Investigating a computer-assisted language learning approach for the development of second language word recognition from speech

Joshua Matthews

University of Newcastle, AUSTRALIA
Investigating a computer-assisted language learning approach for the development of second language word recognition from speech

Joshua Matthews BSc, DipEd

A thesis submitted for the degree of PhD (Education)

School of Education

University of Newcastle, AUSTRALIA

14\textsuperscript{th} July 2015
Statement of Originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to the final version of my thesis being made available worldwide when deposited in the University’s Digital Repository, subject to the provisions of the Copyright Act 1968.

Joshua Matthews 14th July 2015
Statement of Collaboration

I hereby certify that the work embodied in this thesis has been done in collaboration with other researchers. I have included as part of the thesis a statement clearly outlining the extent of collaboration, with whom and under what auspices (see appendix 1.1).

Joshua Matthews 14th July 2015
Acknowledgement of Authorship

I hereby certify that the work embodied in this thesis contains published papers and scholarly work of which I am a joint author. I have included as part of the thesis a written statement, endorsed by my supervisor, attesting to my contribution to the joint publications and scholarly work (see appendix 1.1).

Joshua Matthews

14th July 2015
Thesis by Publication

I hereby certify that this thesis is in the form of a series of published papers of which I am a joint author. I have included as part of the thesis a written statement from each co-author, endorsed by the Faculty Assistant Dean (Research Training), attesting to my contribution to the joint publications (see appendix 1.1).

Joshua Matthews

14th July 2015
Permission Regarding Copyright

I have obtained all necessary permission from the copyright owners of my published work such that use of this published work in my thesis is in line with copyright policies of the relevant journals.

Joshua Matthews 14th July 2015
Acknowledgements

The completion of this thesis has been a demanding task. Along the way I have been helped by a number of people to whom I would like to now put forward my formal acknowledgment.

First and foremost, I would like to thank my dear wife Dushani for her ongoing love and support during this long and time consuming task. Without your encouragement I would never have completed this thesis. You were always there to express your belief in my abilities, even when I was in doubt of them myself. Many thanks and much love to you always.

I must also acknowledge the support provided to me by my son, Jaylan. Although when I wrote this thesis you were quite young, you were always there to fill me with happiness, laughter and love. My dear son, never forget that I think that you are simply wonderful.

I acknowledge the great level of assistance and support provided to me by my principal supervisor, Dr John Mitchell O’Toole. Without Dr O’Toole’s guidance and invaluable assistance this research would not have been possible. Mitch, I especially thank you for the insightful and enjoyable discussions during our regular meetings. They will be missed.

I also thank my supervisors, Associate Professor Shen Chen and Dr Rachel Burke who have each played an important role at various phases of my candidature. I sincerely appreciate your guidance and support.

I would also like to acknowledge the contribution that Dr Cheng Junyu, of South East University, Nanjing, has made to the research undertaken as part of this thesis. Dr. Cheng’s readiness to collaborate and assist has been of fundamental importance to the success of this research.

I would also like to acknowledge the assistance provided to me by Mr. Hara Shintaro. Thank you sir for your collaboration and friendship.

I also thank my sister Dr Nicole Matthews for offering her time and expertise in making so many astute recommendations to manuscripts I have written during my candidature. Much love and many thanks Colie.

Last but not least, I would also like to thank the other members of my family: Chris, Bev and Dan. Thanks for looking after Jaylan, sharing kind words, having chats, organising catch-ups and just being there. Much love to you all!
Publications and Manuscripts Arising from this Thesis


Table of Contents

Statement of Originality ........................................................................................................ iii
Statement of Collaboration ...................................................................................................... iv
Acknowledgement of Authorship ........................................................................................... v
Thesis by Publication .............................................................................................................. vi
Permission Regarding Copyright ............................................................................................ vii
Acknowledgements ............................................................................................................... viii
Publications and Manuscripts Arising from this Thesis ........................................................ ix
Table of Contents .................................................................................................................. x
List of Tables ........................................................................................................................ xvi
List of Figures ........................................................................................................................ xix
List of Abbreviations ............................................................................................................ xix
Abstract .................................................................................................................................. xx

Chapter 1: Introduction ............................................................................................................ 1
  1.1 Background and Context ................................................................................................. 1
  1.2 Research Questions ........................................................................................................ 2
  1.3 Thesis Structure ............................................................................................................ 2
    1.3.1 Overview .................................................................................................................. 2
    1.3.2 Research Paper 1 .................................................................................................... 3
    1.3.3 Research Paper 2 .................................................................................................... 3
    1.3.4 Research Paper 3 .................................................................................................... 4
    1.3.5 Research Paper 4 .................................................................................................... 5
  1.4 Ethics .............................................................................................................................. 5

Chapter 2: Literature Review ................................................................................................. 8
  2.1 Overview ....................................................................................................................... 8
  2.2 Word recognition from speech ....................................................................................... 10
    2.2.1 Acquiring the ability to recognise words from speech: emergentist paradigms ........ 12
    2.2.2 Attributes of the speech signal: a source of difficulty for L2 WRS ...................... 14
      2.2.2.1 Rapidity of the speech signal ................................................................. 14
      2.2.2.2 The blended nature of the speech signal ............................................ 16
2.2.2.3 Phonemic variation intrinsic to the speech signal .......................... 17
2.3 Modelling listening............................................................................. 18
  2.3.1 Why model the cognitive processes of L2 listening? ...................... 18
  2.3.2 Models of listening........................................................................ 19
    2.3.2.1 Anderson’s three phase model .............................................. 20
    2.3.2.2 Interactive bottom-up and top-down model ......................... 20
    2.3.2.3 The decoding and meaning building model ......................... 23
  2.3.3 The cognitive basis of problems experienced by L2 listeners ......... 24
  2.3.4 The characteristics of skilled and less skilled L2 listeners .......... 29
    2.3.4.1 The employment of listening strategies .............................. 30
    2.3.4.2 The primacy of L2 WRS .................................................... 35
2.4 Approaches to teaching L2 listening ................................................ 37
  2.4.1 Brief historical overview of teaching L2 listening ....................... 37
  2.4.2 Contemporary recommendations for teaching L2 listening .......... 40
    2.4.2.1 Application of listening strategies ..................................... 40
    2.4.2.2 Developing L2 WRS ......................................................... 41
    2.4.2.3 Synchronous bimodal input to support L2 listening ............. 42
    2.4.2.4 Text reconstruction and comparison to improve L2 WRS ....... 44
2.5 The value of CALL in the development of L2 WRS .......................... 47
  2.5.1 Previous use of CALL to develop L2 WRS ................................ 48
2.6 Conclusions and future directions .................................................. 50

Chapter 3 ................................................................................................. 52
Preamble to research paper 1: “Recognition of high frequency words from speech as a predictor of L2 listening comprehension” ................................................................. 52
  3.1 Introduction ..................................................................................... 54
  3.2 Literature review ........................................................................... 55
    3.2.1 Models of L2 listening comprehension ..................................... 55
    3.2.2 The application of word knowledge in the prediction of L2 listening comprehension ................................................................. 58
    3.2.3 The present study ................................................................. 60
  3.3 Research questions ......................................................................... 62
  3.4 Materials and methods................................................................... 63
6.5.1 Task interaction ................................................................. 138
6.5.1.1 Task interaction ................................................................. 138
6.5.1.2 Task success ................................................................. 139
6.5.1.3 Word learning outcomes .................................................. 139
6.6 Analysis ............................................................................... 139
6.7 Results ............................................................................... 140
6.7.1 Research question one ...................................................... 141
6.7.2 Research question two ...................................................... 143
6.7.3 Research question three .................................................... 145
6.8 Discussions ....................................................................... 147
6.8.1 Implications for the design of CALL for the development of L2 word recognitions from speech ...................................................... 147
6.8.2 Conclusions and future research ........................................ 151
Summary .................................................................................. 154

Chapter 7: Discussions and recommendations for future research ............ 155
7.1 Overview ............................................................................. 155
7.2 Summary of findings and implications for practice and future research .... 155
  7.2.1 The importance of the construct of WRS in L2 listening comprehension ... 155
    7.2.1.1 Key findings .................................................................. 155
    7.2.1.2 Implications for practice ................................................ 157
    7.2.1.3 Limitations and implications for future research ............... 158
  7.2.2 The utility of computer assisted language learning (CALL) for the development of L2 WRS ................................................................. 160
    7.2.2.1 Key findings .................................................................. 160
    7.2.2.2 Implications for practice ................................................ 161
    7.2.2.3 Limitations and implications for future research ............... 162
  7.2.3 The contribution to the body of knowledge which can be used to inform future computer-mediated approaches to develop L2 WRS ................................. 164
    7.2.3.1 Key findings .................................................................. 164
    7.2.3.2 Implications for practice ................................................ 164
    7.2.3.3 Limitations and implications for future research ............... 165
7.3 Concluding remarks .................................................................. 166
List of tables

Table 1.1 Relationship between the primary research questions, chapters, research paper titles, and individual research questions for each research paper ........................................6-7
Table 3.1: Overview of lexical content of listening comprehension test ......................67
Table 3.2: Descriptive statistics and internal consistencies of tests..............................67
Table 3.3: Descriptive statistics for the subsections of the word recognition test ..........68
Table 3.4: Correlations between listening comprehension tests scores and WRS test scores for the three levels of word frequency and total WRS scores.................................68
Table 3.5: Step-wise multiple regression including 3K and 1K WRS scores (2K WRS scores excluded) ..................................................................................................................70
Table 3.6: Step-wise multiple regression including 2K and 1K WRS scores (3K WRS scores excluded manually) ..................................................................................................................70
Table 3.7: 1K, 2K and 3K WRS scores for modest, competent and good listeners ......71
Table 4.1: General description of the five monologues used in the learning protocol ...86
Table 4.2: Description of attributes of dictation tests .......................................................90
Table 4.3: Summary of three constructs of engagement for each monologue..............94
Table 4.4: Summary of results for dictation pre-test .......................................................96
Table 4.5: Summary of engagement by low and high groups across all monologues ....96
Table 4.6: Summary of attitudinal survey results ..............................................................97
Table 5.1: Summary of internal consistency measures for pre-test and post-tests ......113
Table 5.2: Mean WRS gain scores for participants who used the application and those who did not...................................................................................................................115
Table 5.3: Mean WRS gain scores for treatment one, treatment two and control group .................................................................................................................................116
Table 5.4: Mean WRS gain scores for low, moderate and high exposure to input categories......................................................................................................................117
Table 5.5: Mean WRS gain scores for low, moderate and high modified output categories......................................................................................................................118
Table 6.1: An overview of monologue word length, speech rate and lexical frequency .................................................................................................................................134
Table 6.2: Pre-test mean L2 WRS scores for the total, high frequency and academic words. .................................................................................................................................140
Table 6.3: Mean scores for total, high frequency and academic words for participants from low, moderate and high L2 WRS proficiency levels.............................................. 141
Table 6.4: Comparison of two task interaction variables for three L2 WRS proficiency levels ......................................................................................................................... 142
Table 6.5: Comparison of two variables of task success for three L2 WRS proficiency levels .................................................................................................................... 144
Table 6.6: Comparison of total, high frequency and academic L2 WRS gain scores for three WRS proficiency levels........................................................................................................... 146
List of figures

Figure 2.1: Schematic diagram of literature review ................................................................. 9
Figure 2.2: Waveform of the sentence: “India is the second most populated country in the world” ......................................................................................................................... 16
Figure 4.1: Screen shot of phase two of the learning protocol ............................................. 88
Figure 4.2: Screen shot of phase three of the learning protocol .......................................... 88
Figure 5.1 Opportunities to increase exposure to input and to produce modified output .......................................................................................................................... 108
Figure 5.2 Opportunities to increase exposure to input and to produce modified output .......................................................................................................................... 109
Figure 5.3 Example of feedback mode of treatment level 1 participants (text feedback provided) ......................................................................................................................... 110
Figure 5.4 Example of feedback mode of treatment level 2 participants (no text feedback provided) ......................................................................................................................... 110

List of Abbreviations

ANOVA Analysis of variance
CALL Computer-assisted language learning
ESL English as a second Language
HSD Honest significant difference
L1 First language
L2 Second language
LSD Least significant difference
M Mean
N Number of participants
p Probability (statistical significance level)
r Pearson product-moment correlation coefficient
SD Standard deviation
SLA Second language acquisition
TRC Text reconstruction and comparison
WRS Word recognition from speech
Abstract

Second language (L2) listening is an essential component of L2 proficiency. The most critical aspect of L2 listening is the ability to recognise words from speech. Despite L2 word recognition from speech (WRS) being fundamentally important to successful L2 listening, it has been systematically underemphasised in L2 language learning research. As a consequence of this lack of research emphasis, very little empirical data exists regarding how L2 WRS can be developed in real language learning contexts. This thesis presents four contextualised research papers which aim to begin filling this gap in the existing literature. These papers in their entirety aim to address three primary research objectives:

1) To quantify the importance of the construct of WRS in L2 listening comprehension.
2) To examine the utility of computer assisted language learning (CALL) for the development of L2 WRS
3) To contribute to the body of knowledge which can be used to inform future computer-mediated approaches to develop L2 WRS.

