis procedure was used to distinguish those operators who demonstrated relatively greater visual cue utilisation performance across the *EXPERTise 1.0* cue-based tasks from those who demonstrated relatively lesser levels of cue utilisation performance. The cluster analysis procedures described in Papers 2, 3 and 4 were conducted using the statistical software package, SPSS, and followed the same general procedure as outlined below.

SPSS K-means Cluster Analysis Procedure

1. Ensure that the data is normative by checking the skewness and kurtosis of each of the EXPERTise 1.0 variables.
2. Convert all EXPERTise 1.0 variables to z-scores, as the k-means cluster analysis procedure requires all variables to be in a standardised scale.
3. Conduct a k-means cluster analysis on the EXPERTise 1.0 standardised variables. Click Analyse > Classify > K-Means Cluster. Move all the standardized EXPERTise measures to the Variable field.
4. In the 'Number of Clusters' field in SPSS, enter the number of expected performance levels. This number will depend on the nature of the sample, but will usually be two or three.
5. Click the 'Save' button, and in the next window, check the boxes 'Cluster Membership' and 'Distance from Cluster Centre' and click 'Continue' to close the window.
6. Click the 'Options' button, and then check 'ANOVA' in the next window that opens. Click 'Continue' to close the window.
7. Click 'OK'.

8. Visually inspect the ANOVA table. Ideally, most of the standardised variables should be statistically significant as this indicates that the clusters are significantly different on that variable and merit further investigation.
9. Next, look at the Final Cluster centres output for a cluster that is superior on most, if not all, of the tasks. The table below indicates whether a positive or negative score on each of the EXPERTise 1.0 variables indicates above or below average performance.

Task	Direction	
	Below Average	Above Average
Feature Identification Task Accuracy	Negative	Positive
Paired Association Task Response Latency	Positive	Negative
Paired Association Task Rating Variance	Negative	Positive
Feature Discrimination Task Ratings Variance	Negative	Positive
Transition Task Ratio of Sequential Pairs	Positive	Negative

10. If the analysis shows that one cluster demonstrates superior performance in some, but not all, of the tasks, running another analysis to get a better fit to the data is worth considering.
11. To get a better fit to the data, you can:
    (a) Try a solution with a different number of clusters.
    (b) Remove participants from the data who are outliers (i.e., those participants who are three or more standard deviations above/below the mean).
    (c) Eliminate a task from the cluster analysis, especially those that are not significant in the ANOVA analysis.

Appendix G: Ethics Approval Letters

Ethics Approval Letter for Studies Described in Papers 1 and 2

From: **Ethics Secretariat <ethics.secretariat@mq.edu.au>** Hide
Subject: RE: HS Ethics Amendment - Wiggins-Loveday Inbox - Gmail
Date: 22 March 2011 12:51:21 PM AEDT
To: A/Prof Mark Wiggins <mark.wiggins@mq.edu.au>
Cc: Jaime Auton

Dear A/Prof Wiggins,

Re: Cue perception and expertise in diagnosis (5201000458D)

Thank you for your recent correspondence regarding the amendment request. The amendments have been reviewed and I am pleased to advise you that amendments have been approved.

This approval applies to the following amendments:

1. Personnel added to the project: Dr Ben Searle and Miss Jaime Auton.
2. Revised Information and consent forms.

Please accept this email as formal notification that the amendments have been approved. Please do not hesitate to contact us in case of any further queries.

All the best with your research.

Kind regards,

Kay Bowes-Tseng

Ethics Approval Letter for Studies Described in Papers 3 and 4

Dear A/Prof Wiggins,

Re: "Auditory Cue Perception and Expertise in Diagnosis" (5201300099)

Thank you for your recent correspondence. Your response has addressed the issues raised by the Faculty of Human Sciences Human Research Ethics Sub-Committee and you may now commence your research.

This research meets the requirements of the National Statement on Ethical Conduct in Human Research (2007). The National Statement is available at the following web site:

http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72.pdf.

The following personnel are authorised to conduct this research:

A/Prof Mark Wiggins
Miss Jaime Claire Auton

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007).

2. Approval will be for a period of five (5) years subject to the provision of annual reports.

Progress Report 1 Due: 19th March 2014
Progress Report 2 Due: 19th March 2015
Progress Report 3 Due: 19th March 2016
Progress Report 4 Due: 19th March 2017
Final Report Due: 19th March 2018

NB. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.

Progress reports and Final Reports are available at the following website:
http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

3. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Sub-Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. All amendments to the project must be reviewed and approved by the Sub-Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

5. Please notify the Sub-Committee immediately in the event of any adverse

effects on participants or of any unforeseen events that affect the continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at the following websites:

http://www.mq.edu.au/policy

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of Final Approval to an external organisation as evidence that you have Final Approval, please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of final ethics approval.

Yours sincerely,

Dr Peter Roger
Chair
Faculty of Human Sciences Ethics Review Sub-Committee
Human Research Ethics Committee

****
Faculty of Human Sciences - Ethics
Research Office
Level 3, Research HUB, Building C5C
Macquarie University
NSW 2109