Firstly, findings show L2 WRS is strongly associated with second language listening comprehension. The link between word recognition skill and L2 listening performance provides a strong rationale for research which investigates approaches to develop L2 WRS. Secondly, findings show that CALL does have utility in the development of L2 WRS in real language learning contexts. Two separate investigations which show the effectiveness of a web-based CALL application specifically designed to assist L2 learners to develop improved L2 WRS are presented. The CALL application which is described not only provides a model for CALL for the development of L2 WRS, but also provides a design framework from which future improved iterations can evolve. Lastly, in line with the research objective to improve future CALL approaches to develop L2 WRS, empirical data detailing learner use of the application developed and investigated as part of this thesis are analysed and presented. These empirical data are used as a framework for recommendations for future iterations of the L2 WRS application which may more adequately cater to learners’ individual differences.
Chapter 1: Introduction

1.1 Background and Context

Second language (L2) listening ability, or the ability to process and comprehend information from L2 speech, is central to the development of foreign language acquisition and plays a key role in the acquisition of language skills (Dunkel, 1991; Feyton, 1991; Rost, 2002; Vandergrift, 1999). However, L2 listening is also arguably the most challenging of the four macro-skills. L2 listening is the language skill for which students rate themselves as being least proficient, find the most difficult (Graham 2006) and associate with significant levels of anxiety (Arnold 2000; Elkhafaifi 2005). Listening is also difficult to teach in a manner that provides all the input and individualised support needed for significant improvement (Rost 2002; Vandergrift 2007; Field 2008b). Despite the importance of L2 listening and its known difficulty for teachers and learners, L2 listening has received less research attention than the other three traditional macro-skills (Vandergrift, 2007). This context motivates the need to more fully understand the difficulties associated with L2 listening and to devise approaches to assist language learners to overcome these difficulties.

Of the various skills which support skilled L2 listening, word recognition from speech (WRS) is the most critical (Rost, 2002; Hulstijn, 2003). WRS is defined here as the ability to map information from the speech signal onto the lexical units that information represents. Adequate word recognition enables listeners to make rapid links between the formal and semantic representations of words (Hulstijn, 2002). This capability enables a listener to access the linguistic and semantic information encoded in speech, and is therefore of fundamental importance in facilitating listening comprehension. Due to the transient nature of spoken language, word recognition must occur rapidly, and therefore imposes a significant cognitive load on non-expert listeners. Inadequate automaticity in word recognition is a major impediment to L2 listening success (Goh, 2000). Although the ability to recognise words from speech is a critical aspect of L2 listening comprehension, this specific dimension of L2 listening has received very limited research attention (Broersma & Cutler, 2008). This lack of research emphasis has resulted in a paucity of empirical data which can be used to inform the development of L2 WRS in real language learning contexts.
Although a relatively limited amount of research attention has been directed toward the development of L2 word recognition skills, previous research in this area suggests that L2 WRS may be developed through the provision of structured and extensive opportunities to engage with appropriate speech input (Field, 2003, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003). CALL has the capability to provide learners with extensive and scaffolded exposure to target spoken language in a manner concordant with existing pedagogical recommendations on how best to develop L2 WRS (Hulstijn, 2003; Vandergrift, 2007). Despite the potential CALL may hold, few substantive contributions exist in the L2 listening literature on how to use computers to improve L2 WRS in real language learning contexts.

The primacy of L2 WRS in skilled L2 listening, the difficulty experienced by L2 listeners due to an inadequate capability to recognise words from speech and the strong potential computers have in the development of this challenging L2 skill, combine to provide the overarching motivation for the current research.

1.2 Research questions

Three primary research questions are addressed by this thesis.

1) Does second language word recognition from speech (L2 WRS) have sufficient influence on second language listening comprehension to make research into the development of L2 WRS a productive enterprise?

2) Does computer-assisted language learning (CALL) provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS a useful way forward?

3) Which design features might be most effective in improving the utility of CALL for the development of L2 WRS?

1.3 Thesis Structure

1.3.1 Overview

The thesis begins with a review of literature (Chapter 2) which contextualises the research undertaken as part of this thesis. Next, the four research papers which make up the body of the research undertaken as part of this thesis are each presented in
sequential, individual chapters (Chapters 3 to 6). A brief overview of each research paper presented as part of the study is provided below. Table 1.1 relates each research question to its corresponding thesis chapter, research paper, and also presents the individual research questions addressed by each research paper. Key findings from each of the research papers are then used to address each of the three primary research questions in Chapter 7. Implications for practice, limitations of the current research as well as implications for future research are also addressed in Chapter 7.

1.3.2 Research paper 1

The first primary research question is addressed in research paper 1, which is entitled “Recognition of high frequency words from speech as a predictor of L2 listening comprehension”. This research paper is presented in Chapter 3 and provides an empirical basis of the importance of L2 WRS in L2 listening comprehension success. To do so, tests measuring L2 WRS and global L2 listening comprehension were administered to a cohort of Chinese tertiary level students (N = 167) studying English as a second language. Results indicated that L2 WRS scores were strongly correlated with, and highly predictive of, L2 listening comprehension scores. A detailed account of the context, research questions, methods, results, and conclusions of this research paper are presented in Chapter 3. The strong link observed between L2 WRS and L2 listening comprehension highlights the importance of emphasising L2 WRS in teaching and learning aimed at improving L2 listening comprehension. Specifically, these results provide an explicit rationale for investigating learning approaches which effectively develop the capacity of L2 learners to recognise L2 words from speech. Thus, research paper 1 provides an empirical rationale for research paper 2 (Chapter 4) and research paper 3 (Chapter 5), each of which address primary research question 2.

1.3.3 Research paper 2

The second primary research question is addressed in the research presented in the second and third research papers. The second research paper entitled, “Investigating an innovative computer application to improve L2 word recognition from speech” is presented in Chapter 4. This research paper involved the development and implementation of a computer application designed to improve L2 WRS among a cohort of international students (N = 33) studying English as a second language in a
Thai university. A detailed description of this computer application is provided in section 4.3.3 of Chapter 4.

An exploratory, nine week, one group research design was adopted. The study gathered a range of data in order to establish: the L2 WRS improvement experienced by participants who used the application, the manner by which those participants engaged with the application and the application’s general functionality. Results from a battery of word recognition tests suggested the utility of the application in developing L2 WRS.

A detailed account of the context, research questions, methods, results, and conclusions of this research paper are presented in Chapter 4. The positive learning outcomes and functionality of the application noted in research paper 2 establishes the context for a more rigorous investigation of the CALL application among a larger cohort of learners with a quasi-experimental research design, which was the nature of the research described in research paper 3.

1.3.4 Research paper 3

The second primary research question is also addressed in the third research paper entitled, “Computer-mediated input, output and feedback in the development of L2 word recognition from speech”. Research paper 3, which is presented in Chapter 5, involved the implementation of the same computer application described in research paper 2 but this time among a cohort of Chinese tertiary level ESL students ($N = 96$). A quasi-experimental, pre-test/treatment/post-test research design involving three groups was applied. The control group ($n = 31$) did not use the application, whereas the members of the two treatment groups ($n = 30$ and $n = 35$) used the application over a period of five weeks. Treatment group one and treatment group two each received different levels of feedback while using the computer application. The study sought to evaluate the word recognition improvements associated with use of the application. Further, an analysis of participant interaction and feedback mode enabled conclusions to be drawn in regards to the relative value of the design features of the application which were modelled on the key second language acquisition (SLA) constructs of input, output and feedback. A detailed account of the context, research questions, methods, results, and conclusions of this research paper are presented in Chapter 5. Results from this research paper provide empirical support for the efficacy of the CALL approach in the development of L2 WRS. These results validate the use of this application as a prototype for future
iterations of CALL designed to improve L2 WRS and as such provide a context for the research aims of research paper 4.

1.3.5 Research paper 4

The third primary research question is addressed in the fourth research paper which is presented in Chapter 6. The fourth research paper is entitled “The impact of proficiency level on interaction, task success and word learning: design implications for CALL to develop L2 word recognition from speech”. This research paper involved analysing data taken from a cohort of Chinese tertiary level ESL students (N = 65) who used the CALL application described and applied in research paper 2 and research paper 3. The study sought to interpret this empirical data in order to put forward design recommendations which would inform future iterations of the CALL application. Specifically, this research paper sought to establish design principles which would enable the application to more adequately cater to difference in learner L2 proficiency. The participants were categorised based on their relative baseline WRS proficiency level and their interactions with the application and their word learning outcomes were analysed. Results of this analysis provided an empirical basis from which design recommendations were put forward. A detailed account of the context, research questions, methods, results, and conclusions of this research paper are presented in Chapter 6.

1.4 Ethics

Ethics approval for the research outlined in section 1.3 was obtained from the University of Newcastle Human Research Ethics Committee (H-2010-1335)
Table 1.1 Relationship between the primary research questions, chapters, research paper titles, and individual research questions for each research paper

<table>
<thead>
<tr>
<th>Primary Research questions</th>
<th>Chapter / Research Paper title</th>
<th>Individual research questions for each research paper</th>
</tr>
</thead>
</table>
| Research question 1:      | Chapter 3: Recognition of high frequency words from speech as a predictor of L2 listening comprehension | 1. To what extent does the ability to recognise high frequency words from speech correlate with listening comprehension scores?  
2. To what extent does the ability to recognise words from the first, second and third thousand frequency level contribute to the prediction of L2 listening comprehension scores?  
3. What is the comparative profile of word recognition from speech performance across the first, second and third thousand frequency levels for those participants attaining modest, competent and good listening comprehension scores? |
| Does second language word recognition from speech (L2 WRS) have sufficient influence on second language listening comprehension to make research into the development of L2 WRS a productive enterprise? | | |
| Research question 2:      | Chapter 4: Investigating an innovative computer application to improve L2 word recognition from speech | 1. Will measurable improvements in word recognition be observed among a cohort of learners after using the application?  
2. How will self-determined exposure to input (repetitions), modified output productivity (revisions) and learning protocol word recognition scores (accuracy) vary as learners progress through the computer application?  
3. How will self-determined exposure to input (repetitions), modified output productivity (revisions) and learning protocol word recognition scores (accuracy) vary between learners of relatively low and high word recognition ability? |
| Does computer-assisted language learning (CALL) provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS a useful way forward? | | |
Table 1.1 (cont.) Relationship between the primary research questions, chapters, research paper titles, and individual research questions for each research paper

<table>
<thead>
<tr>
<th>Primary Research questions</th>
<th>Chapter / Research Paper title</th>
<th>Individual research questions for each research paper</th>
</tr>
</thead>
</table>
| Research question 2: Does computer-assisted language learning (CALL) provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS a useful way forward? | **Chapter 5:** Computer-mediated input, output and feedback in the development of L2 word recognition from speech | 1. What is the impact of the computer-mediated approach described here in the development of L2 word recognition from speech?  
2. What is the impact of computer-mediated text feedback on L2 word recognition from speech?  
3. What is the relationship between computer-mediated exposure to input and L2 word recognition from speech?  
4. What is the relationship between computer-mediated production of modified output and L2 word recognition from speech? |
| Research question 3: Which design features might be most effective in improving the utility of CALL for the development of L2 WRS? | **Chapter 6:** The impact of proficiency level on interaction, task success and word learning: design implications for CALL to develop L2 word recognition from speech | 1. What is the impact of learner WRS proficiency level on the manner by which learners interact with a CALL application for the development of L2 WRS?  
2. What is the impact of learner WRS proficiency level on the varying levels of task success experienced by learners during interaction with a CALL application for the development of L2 WRS?  
3. What is the impact of learner WRS proficiency level on the word learning outcomes attained by learners after interaction with a CALL application for the development of L2 WRS? |
Chapter 2: Literature Review

2.1 Overview

The literature review presented in this chapter is divided into four sections. The first section focusses on the construct of word recognition from speech (WRS). Word recognition from speech and the relevant cognitive processes involved in word recognition from speech are described. In the later part of this section, the attributes of the speech signal which make word recognition from speech a challenging task for L2 learners are described.

The second section focusses on models of L2 listening. The general value of modelling the processes involved in listening is highlighted and an overview of three models of listening are presented. This provides a context for an overview of the difficulties experienced by L2 listeners and a description of the attributes of skilled and less skilled L2 listeners. The key objective of this section of the literature review is to provide an overview of the processes involved in L2 listening in order to establish the primacy of L2 word recognition from speech in L2 listening success.

The third section focusses on selected approaches to teaching L2 listening. It begins with a brief historical overview of approaches to teaching listening which leads into a discussion of contemporary recommendations for teaching L2 listening. From here, approaches to the development of L2 listening which provide context to the current study are presented. The latter part of this section addresses the value of using CALL in the development of L2 WRS and provides an outline of the few previous studies which have addressed the use of CALL for this specific learning objective.

The fourth section presents the conclusion of the literature review and identifies the gap in knowledge which motivates the research undertaken as part of this thesis. Figure 2.1 presents an overview of the literature review.
Section 1: L2 word recognition from speech

- Acquiring the ability to recognise word from speech
- Attributes of the speech signal: a source of difficulty for L2 WRS

Section 2: Modelling listening

- Why model listening?
- Models of listening
  - Anderson’s 3 phase model
  - Bottom-up/Top-down model
  - Decoding/meaning building model

Section 3: Approaches to teaching L2 listening

- Historical overview
- Contemporary recommendations
- CALL in the development of L2 WRS

Section 4: Conclusions

- Identification of the gap in knowledge which motivates this research

Figure 2.1: Schematic diagram of literature review
2.2 Word recognition from speech

According to information processing theory, listening comprehension is mediated by rapid, accurate and socially contextualised cognition which draws upon linguistic and non-linguistic knowledge sources (Buck, 2001). The term listening comprehension encompasses the perception of speech sounds and the mental conversion of those sounds into units of abstract meaning. The ability to extract the meaning of spoken language is fundamentally dependant on the ability to recognise the composite words present within any given spoken utterance (Rost, 2002). Words hold special significance in language learning as the lexical level is the lowest level at which linguistic and semantic forms can be stably matched (Hulstijn, 2002). Words therefore represent the lowest level of representation at which the meaning of linguistic units present in the speech input can be unequivocally accessed by the listener. Word recognition is therefore an important mechanism by which the raw linguistic information encoded in the speech signal can be accessed, processed and interpreted. As word recognition is a fundamentally important component of listening comprehension, the cognitive mechanisms which underpin word recognition are of relevance to language teachers.

The assumption made here is that establishing an understanding of the cognitive processes involved in word recognition from speech will enable a clearer understanding of the ways in which spoken word recognition can be developed among L2 learners.

Various models of word recognition have been formulated in an effort to describe the cognitive events which facilitate spoken word recognition in native speakers (Luce & Pisoni, 1998; McCelland & Elman, 1986; Norris & McQueen, 2008). Although these models vary considerably in their fine detail, a synthesis of their unifying themes provides a useful foundation from which to understand the broadly accepted mechanisms of word recognition. Fundamentally, word recognition depends on the acoustic elements of the speech being mapped onto pre-existing lexical representations stored in the listener’s brain. These lexical representations are referred to as lexical entries and the sum total of a language user’s lexical entries are referred to as the mental lexicon. The lexical entries within a native speaker’s mental lexicon are typically associated with various types of information relating to the words each entry represents.

It is the depth and degree of integration of the information relating to the lexical entry for any given word which equates to an individual’s knowledge of that word. Lexical
entries contain information about the semantic, syntactic, morphological, orthographic and phonological aspects of the word (Jiang, 2000; Levelt, 1989).

Of specific importance in regards to word recognition from speech are the phonological representations of words held in the mental lexicon. Phonological representations held in the mental lexicon are the abstract representations which underpin phonological lexical knowledge. Phonological lexical knowledge describes the implicit knowledge of the phonological properties of words (Jones & Witherstone, 2011). The possession of the phonological knowledge of a particular word, will enable a listener to perceive the sequence of phonemes that represents that word, and thus result in activation of the appropriate word held in the mental lexicon. It is language specific, phonological knowledge that underpins the language performance construct of *receptive phonological ability*. Drawing from a broader definition put forward by Trofimovich (2005, page 479), here receptive phonological ability is defined as the “language user’s ability to perceive…the auditory dimension of spoken language like native speakers of that language”.

Word recognition, or the process of mapping speech sounds onto the representation of words held in the mental lexicon, always involves the activation of candidate lexical entries and competition between those lexical entries (Broersma & Cutler, 2008). Activation occurs when perceptual information triggers the excitation of potential lexical entry candidates in the listener’s mental lexicon (McCelland & Elman, 1986). The degree to which lexical entries are activated depends on the degree of congruence between the candidate’s lexical entries and the perceived phonological attributes of the speech signal (Weber & Scharenborg, 2012). The process of recognition is incremental and unfolds in a probabilistic manner as more and more of the speech input is perceived (Rost, 2002). Therefore, on perceiving the syllable /laɪ/, providing they are present in the listener’s mental lexicon, lexical entries for words such as light, lightning, lie, lying, lion and lice are automatically activated. Candidate words from the activated lexical entries compete as more of the target word becomes perceptually available. Therefore as the syllable /laɪ/ is enunciated and perceived as /laɪt/, light and lightning will remain as candidates from the previous list but lexical entries of words such as lie, lying, lion, lice and so forth will be excluded. The process of activation and competition between candidate words enables inappropriate entries to be discarded and the most strongly
activated candidate words to be aligned with the incoming speech signal (Magnuson, Dixon, Tanenhaus, & Aslin, 2007). The mechanisms of activation and competition which drive word recognition from speech are therefore strongly contingent on the breadth and depth of the lexical knowledge and therefore the receptive phonological ability of the listener.

2.2.1 Acquiring the ability to recognise words from speech: emergentist paradigms

As has been briefly described the mechanisms which enable word recognition from speech depend on the alignment of components of the speech signal with pre-existing representations stored in the brain of the listener. As rapid and accurate word recognition is a fundamentally important component of successful listening comprehension, considering the manner by which this capability becomes acquired by a language user is of relevance to language teachers and researchers. An examination of the learning mechanisms which typically enable first language users to rapidly and accurately recognise words from speech will provide a theoretical frame from which pedagogical approaches aimed at developing word recognition of second language (L2) speech can be devised.

An important starting point is a consideration of how the phonological representations which facilitate spoken word recognition become established in the first language (L1) listener’s cognitive apparatus. Experimental evidence strongly suggests that the development of phonological knowledge which enables L1 listeners to rapidly identify words in speech emerges from language usage (Ellis, 2002; Frisch, Large, Zawaydeh, & Pisoni, 2001). Unlike theories of generative grammar, the emergentist view of language learning contends that the domain non-specific cognitive capabilities to perceive and store information in memory are a sufficient mechanism to explain language acquisition (Hulstijn, 2002). An emergentist view of language maintains that language is best described by aspects of cognition that lie beneath the level of grammatical rules (O’Grady, 2008). According to emergentist approaches to language acquisition, implicit knowledge or the underlying linguistic competence which facilitates automatic language processing, is primarily driven by cognitive engagement with linguistic input (Hulstijn, 2002).
The development of the phonological ability typical of native language users begins in the very earliest stages of infancy. Within the first years of life this sensitivity becomes language specific as a consequence of exposure to the native language in the aural modality (Jusczyk, 1997; Werker & Tees, 1999). Such emergentist theories of language acquisition contend that underlying linguistic competence is the result of “piecemeal learning of many thousands of constructions and the frequency-biased abstraction of regularities within them” (Ellis, 2002, page. 144). Typically, adult native listeners have been exposed to extensive examples of the target language input in the aural modality. The resultant units of phonological representation held in the native listener’s mental lexicon are highly ordered and have developed in direct response to the statistical regularities of target language speech (Ellis, 2002; Goldinger, Luce, & Pisoni, 1989).

The resultant implicit knowledge acquired through extensive exposure to the native language is evidenced in language specific sub-lexical and lexical knowledge which aids in spoken word recognition (Vitevitch, 2003). For example, native speakers, even in the early phases of life, are perceptually attuned to the phonotactic probabilities of the native language. Phonotactic probability refers to the relative frequency with which the sequences of phonological segments appear in a spoken language (Jusczyk, Luce, & Charles-Luce, 1994). Such implicit phonological knowledge facilitates the differentiation of the phonemic elements of the target language and lexical segmentation of words in the speech stream (Al-jasser, 2008).

Another important feature of the native listener’s phonological knowledge is the strong degree of integration of the various forms of word knowledge associated with each lexical entry. This high degree of integration of different forms of word knowledge enables a listener to automatically access all other dimensions of word knowledge associated with a particular lexical entry once the phonological representation of that word has been activated (Jiang, 2000). For example, the activation of the phonological information of a lexical entry automatically makes available the corresponding semantic representations for the same lexical entry in what is referred to as an information cascade (Huang & Snedeker, 2011). The ability to automatically recognise words, and the resultant information cascades which link the various forms of word knowledge which are possessed by the listener, are the result of exposure to highly contextualised target language in the aural modality (Jiang, 2000).
2.2.2 Attributes of the speech signal: a source of difficulty for L2 WRS

Thus emergentist paradigms suggest that the phonological knowledge typical of an L1 listener, which facilitates rapid word recognition from speech, is a product of extensive contextualised language usage. As with L1 listeners, the L2 learners’ sensitivity to phonological properties of L2 spoken words also depends on implicit learning facilitated by exposure to L2 linguistic input (Trofimovich, 2008). However, when turning the attention of this discussion to the case of second language (L2) listeners, a consideration of the differences in attributes of the typical learning context of the L2 and L1 listener suggests pause for thought. When compared to native speakers, L2 learners typically have far fewer opportunities to be cognitively engaged with L2 target language speech. As a result an L2 language learner, when compared to an L1 language user, has less robustly integrated word knowledge within the mental lexicon. Specifically, the relatively limited opportunity to engage with contextualised spoken input in the target language results in a paucity of phonological word knowledge in the L2 mental lexicon (Jiang, 2000). Inadequate levels of phonological word knowledge have a measurable impact on the ability of L2 listeners to comprehend L2 speech (Goh, 2000).

The lack of opportunities for L2 listeners to develop phonological knowledge of L2 words in part relates to a paucity of exposure to cognitively accessible spoken input. However, the difficulties L2 learners experience in developing phonological knowledge of the target language, and by extension the ability to recognise words from speech, also in large part relate to the intrinsic nature of spoken language. As the central objective of this research is to investigate an approach to develop L2 WRS, a consideration of the intrinsic attributes of speech which make this skill challenging for L2 listeners is warranted.

2.2.2.1 Rapidity of the speech signal

Unlike written words on a page which are temporally stable, spoken language is “distributed in time and fades quickly from the perceptual field” (Weber & Scharenborg, 2012, page 387). As such, reading affords far greater control over the rate at which words must be processed when compared to that required when processing spoken language. Readers can stop, re-read, skip ahead or refer to co-text in order to
acquire cognitive support while attempting to comprehend written text (Leeser, 2004). The listener is not afforded such opportunities due to the rapid and ephemeral nature of speech (Roussel, 2011). Moderately fast rates of speech for English conversation may range from 230 to 260 words per minute (Tauroza & Allison, 1990). To successfully manage the processing load imposed on the listener by the rapidity of speech, the processes of activation and competition must occur very rapidly. Despite the rapidity of the cognitive processes required, L1 word recognition is accomplished effortlessly by adult listeners and is characterised by a high degree of automaticity (Segalowitz & Hulstijn, 2005). The automaticity of word recognition typical of L1 listeners is facilitated by the possession of implicit knowledge, the construct of knowledge understood broadly to equate to linguistic competence (R. Ellis, 2005). The application of implicit knowledge is evident when language performance is fluent, automatic and intuitive (Ellis, 2008). For the typical adult native listener, L1 spoken words are recognised largely below the level of conscious control and therefore draw very little on finite cognitive resources such as attention. The automatic and therefore relatively effortless word recognition typical of a native listener is significant in two important regards. First, it facilitates accurate and thorough perception of the composite words present in an utterance in a manner that draws very little on the cognitive resources of the listener. Secondly, it facilitates the ongoing availability of the cognitive resources needed for the interpretation of the meaning of an utterance (Field, 2008b; Segalowitz & Hulstijn, 2005).

In contrast to the effortlessness typical of L1 word recognition, L2 listeners have great difficulty in recognising the words in connected speech (Goh, 2000; Graham, 2006). Without the degree of automaticity of word recognition typical of native listeners, L2 listeners must expend considerable powers of attention to extract linguistic data from the speech stream. The cognitive burden brought about during this process, often results in the linguistic input encoded within continuous speech being difficult or impossible to access for L2 listeners (Goh, 2000). An inability to map the components of the speech stream rapidly enough results in L2 listeners having to depend on contextual knowledge and guessing in an effort to extract meaning from speech (Field, 2008a).
2.2.2.2 The blended nature of the speech signal

Another key difficulty associated with L2 word recognition from speech is the blended nature of spoken language. Whereas many written languages have blank spaces which mark the boundaries between adjacent words, no such overt boundaries consistently appear between the composite words of fluently articulated speech. The example provided in Figure 2.2 below shows a relatively slowly and clearly uttered sample of speech represented visually as a waveform. The vertical axis represents the amplitude of the sound wave and the horizontal axis represents time. The approximate time duration of this utterance was 3.2 seconds. Breaks in the amplitude of the waveform represent momentary intervals of silence which occurred during the utterance. As can be noted these intervals on occasion fall between words (as with the interval within the words “India” and “is”), on other occasions within words (as with the interval within the word “second”), and on other occasions no interval is present between clusters of words (as with the word cluster “country in the word”).

Figure 2.2: Waveform of the sentence: “India is the second most populated country in the world”

Native listeners have little problem in segmenting the lexical units present within streams of continuous speech by virtue of implicit L1 phonotactic knowledge. In contrast, non-expert L2 listeners typically lack L2 phonotactic knowledge and therefore recognising words from L2 speech is problematic (Al-jasser, 2008). Further, the phonotactic constraints which can be applied in the segmentation of words are language
specific. Thus, listeners’ L1 phonotactic knowledge can interfere with L2 spoken language processing and therefore lead to difficulties in the recognition of words from L2 speech (Weber & Cutler, 2006).

2.2.2.3 Phonemic variation intrinsic to the speech signal

The phonemes and words in normal connected speech are blended, with sequential speech sounds influencing one another’s acoustic nature (Al-jasser, 2008; Buck, 2001; Hulstijn, 2003). This blending is brought about through the efficiency of movement of speech articulators as they are continuously transitioned from one position to the next during fluent speech. The resultant change to the pronunciation of fluent speech is referred to as co-articulation. The phonological form of a word as it is pronounced in fluent spoken language therefore can be significantly different than that of that same word pronounced in isolation (Buck, 2001; Field, 2008a; Rost, 2002; Vandergrift, 2007). There are several categories of such phonemic variation including: assimilation, reduction and elision. Assimilation describes the influence of one phoneme on the next, such that their phonetic identity becomes fused (citation form: “terrorist cell” /ˈtɛrərɪst sɛl/, assimilated form: “terr(or)ist cell” /ˈterɪst sɛl/). Elision refers to the omission of phonemes which facilitate ease of pronunciation (citation form: “last song” /ˈlɑːst sɔŋ/, elided form: “las(t) song”, /lɑːsɔŋ/). Reduction typically occurs as a result of the emphasis of vowels at the expense of weaker consonants (citation form: “there he goes” /ˈðeə(ʔ)iːgəʊz/, reduced form: “there (h)e goes” /ˈðeə(ʔ)iːəʊz/ (Adapted from Rost, 2002). The phonological modification of fluently articulated speech can result in learners being unable to recognise words present in the speech stream (Henrichsen, 1984). Indeed, a notable problem for second language listeners is the inability to recognise words in the aural modality despite having existing knowledge of those words (Goh, 2000).

As outlined, the intrinsic features of speech present specific difficulties for L2 listeners. Fluently articulated speech is rapid and does not have the reliable boundaries between words typical of most written languages. Additionally, the phonological form of words spoken in fluent speech is prone to alteration due to co-articulation. These attributes of speech pose few difficulties for the native listener. This is the case as extensive exposure to contextualised target language has conferred implicit phonological knowledge which enables native speakers to automatically recognise words in streams
of speech. If word recognition is automatic, it draws very little on the finite cognitive resources of the listener. Automaticity in word recognition enables the attentional capacity to be directed towards processing the meaning of a spoken message (Segalowitz & Hulstijn, 2005). Conversely, the early stage L2 listener typically struggles with the cognitive load imposed on the short term memory by streams of fluent L2 speech (Goh, 2000, Graham, 2006). With far less contextualised exposure to the target language in the aural modality, the L2 listener’s store of implicit lexical knowledge is less strongly structured according to the attributes of the target language. As a consequence, words are difficult to process quickly enough, the boundaries between words are difficult to establish and the phonological form of connected speech adds an additional level of difficulty to spoken word recognition (Field, 2008b)

2.3 Modelling Listening

2.3.1 Why model the cognitive processes of L2 listening?

Listening comprehension presents a unique set of challenges for language learners and teachers. Many of these challenges stem from the largely “internal” cognitive nature of the skill (Vandergrift, 2007). Listening is arguably the most covert of the four macro-skills (Rost, 2002; Vandergrift, 2007). Of course speaking, reading and writing also depend on underlying cognitive processes; however, there are key differences between listening and these other skills. For example, unlike the skills of speaking and writing, listening does not necessarily involve overt physical manifestations of the skill’s successful completion. Speaking involves the physical manipulation of speech articulators and the generation of acoustic signals. Accordingly, manifestations of performance such as pronunciation, fluency, and language usage can be readily modeled by teachers. Similarly, the speech performance of language learners can be directly accessed and assessed. Writing involves the manipulation of tools and the resultant production of graphemes. In a similar sense to speaking, writing can also be modelled by teachers and the written products of learners can be readily recorded and assessed. In contrast “the product of listening comprehension is a construction or representation of meaning in the mind” (Buck, 2001, page 99). Consequently, the processes involved in
listening comprehension alone are largely inaccessible to direct quantification (Rost, 2002).

The “invisible” nature of listening poses a great practical challenge for language teachers attempting to teach and assess this skill (Field, 2008b). Assessment of listening necessarily depends on gathering empirical data based on observation of task completion (Vandergrift, 2007). Accordingly, the validity of modes of assessment of L2 listening are prone to threats from construct-irrelevant variance (Messick, 1996) as abilities other than those central to listening comprehension must be tapped as part of the assessment process. As a result, care must be taken to ensure L2 learning and assessment programs amply reflect the key elements of L2 listening performance. In order to identify the key elements of successful listening, robust models which describe the processes involved in listening are of strong value.

2.3.2 Models of listening

Psycholinguistics, the branch of cognitive psychology which brings together the fields of linguistics and psychology, is uniquely poised to address the difficulties associated with L2 listening. Psycholinguistics seeks to understand the manner in which human language is stored, processed and acquired (Field, 2004b). A strength of psycholinguistic research as a foundation for pedagogical approaches is the tradition’s strong dependence on experimental evidence in the development of its theoretical models. As such, psycholinguistic models of language processing provide a robust theoretical foundation from which to conceptualise the cognitive processes of listening. To date the influence of psycholinguistics on the field of SLA has been relatively limited. However, there is a growing appreciation from within the field of SLA for the need to carefully consider the internal cognitive processes which underpin language skills. As such the value of applying psycholinguistic models of language processing to supplement existing practice in the field of second language acquisition is becoming more explicitly recognised (Field, 2008d).

Below is an overview of three models which have been previously used to analyse components of the listening process. A focus has been maintained on those models which have been applied in research which has directly investigated L2 listening.
2.3.2.1 Anderson’s three phase model

Drawn from the field of cognitive psychology and applied originally to first language listening, Anderson’s (Anderson, 2000) three phase model was a relatively early theoretical model of listening used in L2 listening research (Goh, 2000; O’Malley, Chamot, & Kupper, 1989). This model posits listening comprehension as consisting of three discrete phases: perception, parsing and utilisation. According to this model the lowest level of processing is perception, which involves the segmentation of the components of the acoustic signal. During perception, phonemes are extracted from speech and held in echoic memory (Anderson, 2000). Parsing involves the coalescence of words in a meaningful sequence by association with linguistic representations already present in the listener’s mental lexicon. Utilisation is the highest level of representation according to the three phase model. Utilisation involves the establishment of the meaning of an utterance through association of the linguistic information with schemata present in long term memory. O’Malley, et al., (1989) used the three phase framework to investigate the cognitive processes of L2 listeners and to draw conclusions regarding the attributes of more and less skilled listeners. Goh (2000) used the same framework to categorise the cognitive origin of the difficulties experienced by L2 listeners. Although the three phase model has contributed to early understandings of the processes involved in L2 listening, it can be criticised on the basis of its assertion that listening comprehension involves a predominantly unidirectional cognitive process which proceeds sequentially from lower to higher order processes (Graham & Macaro, 2008). This attribute of the model is indicative of early information processing models of listening which suggested a relatively fixed sequence of interaction between levels of representation. Such models asserted listening comprehension involves a fundamentally “low to high” linear progression of information processing (Marslen-Wilson, 1984). Anderson’s model has been largely superseded by those models that adopt a less strictly ordered sequence of processing events (Marslen-Wilson, 1989). As will be explained, more contemporary models of L2 listening emphasise the bi-directionality and interactive nature of information processing.

2.3.2.2 Interactive bottom-up and top-down model

The concepts of bottom-up and top-down information processing represent the foundation of most contemporary models of L2 listening comprehension. The broad
influence that the top-down and bottom-up processing paradigm has exerted on the field is indicted by its recurrence in a range of academic literature relating to L2 listening (Bonk, 2000; Buck, 2001; Field, 2003; Goh, 2000; Graham, 2006; Graham, Santos, & Vanderplank, 2010; Robin, 2007; Tsui & Fullilove, 1998; Vandergrift, 2007; Vandergrift & Tafaghodtari, 2010; Wilson, 2003). In contrast to the previously outlined three phase model of listening comprehension, which conceives listening as comprising of discrete levels of processing, the top-down and bottom-up processing model emphasises the bi-directionality of mental processes which facilitate listening comprehension (Field, 2004a).

The bottom-up component emphasises the direction of information transformation from lower levels of representation to higher levels. Therefore bottom-up processing involves the perception and coalescences of phonemes to syllables, syllables to words, words to phrases and so on. Fundamental to bottom-up listening processes is the listener’s ability to differentiate and recognise the small scale components of speech such as phonemes and words (word recognition). Conceptually, bottom-up processes involve the concrete linguistic units of speech being transformed to larger and larger units of language.

Various definitions of top-down processing have been put forward in regards to L2 listening. Top-down processes are largely associated with the use of non-linguistic sources of knowledge to establish the meaning of speech. A recurring theme relating to top-down processing is the employment of pre-existing contextual knowledge in the establishment of meaning (Flowerdew & Miller, 2005). Field (2008a) equates top-down processing with the use of context in the identification of smaller units of meaning. Field (2008a) differentiates two sources of contextual information: a) the knowledge a listener possesses through general experience, and b) knowledge of what has been previously said in the conversation. Rost (2002, page 284) defines top-down processing as the “form of language processing that bases inferences on expectations and predictable generalisations cued by the incoming language”.

Such definitions draw attention to the inherent connectedness of bottom-up and top-down processing events. Bottom-up processing is traditionally understood to involve the establishment of larger and larger units of language by serially connecting speech data. However these processes are also facilitated by the application of appropriate contextual knowledge. Similarly, the construct of top-down processing or the employment of
appropriate contextual information is typically understood to be dependent on the application of non-linguistic domains of knowledge. However the engagement of appropriate pre-existing knowledge as part of the comprehension of spoken language must by necessity also depend on the recognition of linguistic cues in the speech signal.

The interconnectedness of bottom-up and top-down processing is readily understood by considering simple language usage. For example, bottom-up processing would enable the utterance “That’s extraordinary!” to be decoded and understood at the linguistic level. However, in order to establish comprehension on a holistic level, the listener would also have to be privy to contextual information. For example, only by seeing or knowing what the person who uttered “That’s extraordinary!” was referring to, would the listener truly be able to comprehend the utterance in the fullest sense. Top-down processing also provides predictive power during listening which in turn facilitates more efficient bottom-up processing. For example, on hearing: “It’s freezing outside you’d better put on a ….”, a variety of potential candidate words to finish the sentence are activated in the mind, for example coat, jacket or scarf. This is the case despite the listener having no direct linguistic input relating to the phonological form of the terminal word in this example sentence (Buck, 2001; Grosjean, 1980). As can be appreciated from the previous examples, bottom-up and top-down listening processes have a strong degree of interdependence.

There is broad acknowledgement of the interdependence between bottom-up and top-down processes in successful L2 listening comprehension (Buck, 2001; Rost 2002). Interactive information processing models assume that information relating to a speech signal can be processed at a number of different levels of representation at the same time. So-called called parallel processing implies that top-down and bottom-up processing occurs simultaneously (Field, 2004b). Such interactive models emphasise the mutually facilitative role of bottom-up and top-down listening processes. In this way listening comprehension may be viewed as occurring at multiple points along a continuum between bottom-up and top-down processes. The exact nature of the interaction between bottom-up and top-down processing in listening comprehension is not clearly understood (Tsui & Fullilove, 1998). However the degree to which listening processes draw upon either end of this processing spectrum is likely to be dependent upon a listener’s phase of learning and knowledge status, the specific nature and content
of the auditory message, and the specific purpose for which the listening task is being undertaken (Vandergrift, 2007). The relative importance assigned to each end of the processing spectrum has provided a formative context from which current perspectives on teaching L2 listening have emerged, with efforts to establish the relative importance of bottom-up and top-down processing in listening comprehension providing a significant agenda for L2 listening researchers (Field, 2004a; Goh, 2000; Hasan, 2000; Tsui & Fullilove, 1998). To date a significant proportion of the research relating to L2 listening pedagogy has been directed toward investigations of the role of top-down processing (Wilson, 2003). As a consequence, and of relevance to the overarching objectives of this research, investigations into the role of bottom-up processing in L2 listening pedagogy have been relatively underemphasised (Broersma & Cutler, 2008).

2.3.2.3 The decoding and meaning building model

An alternative framework used to conceptualise the processes involved in listening, put forward by Field (2008b), identifies listening as consisting of two categories: decoding processes and meaning building processes. Within this framework decoding processes are characterised by sub-lexical, lexical and supra-lexical levels of processing which enable the establishment of the literal meaning of an utterance. According to this model, the primary component of decoding is word recognition which involves the mapping of speech sounds onto the words that those speech sounds represent. The decoding arm of this two-part model provides a useful adjunct to the concept of bottom-up processing evident in bottom-up and top-down models of L2 listening. This is the case as this model identifies the establishment of the literal meaning of an utterance as an endpoint for the decoding phase and as such provides a more constrained definition of lower-level processes (Field, 2004a).

Meaning building processes are characterised by cognitive actions which enable the decoded message to be interpreted beyond its literal meaning. At the base of the meaning building process is the conversion of the literal meaning of an utterance to an abstract semantic representation. This abstract representation is enriched by alignment with contextual knowledge. It is also enriched by that which has been understood by the listener up to that current point of listening (text-so-far or co-text). The resultant contextualised abstract representation is referred to as “meaning representation”. Meaning representation continually feeds the on-going representation of what has been
understood so far, to give rise to what Field (2008b) refers to as the discourse representation.

The two part model involving decoding and meaning building processes provides additional depth to the interactive top-down bottom-up model of listening. A key advantage of this model is that it clearly positions the role of decoding as central to the listening processes overall. This is significant as it establishes a paradigm of listening which reflects the mounting body of evidence suggesting the primacy of lower order perceptual decoding processes in skilled L2 listening (Field, 2008a; Goh, 2000; Lynch, 2006; Tsui & Fullilove, 1998).

2.3.3 The cognitive basis of problems experienced by L2 listeners

As outlined in section 2.2.2, the specific attributes of spoken language help to explain why L2 listeners experience difficulty with listening. For example, speech rate has a measurable impact on listeners’ ability to extract information from spoken text (Zhao, 1997) as does the phonological modification of the words encoded in fluent speech (Al-jasser, 2008; Buck, 2001; Hulstijn, 2003). However, the difficulties experienced by L2 listeners can not be adequately explained by only considering the attributes of a spoken text (Brindley & Slatyer, 2002). As Buck (1994, page 162) asserts, “two listeners hearing exactly the same text in the same situation are likely to have two very different contexts for interpreting what they hear”. It is the “cognitive environment” of the listener which provides the greatest explanatory value for the ability of a listener to interpret L2 speech. Cognition is defined as the human capability to “perceive information, to encode and store perceived information, to retrieve information”, to use thought to solve problems, “and to behave skillfully on the basis of stored information” (Hulstijn, 2003, page 414).

The following section aims to review the existing literature which has specifically aimed to determine the cognitive basis of L2 listening problems. Establishing the cognitive basis of the problems experienced by L2 listeners, will provide a context from which pedagogical decisions about teaching L2 listening can be based. In turn this review will establish a foundation from which to exposit the cognitive processes indicative of skilled and less skilled listeners described in Section 2.3.4.
Goh (2000) examined the listening difficulties experienced by 40 Chinese ESL students using recall protocols, learner diaries and interviews. This study was aimed at determining the comprehension problems that occurred while the L2 listeners were attempting to comprehend a L2 listening text. Anderson’s (2000) three phase cognitive model was used to categorise the listening problems which were identified. Of the ten problems categorised in this study, five were considered as perceptual problems, three were problems associated with parsing and two were problems associated with utilisation. Generally, this indicates that the problems associated with L2 listening span lower and higher-level cognitive processes.

Goh identified problems reported by a majority of students of both high and low ability. These findings are of interest as they cast light on the most common problems experienced among the sample group involved in the study. The most common problem reported by both ability levels was that the speech input that was successfully perceived was quickly forgotten. It appears that in many cases the words that students recognised could not be retained in memory long enough to enable that linguistic information to be processed at higher levels of semantic interpretation. The source of this problem appeared to stem from the cognitive overload imposed on students as they continually needed to perceive new sections of the input as the listening texts proceeded. Previously perceived elements of that speech signal were purged from the working memory of the listener. This problem suggests that the cognitive load imposed by the perceptual or lower-level processes was so great that few cognitive resources were available for higher-level processing. Chunks of decoded language could not proceed from the literal to abstract levels of representation and thus could not be processed for meaning.

The second most common problem identified by a majority of high and low ability participants related to word recognition. This problem related to the degree of automaticity in the recognition of the phonological form of words. Students reported that they were unable to recognise words from speech despite reporting that they had knowledge of those words. This finding indicates that a major problem experienced by the students in this study was insufficient knowledge of the phonological form of words and thus highlights the importance of the knowledge of the phonological form of words in L2 listening comprehension. Goh concludes that the learners’ self-reported method of learning may have contributed to the lack of depth of phonological knowledge
associated with the words in the listening text. Students reported knowing some of the words in orthographic form but could not associate this knowledge with the phonological form of the word when those words were present in speech. Without an accurate mental representation of the word’s phonological form, the ability to automatically recognise words and therefore strategically manage cognitive resources will be limited (Segalowitz & Hulstijn, 2005). This finding suggests that knowledge of the meaning of a word, knowledge of the orthographic form of a word, or knowing the translation of a word may not guarantee that the word can be recognised when presented in connected L2 speech.

The inability to automatically recognise words in connected speech is a significant L2 listening problem with a cognitive origin. These problems can be explained most directly by considering processing at a lexical and sub-lexical level. On a lexical level, word recognition may be hindered by limited vocabulary knowledge. This is the case as pre-existing word knowledge facilitates the rapid alignment of incoming lexical input with lexical representations stored in the mental lexicon. Words that are unknown to a listener will be difficult to recognise (Hulstijn, 2003). Breadth of vocabulary knowledge is an essential component of successful listening comprehension (Bonk, 2000) as highlighted by findings that 98% lexical coverage was required for test takers to cope with spoken texts (Stæhr, 2009). As shown previously, this knowledge needs to be inclusive of the phonological form words take in fluent speech (Goh, 2000). Without such knowledge words that are known may not be recognised.

The ability of L2 listeners to recognise words from speech is also negatively impacted by increased competition between candidate words held in the mental lexicon (Broersma & Cutler, 2008). As outlined in section 2.2, the cognitive operations which facilitate word recognition involve the processes of activation and competition. Activation occurs when perceptual information triggers the excitation of potential word candidates in the listener’s mental lexicon (McCelland & Elman, 1986). Again returning to the example used in section 2.2, on perceiving the syllable /laɪ/, known words such as light, lightning, lie, lying, lion, or lice are automatically activated. As more of the word is perceived, candidate words compete for selection. However the specificity of activation and competition between candidate words is dependent on the reliable perception of phonemic categories. For example, if an L2 listener has difficulty in
reliably distinguishing /laɪ/ and /raɪ/ then on hearing /laɪ/ not only will the preceding list of words be activated but also will words such as right, righteous, rye, rice, write, writer, or writhe (Broersma & Cutler, 2008). The activation of more candidates in response to aural stimulus invariably leads to slower word recognition (Norris, McQueen, & Cutler, 1995).

Recognition of L2 words in speech is also impeded by previously acquired implicit knowledge from the L2 listener’s first language. Each language has its own suite of phonotactic rules which determine the sound sequences that can occur at the beginning (onset) and end (coda) position of a syllable. In English for example the speech sound represented by the consonant cluster /pr/ can occur as an onset syllable, as in the word “prey”; however, it is an illegal coda cluster. In English, the speech sound /ŋ/ can’t occur as a syllable onset but it can occur as a syllable coda, as in the word “sing”. However the speech sound /ŋ/ can occur as either an onset or a coda in Thai. Phonotactic rules make up part of the implicit phonological knowledge which enables listeners to recognise words in speech by distinguishing when one word ends and another begins (lexical segmentation) (Al-jasser, 2008). However as phonotactic constraints are language specific, implicit first language phonotactic knowledge can lead to less rapid word recognition in L2 listeners (Weber & Cutler, 2006). The language specific constraints, such as phonotactic rules, which guide the recognition of words in one language may negatively impact on the ability to decode another language with different phonotactic rules (Al-jasser, 2008). The limited speed at which L2 listeners can engage lower order listening processes in order to recognise words is a significant cognitive difficulty associated with L2 listening.

Some theorists have argued for the primacy of bottom-up processes in listening as the linguistic information derived from these processes is facilitative in the integrated employment of higher and lower order listening processes (Marslen-Wilson, 1989; Wilson, 2003). However, an overdependence on bottom-up listening processes can also cause problems for L2 listeners (Hasan, 2000). Expending excessive attention to bottom-up processes may negatively impact a listener’s ability to apply top-down processes, which in some instances may compensate for the inability to recognise words in fluent speech (Graham, et al., 2010; Vandergrift, 2007). The application of processes which are used to make up for gaps in linguistic data are referred to as listening
strategies and involve exercising conscious control over cognitive processes in order to achieve learning goals (Macaro, 2006). Such listening strategies involve the engagement of mental processes in an effort to comprehend information that is ambiguous (Field, 2008b; O'Malley, et al., 1989). Considering the previously described difficulty L2 listeners have in lower-level processes such as automatically recognising words in connected speech, successful early stage L2 listening will depend strongly on the ability to strategically apply contextual information to make up for gaps in incompletely decoded linguistic information (Hasan, 2000). An ability to draw on top-down information sources in a manner which effectively manages the finite cognitive resource available to the listener has been shown to be facultative in L2 comprehension (Vandergrift, 2003b).

The strategic use of contextual information is likely to involve the employment of metacognitive strategies (Graham, et al., 2010). Metacognitive strategies involve the executive control of cognitive processes including planning for, monitoring, and evaluating the success of a learning experience (O'Malley & Chamot, 1990). Such strategies enable listeners to allocate cognitive resources in a strategic manner in accordance with ongoing monitoring of comprehension levels (Graham, et al., 2010). Listeners who are unable to effectively employ metacognitive strategies are unlikely to be able to effectively manage the cognitive demands intrinsic to L2 listening in the early stages of learning. Without the employment of strategies to manage finite cognitive resources, L2 listeners may depend disproportionately on bottom-up processes (Osada, 2001).

The ability to effectively apply listening strategies is linked to the underlying linguistic competence of the L2 listener. Graham, et al., (2010, page 14), despite advocating for the value of strategies in listening pedagogy, additionally assert that “a minimum level of vocabulary recognition is required before non-linguistic knowledge in the form of prior or world knowledge or higher order reasoning” can be used effectively. Therefore a significant problem experienced by L2 listeners is the inability to effectively apply listening strategies; however, the ability to apply such strategies is dependent on the adequacy of word recognition and the underlying implicit linguistic competence that underpins this ability.
When considering the relative significance of the cognitive problems experienced by L2 listeners, it is important to consider the manner by which lower-level and higher-level cognitive processes interact. Field (2004a) drew on Stanovich’s (1980) interactive-compensatory mechanism to describe the interaction between top-down and bottom-up processing involved in L2 listening comprehension. According to this framework, top-down or contextual information is compensatory in nature. By this it is meant that when bottom-up information is flawed or incomplete, such as when word recognition is not automatic, top-down or contextual sources of information are drawn on more strongly to compensate for uncertainty about the content of the spoken message. Field (2008a) added empirical weight to this assertion through implementing a gating research paradigm. This research methodology involved short sections of a spoken text being listened to by L2 listeners in stages. On each occasion the length of the spoken text was extended such that more and more of the linguistic information encoded in the text was revealed to participants. After listening to each section, participants predicted what they had heard. Results indicated that L2 listeners, when compare to L1 listeners, were more reluctant to abandon hypotheses about the content of what they had heard despite the presence of linguistic evidence that runs counter to their hypotheses. Field concludes that a significant problem for L2 listeners is the inability to rapidly amend tentative hypotheses about the content of a spoken message in light of the available evidence present in the speech signal. Therefore, using contextual information that is not supported by reliable decoding processes can lead to breaks in comprehension (Tsui & Fullilove, 1998)

2.3.4 The characteristics of skilled and less skilled L2 listeners

In the preceding section a review of relevant literature was presented in order to highlight the cognitive basis of the problems which impede L2 listening comprehension. From a consideration of these problems, it can be understood that the ability to comprehend L2 speech is primarily dependent on the ability to overcome the cognitive constraints imposed on the listener by virtue of the attributes intrinsic to fluent speech (Vandergrift, 1998). Sufficient levels of automaticity in word recognition are needed to ensure listeners do not expend excessive attention in surface level decoding. Spare cognitive resources are required to pass literal chunks of recognised speech through to higher levels of meaning building. Top-down processes, including higher order
executive cognitive processes, are required to manage contextual knowledge and to manage cognitive resources in a strategic manner. Furthermore these fundamental cognitive processes need to be applied in an integrated manner.

Assuming a key goal of language teachers is to facilitate learning experiences which contribute to a learner’s L2 language competence, determining the factors which differentiate more and less skilled L2 listening is of strong interest. In this regard, comparisons of the cognitive operations undertaken by L2 listeners of different proficiency levels represent a valuable source of empirical data for language educators. It is asserted here that an understanding of the cognitive operations typical of lower and higher proficiency L2 listeners provides a valid foundation from which to devise targeted pedagogical interventions.

What follows is a review of literature which has aimed to draw conclusions about the cognitive processes which can be attributed to higher and lower levels of L2 listening comprehension.

2.3.4.1 The employment of listening strategies

Much of the research which has specifically focussed on the attributes of more and less skilled listeners has addressed the differential use of listening strategies (Goh, 1998, 2002; Graham, Santos, & Vanderplank, 2008; Graham, et al., 2010; O’Malley, et al., 1989). The strong emphasis on listening strategies in this body of research reflects the overall centrality of the role of non-linguistic knowledge or top-down processing in L2 listening research (Macaro, Graham, & Vanderplank, 2007).

O’Malley, et al., (1989) investigated the learning strategies of effective and less effective listeners using Anderson’s (2000) three phase cognitive model to structure their analysis. The study was small (N = 11) and drew qualitative data from learners using think-aloud protocols. According to their findings the use of listening strategies differentiated effective and less effective listeners. Effective listeners used self-monitoring of understanding to negotiate sections of the listening text which were not understood. Ineffective listeners on the other hand, usually stopped listening at a point where an unknown word or phrase was encountered. Effective listeners were reported to be able to attend to larger chunks of language, whereas ineffective listeners attended to
language chunks at the word level. Effective listeners were also noted to be able to strategically apply prior knowledge to extract meaning from the spoken texts.

Goh (2000) investigated the listening strategies applied by a larger group L2 listeners ($N = 40$) of differing proficiency level. Data were attained in a number of ways: learner diaries, small group interviews and think-aloud protocols. In line with the earlier results of O’Malley et al., (1989) high ability listeners were shown to apply metacognitive monitoring strategies to ensure that unknown words and phrases did not impede ongoing efforts to understand the spoken text. In contrast, low ability learners did not apply such strategies when they encountered an unknown section in the listening text and as a result remained focused on those unknown sections. Goh asserts that processing strategies can be applied to facilitate the interpretation of an utterance even when all words are not recognised.

These findings also align with the attributes of less skilled listeners according to Vandergrift (1998). According to his results based on analysis of think-aloud protocols, successful listening depends on the ability to allocate less attention to the processes of decoding words thus freeing up mental resources for higher levels of interpretation. Vandergrift contends that because listening is undertaken in real time, successful listening depends on selective processes. Vandergrift emphasizes the role of strategies, especially comprehension monitoring in successful L2 listening comprehension. Successful listening requires the ability to continually monitor the validity of what has been comprehended in accordance with newly perceived information that may run counter to the original interpretation. Vandergrift also asserts that less successful listeners rely more heavily on background knowledge in an effort to interpret unfamiliar linguistic information. This assertion is in line with the “compensatory” role of top-down processing hypothesised by Field (2008a). Again Vandergrift (2003b) used think-aloud protocols to elicit the types of strategies employed by more and less skilled listeners. It was found that skilled listeners more flexibly applied both top-down and bottom-up listening strategies. Additionally successful listening was characterised by a lack of mental translation. Vandergrift also found that more skilled listeners used metacognitive strategies significantly more often than less skilled listeners.

Graham, et al., (2008) undertook a case study of two L2 students, one of relatively high and one of relatively low listening proficiency. This longitudinal study aimed at
discerning how L2 listening strategy use changed over time. Again think-aloud protocols were used to illicit the strategies engaged by the participants during listening tasks. The lower ability student predicted words and used selective attention which primarily involved listening for specific vocabulary items. The low proficiency listener’s strategic approach was typified by deductive processes based on the types of linguistic data that were not reliably perceived. Presumable low lexical knowledge of the listening text and limited automaticity of word recognition led the low proficiency learner to adopt this strategy.

The higher proficiency listener adopted a strategy of comprehension monitoring by which the interpretation of the text was treated as tentative. The higher proficiency listener regularly checked his interpretation and thus exercised executive control over his listening processes. This was in contrast to the low proficiency listener who rarely refined interpretation based on uncertainty of the text. The high proficiency listener exercised selective attention, however it was implemented once overall understanding of chunks of language were established. The high proficiency listener was therefore able to identify salient information from what had been already decoded. Apparently, sufficient cognitive resources were available to hold the decoded chunk of language in memory while strategic application of comprehension monitoring of the decoded language was implemented. It was noted that on several occasions the high proficiency listener automatically recognised chunks of language.

Graham, et al., (2008) found that participant strategy use remained stable over the duration of their 6 month study. Interestingly the higher proficiency listener had developed strategic listening without explicit instruction. Additionally, the lower ability student implemented a number of strategies that have been regularly acknowledged as those indicative of an effective listener (Vandergrift, 2003a).

Graham, et al., (2010) analysed data from 14 L2 listeners to investigate the listening strategies and sources of knowledge tapped by learners as they undertook listening comprehension tasks. The study found both high and low L2 listening ability students tended to rely heavily on linguistic knowledge. Listeners with high linguistic knowledge were however more likely to draw from both linguistic and non-linguistic knowledge sources. The highest performing listener in the study demonstrated the ability to integrate linguistic knowledge, perception and contextual knowledge. This contrasted
the finding that low proficiency listeners typically relied on only one source of knowledge which was generally linguistic knowledge. Overall it was concluded that students with higher linguistic knowledge were more able to integrate linguistic and non-linguistic information in order to establish comprehension of a text. More skilled listeners were able to match background knowledge to the content of the text through accurate linguistic recognition. The ability to recognise chunks of words was shown to be facilitative of effective strategy use.

The preceding studies have provided a clearer picture of the types of listening strategies typically employed by higher and lower proficiency listeners. However, it should be noted that the studies mentioned in this section have some methodological and theoretical weaknesses. Firstly, each of the studies had a relatively small sample size. Of the studies reviewed in this section, the numbers of participants ranged between two and forty. As such, interpreting these results in broader contexts should be undertaken with caution.

Another important consideration in the studies detailed above is the manner by which data was acquired. All of the studies depended strongly on think-aloud protocols in order to illicit the mental processes at play while, or just after, L2 listening events had taken place. As listening is a difficult construct to measure directly, think-aloud protocols present a valuable source of information regarding the thought processes which are being undertaken during the listening process (Buck, 1991). However, the verbal protocol data collection method will primarily tap into controlled, conscious processes that are able to be verbalised by the learner. This method of data collection is well suited to elicit listening strategies as such strategies are consciously controlled and are therefore dependant on explicit knowledge. Explicit knowledge is able to be verbalised and is “typically accessed through controlled processing when L2 learners experience some kind of linguistic difficulty in the use of the L2” (Ellis, 2004, page 245). Graham, et al., (2010) acknowledge that listening strategies are primarily engaged when learners experience a degree of difficulty with the listening task. In their 2010 study they selected listening texts that were intentionally challenging for their study participants such that strategy use was elicited, as tasks which are able to be completed easily and automatically tend not to elicit strategy use (Graham, et al., 2010).
Think-aloud protocols are unlikely to provide direct information about the knowledge and processing types strongly linked to implicit knowledge. According to Hulstijn (2007, page 706) “implicit knowledge is not open to conscious inspection” and “its processing components cannot be verbalised”. The ability to automatically and fluently recognise units of speech occurs at an unconscious level and is therefore not readily monitored in a conscious manner (Hulstijn, 2003). Therefore, think-aloud protocols are unlikely to elicit information about the degree to which success in listening can be attributed to processes such as word recognition, which depends on automatic processes. This limitation of think-aloud protocols was alluded to by Graham, et al., (2010) as they acknowledge that in a significant number of instances the think-aloud protocols could not elicit the means by which learners had answered the listening comprehension questions in their study.

Another point of general criticism relating to the studies described was a lack of analysis or commentary on the potential reactivity of the think-aloud protocols. Reactivity refers to the potential effect that the act of verbalising thoughts has on the task being undertaken and reported upon (Barkaoui, 2010). The effect of think-aloud protocols on time constrained tasks is particularly strong (Bowles, 2010) and therefore a lack of consideration of these potential threats to validity represents a flaw in the methodological approach of the previously mentioned studies.

The preceding investigations considered the attributes of more and less skilled listeners with a primary focus on the listening strategies employed by L2 listeners, in particular the different ways listeners employ top-down strategies and how these differences relate to L2 listening comprehension levels. Much of the research in this area has emphasised the different application of listening strategies as an indicator of more and less skilled listeners. However, there is also a smaller body of empirical evidence which suggests that the ability to recognise words in connected speech is an underemphasised construct which also differentiates more skilled and less skilled listeners. It is acknowledged that the application of listening strategies is an important component of skilled L2 listening, however, as highlighted by Lynch (2006) and detailed in the studies reviewed below, there is a body of empirical evidence which suggests the primacy of lower order decoding processes, such as word recognition from speech, in determining L2 listening comprehension success.
2.3.4.2 The primacy of L2 WRS

Tsui and Fulliove (1998) is an early yet significant study which specifically investigated the attributes of more and less skilled L2 listeners from a knowledge source perspective. This study was primarily aimed at determining whether bottom-up or top-down processing was more important in determining higher listening test performance in English as a Second Language (ESL) students. Approximately 20,000 participants in public examinations in Hong Kong were involved in this study, which involved analysis of two types of listening test question: “matching schema” and “non-matching schema”. In matching schema test items, the listeners who formed hypotheses based on the initial input could primarily use top-down processes to select the correct answer. Non-matching schema test items required candidates to rapidly process linguistic input which ran counter to the initial input. Therefore non-matching schema test items could not be answered by depending on contextual information in the listening text. For these question types, the selection of the correct answer primarily depended on the test takers’ ability to rapidly process the linguistic input in the stimulus listening text. In short, these question types were written such that if the critical pieces of linguistic information could not be successful processed through use of bottom-up processes, strategic use of background knowledge or context could not make up the gap. Analysis of the results indicated that those candidates who correctly answered the non-matching question types were the more skilled test takers. It is possible to level some criticism at the study in that the format of the questions types described may have resulted in an excessive emphasis on the short-term memory of candidates (Call, 1985). Furthermore, the listening texts were written and read and did not have the typical features of “authentic” spoken text (Gilmore, 2007; Rost, 2002). However, the strength of this study was its size: with 117 test items and approximately 20,000 participants, it represents a major study suggesting the importance of bottom-up processing for successful L2 listening.

Yi’an (1998) also provides empirical support for the primacy of lower order linguistic processing in L2 listening comprehension. The small scale study was undertaken among Chinese-speaking ESL students (N = 10) and involved the students listening to a short English language listening text in order to answer a series of multiple choice questions. A key objective of the study was to investigate the manner by which linguistic and non-linguistic knowledge sources interacted as students answered the multiple choice...
questions. The interaction of these knowledge sources was tapped through retrospective verbalisations gathered while learners selected the multiple choice answers. An analysis of the retrospective verbalisations suggested that when the salient section of the listening text could not be adequately decoded, learners tended to rely on the engagement of non-linguistic knowledge to select the answer. Thus non-linguistic knowledge was applied in order to compensate for gaps in knowledge resultant from incomplete lexical decoding. A key finding was that the application of non-linguistic knowledge did not guarantee the selection of the correct answer. Further, the inability to decode linguistic information from the text, and the resultant engagement of non-linguistic information in an effort to compensate for limitations to lexical decoding, often led to the sections of the text which had been successfully decoded to be overridden. Conversely, confident and accurate decoding of linguistic input enabled listeners to override pre-existing beliefs which ran counter to the message present in the speech signal. Competent linguistic processing enabled a refined use of contextual or non-linguistic information in a manner which facilitated comprehension. This interaction between accurate linguistic processing and non-linguistic knowledge activation was indicative of more able L2 listening performance. This was contrasted with the compensatory use of non-linguistic knowledge which led to inaccurate guessing which was more strongly associated with less able listeners.

The ability to accurately decode linguistic information as an indicator of superior listening ability has also been suggested by more recent research by Field (2008a). Field employed a version of the gating research approach to investigate the relative levels of sensitivity of native and non-native listeners to spoken language. The goal of the research was to determine how sensitive L2 listeners were to acoustic cues which disambiguated a previously heard, ambiguous sample of speech. For example, “drəvə” can be interpreted as either “driver” or “drive a”, but can be narrowed down to the latter option once the spoken input “a long” is perceived subsequent to the initial input (Field, 2008a, page 21). Participants listened to several such utterances which were initially ambiguous but became less so as more of the utterance was heard. A key finding was that native listeners rapidly changed their initially incorrect assumptions about the content of the utterance as disambiguating linguistic information was heard. In contrast, non-native listeners were far less sensitive to incoming linguistic information and were thus less able to rapidly change their hypotheses about the content of the utterance.
despite the existence of linguistic information which indicated that an error in recognition was present. These findings complement the studies of Tsui and Fullilove (1998) and Yi’an (1998) in that they clearly highlight the primacy of rapid and automatic decoding routines as an attribute of expert listening behaviour. These results in sum suggest that a key attribute of successful listening performance is the ability to rapidly and accurately decode linguistic information present within the speech signal.

To conclude, the preceding section highlighted the application of listening strategies and the ability to recognise words from speech as important attributes of skilled L2 listeners. The studies reviewed suggest that the capacities to both apply appropriate listening strategies and to recognise words in speech are important components of successful L2 listening. In the following section a focus on approaches to teaching L2 listening will be briefly overviewed. Furthermore the section will provide a summary of contemporary recommendations on how to teach listening in general. The latter part of this section will provide support for the view that there is currently an under-emphasis on pedagogical approaches aimed at developing word recognition from L2 speech.

2.4 Approaches to teaching L2 listening

2.4.1 Brief historical overview of teaching L2 listening

Traditionally listening was not dealt with in a direct manner in the modern language classroom. Early approaches to language learning were influenced by the imperatives of grammar-translation syllabi which stressed reading and writing of languages such as Latin and Greek (Flowerdew & Miller, 2005). The grammar-translation approach to language learning emphasised the ability to read and write the target language and was facilitated by explanation in the first language (L1) and by translation to and from the language being learnt (Richards & Rogers, 2001). Traditional western pedagogical practice was influenced strongly by the objective to be able to understand classic literature, such as that of Latin and Greek, thus the ability to comprehend fluent second language speech was not a priority in these relatively early learning contexts.

Language learning approaches subsequent to the grammar-translation approach did use the aural modality as a fundamentally important component of language teaching.
Notable among these was the audio-lingual method which involved the learners listening to and repeating grammatical structures in a strictly determined manner. The aural modality was primarily used to present new grammatical structures to be learnt (Field, 2008b). The audio-lingual method in large part relied on learners listening to and repeating spoken utterances, with the goal being to gain control of grammatical forms rather than to develop listening skills in the broader sense (Flowerdew & Miller, 2005). Audio-lingual approaches such as those suggested by Lado and Fries (1958) were strongly influenced by contemporary behaviourist paradigms of language acquisition (Skinner, 1956). In the field of language learning, approaches associated with behaviourist theories became effectively defunct after influential criticism (Chomsky, 1959) and the associated trends of the “cognitive revolution” (Miller, 2003). Behaviourist approaches, such as the audio-lingual method, were criticised for emphasising repetition without cognitive engagement or attention to meaning (Flowerdew & Miller, 2005).

From the late 1960s the skill of listening started to become the specific concern of language teachers (Field, 2008b). Interactionist and sociolinguistic paradigms of language learning became increasingly important (Goh, 2008) and gave rise to communicative language teaching. Communicative language teaching emphasised that language teaching must stress links to authentic, transactional language-mediated communication (Flowerdew & Miller, 2005). With an increased emphasis on the social aspects of interaction in language use, the aural modality as a means to facilitate comprehension became increasingly important in the language classroom.

The goal of communicative language classes was to facilitate interaction with a speaker or a spoken text and respond in a socio-linguistically appropriate manner. In the period from the 1970s to the 1980s there was a documented trend in the use of authentic spoken texts with a clear emphasis on the use of context to interpret a spoken text in a manner which facilitates comprehension (Goh, 2008; Wilson, 2003). Field (2008b) provides an overview of the format of conventional listening lessons as being made up of the following components: pre-listening, extensive listening, intensive listening and post listening. Pre-listening involves the establishment of context, motivation for listening and an introduction of key vocabulary items present in the listening text. Extensive listening involves listening to the text for the purpose of establishing a global
orientation to the text. Intensive listening typically involves questions being set before listening to the text again. As part of the intensive listening phase learners are given an opportunity to answer the written questions based on the listening text. It is during this phase that the listeners’ comprehension is assessed according to the degree of “correctness” of written answers to questions. Finally, a post listening phase may be implemented where the function of the language used in the text, unknown vocabulary or the text’s written transcript may be used in association with a final play of the listening text.

The current orthodoxy of L2 listening instruction has been criticised on the basis of its focus on the product of listening – that is, on the correct answer to comprehension questions (Field, 2008b; Vandergrift, 2007). An important turning point in approaches to teaching L2 listening relates to the establishment of the distinction between the pedagogical approaches which emphasise the products of listening and those that emphasise the processes of listening (Field, 1998), which suggest that a focus on product actually “does little to help students understand and control the processes leading to comprehension” (Vandergrift, 2007, page 191). The role of comprehensibility in language acquisition was an influential theoretical assertion in second language acquisition from the early 1980s (Krashen, 1982) and has been expressed in specific reference to recommendations relating to L2 listening. Ridgeway (2000) emphasised that listening is developed through increased practice and exposure to comprehensible listening texts. However, the view that more comprehension based listening practice is sufficient for second language acquisition has been drawn into question by early research from immersion second language programmes (Swain & Lapkin, 1995). Teaching practices which have primarily involved increasing the exposure of listeners to more and more texts and judging the listeners’ performance based on measures of comprehension, have been strongly criticised for not providing diagnostic information as to why errors in comprehension occur during the process of listening (Field, 1998, 2008b). The shallowness of diagnostic information provided by the comprehension approach to teaching listening has led to criticism that this approach provides listening practice but does little to facilitate the development of listening skills (Brown, 1986). Such criticism encourages the employment of speech input in L2 listening pedagogy to encompass the purposes of facilitating not only comprehension, but also the acquisition of the underlying linguistic competence essential for listening (Sharwood Smith, 1986).
2.4.2 Contemporary recommendations for teaching L2 listening

Contemporary recommendations regarding teaching L2 listening have evolved from an understanding of the cognitive processes which contribute to successful L2 listening. Successful L2 listening comprehension depends on the ability to apply lower order processes accurately and rapidly as well as on the ability to apply listening strategies in order to manage the cognitive load typically imparted by virtue of the nature of the speech signal. Contemporary recommendations regarding L2 listening emphasise two arms of instruction: strategy instruction and word and phrase recognition (Graham, et al., 2010; Vandergrift, 2007). As will be explained in the following sections, a focus on the application of listening strategies has dominated the research relating to the L2 listening pedagogy. The second focus for L2 listening teaching, word recognition, has received far less research attention.

2.4.2.1 Application of listening strategies

The predominant emphasis in the relevant research literature is the role of non-linguistic knowledge or top-down processing in facilitating L2 listening success (Macaro, et al., 2007) through implementation of L2 listening strategies. Recurring elements of strategy instruction include encouraging listeners to make predictions about the content of a listening text, selectively attending to salient elements using inferences to determine unknown elements of the speech signal and monitoring levels of comprehension (Goh, 2002; Graham & Macaro, 2008; O'Malley, et al., 1989; Vandergrift, 2003a, 2003b, 2007; Vandergrift & Tafaghodtari, 2010).

Vandergrift (2004, 2007) recommends an approach which involves an integrated pedagogical cycle incorporating a number of phases aimed primarily at developing increased metacognitive awareness of listening processes. The cycle involves a prediction stage, three verification stages and a reflection stage. In the prediction stage students discuss the general topic of the listening text and predict elements of language and content that they expect to hear. In the first verification stage students listen to and verify the hypotheses they put forward initially. During this stage listeners refine their collective understanding of the components of the listening text that they need to attend to selectively in order to resolve uncertainties about the content of the listening text. During the second verification stage students selectively attend to problematic sections
of the listening text. In the last verification stages the students discuss what remains unclear regarding the content of the text and have the opportunity to compare the listening text to a written transcription of the text. In the reflection stage students are encouraged to reflect on the strategies they used during the listening activity and formalise goals for the next listening activity.

Vandergrift’s (2007) pedagogical cycle is informative in that it is representative of the current importance placed on the two components of listening instruction recommended by Graham, et al., (2010) namely the application of strategy instruction and the development of word and phrase recognition. As with other researchers primarily interested in the role of the application of strategies in L2 listening pedagogical approaches (Cross, 2009; Goh, 2002, 2008; Goh & Taib, 2006; Graham & Macaro, 2008; Graham, et al., 2008; Vandergrift, 1999, 2003a, 2003b) a strong emphasis on top-down, non-linguistic knowledge is evident in the approach recommended and research paradigms adopted.

2.4.2.2 Developing L2 WRS

Despite empirical evidence which suggests that lower order listening skills, such as word recognition, are of primary importance in skilled L2 listening comprehension, the SLA research devoted to word recognition is relatively scarce (Broersma & Cutler, 2008; Field, 2008a, 2008c; Goh, 2000; Graham, et al., 2010; Tsui & Fullilove, 1998; Yi’an, 1998). Field (2008d) asserts that until recently a received view persisted among researchers and practitioners that L2 decoding problems could be compensated for through the application of top-down listening strategies.

In response to renewed appreciation of the importance of perceptual processing as a central component of successful L2 listening (Field, 2008c) and a perceived overemphasis on top-down listening skills in L2 pedagogy (Wilson, 2003), calls have been made for greater emphasis on lower order listening processes in L2 listening pedagogy, particularly word recognition (Al-jasser, 2008; Field, 1998, 2003, 2008a; Hulstijn, 2003; Wilson, 2003). With the assertion that decoding processes are of primary importance in L2 listening proficiency comes an imperative to establish pedagogical approaches that assist learners to become more accurate and automatic in their L2 speech decoding capabilities (Lynch, 2006). Specific guidelines on teaching
these decoding processes have been put forward (Field, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003). Although these recommendations have yet to be broadly implemented and validated in real language learning contexts, they do possess a number of common features.

Having identified the importance of L2 word recognition from speech, the challenge is to devise and evaluate methods to develop this language construct in a manner that is suitable for use in real language learning contexts. The existing orthodoxy of listening instruction emphasises global listening skills and depends strongly on the use of contextual information. Therefore, there is a need to complement these existing approaches with learning approaches which adequately emphasise the development of decoding skills. The pedagogical task at hand is to devise approaches which attune the L2 listener’s perceptual sensitivity to the target language speech such that spoken words can be more accurately and rapidly decoded.

2.4.2.3 Synchronous bimodal input to support L2 listening

Learning opportunities that facilitate association of the transient speech signal with the corresponding orthographic form hold strong potential for the development of L2 word recognition from speech (Bird & Williams, 2002; Field, 2008b; Goh, 2000; Hulstijn, 2003; Mitterer & McQueen, 2009; Wilson, 2003). The written form of words provides an explicit linguistic framework which clearly indicates the form of the composite words present in speech. This type of explicit framework would appear to be particularly valuable to L2 listeners in light of the rapid, variable and blended nature of fluent speech. Several studies have found that supplementing L2 speech with the written form functions to improve links between knowledge of aural and written forms of language (Borras & Lafayette, 1994) and perception of L2 speech (Bird & Williams, 2002; Mitterer & McQueen, 2009).

Most previous studies considering the benefits of bimodal input in language learning have investigated the assessment of these improvements in relation to global L2 listening comprehension (Baltova, 1999; Huang & Eskey, 1999-2000; Jones & Plass, 2002). For example, Guichon & McLornan (2008) sought to determine the effect of input modality of stimulus material on L2 learner global comprehension. Participants (N = 40) of the same proficiency level were divided into alternative treatment groups.
Treatment groups were asked to engage with the same basic stimulus material presented in a variety of combinations of modality: audio only, video with audio, video with audio and L2 subtitles, and video and audio with L1 subtitles. Participants were asked to demonstrate their comprehension of the text by writing a twenty minute summary of the content of the stimulus material. Summaries were then scored to determine comprehension level for each student by counting semantic units in the students’ written summaries which concurred with the content of the stimulus material. Results indicated that the presentation of the stimulus material in a form which included written subtitles was associated with greater levels of text comprehension. Such results provide empirical support for the pedagogical value of the use of bimodal input in the support of global comprehension of spoken texts. However, the degree to which the results from these studies apply specifically to the development L2 word recognition from speech is unclear for a number of reasons. In the Guichon and McLornan (2008) study, and others like it, both aural and textual modalities are made available to learners as a means to support comprehension. Under such conditions it is unclear as to whether the presentation of subtitles facilitated increased comprehension of speech or whether the learners established comprehension of the text by compensating for limitations in listening capabilities by predominantly processing information via the written form. Under these bimodal treatment conditions, participants are theoretically able to totally ignore information in the aural modality, and use the information in the written (or visual form) to establish comprehension of the stimulus material (Bird & Williams, 2002), this renders such research designs ill-equipped to provide clear information regarding the efficacy of bimodal input in the development of word recognition from speech.

Mitterer and McQueen’s (2009) study provides stronger evidence for the value of bimodal input in the development of L2 word recognition from speech. This study aimed to determine the relative effectiveness of L1 and L2 subtitles on the perceptual processing of L2 speech as evidenced by participants’ ability to repeat verbatim excerpts from stimulus spoken material. A comparison of the effect of audio only treatment and the audio and L2 subtitles treatment revealed that the presentation of L2 subtitles significantly enhanced the ability to correctly repeat words. Mitterer and McQueen (2009, page 2) assert that “printed English words can provide an additional
source of information about words being spoken, and hence about the sounds being heard.”

2.4.2.4 Text reconstruction and comparison to improve L2 WRS

Several researchers have suggested the value of asynchronous presentation of aural and intra-lingual subtitles as a means by which to develop L2 word recognition from speech (Field, 2003, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003). These involve initially presenting the target text in the aural modality alone. Listening to sections of the aural text on multiple occasions provides an opportunity for the learner to reconstruct the content of the text and thereby engage in decoding processes at the word level. After ample opportunity has been provided for learners to process the target language in the aural modality, learners use the target text transcript as a means by which to self-diagnose errors in perception. These methodological steps circumvent the issue associated with synchronous bimodal presentation which may result in learners attending primarily to written subtitiles at the expense of aural processing.

Field (1998, 2004a; 2008b) has consistently advocated the value of asynchronous presentation of first speech and then the presentation of the written transcript as an efficient means to develop lower-level listening skills such as L2 WRS. Field (2008b) recommends that in the early stages of L2 acquisition a significant proportion of a learner’s effort should be directed towards the improvement of the automaticity of lower-order decoding routines. Field (2003) advocates the use of transcripts as a means by which learners can improve their ability to segment linguistic information from the aural modality at the word level. Field (2008a) asserts that a variation of the traditional dictation methodology presents the most effective way of raising listeners’ awareness of their decoding problems. Field (2008a) recommends that listeners are given opportunities to listen repeatedly to segments of connected speech and actively engage in attempting to recognise the composite words of the speech signal. These recommendations stress the importance of pedagogical approaches which foster a critical approach to learner decoding where listeners “construct and carry forward provisional interpretations” (Field, 2008a, page 49). Field’s recommendations are clearly focused on the development of the L2 decoding processes fundamental to the ability to perceive and map speech into its composite words.
Wilson (2003) and Hulstijn (2003) independently identify the pedagogical value of approaches which enhance the learner’s capability to decode words from speech and each recommend very similar pedagogical approaches which largely concord with those put forward by Field (2008b). Wilson (2003) advocates dictogloss (Wajnryb, 1990) or discovery listening, a method tailored specifically to provide opportunities to practice recognising the words in speech, but also to facilitate attention to meaning. This approach consists of three stages: listening, reconstructing and discovering. In the first stage learners listen to a short sample of connected speech, first without taking notes and then a second and third time noting down the lexical content of the speech. In the second stage students work in small groups in an effort to reconstruct the sample of speech that they have heard. Stage three involves students critically assessing their word recognition by comparing their reconstruction with the transcript of the target speech. The final stage also involves listening to the target speech again without looking at the transcript. Again the broad pedagogical theme of using the stable form of the written transcript to facilitate diagnosis of decoding errors is evident within these recommendations.

Hulstijn (2003) also strongly emphasizes the importance of learning activities which attune the L2 listener’s perception to the phonological form of words and thus improve the automaticity of word recognition. Hulstijn recommends that L2 learners wishing to improve automaticity of word by word understanding of the L2 target speech should listen to short samples of speech input and reconstruct the composite words of the speech sample. Hulstijn’s recommendations involve: 1) repeated listening, 2) mental reconstruction of the lexis of the utterance, 3) multiple replays of the utterance to check perception, 4) use of the written transcript of the utterance in order to check perception, 5) analysis of errors in perception, and 6) listen repeatedly in order to understand the utterance without the aid of the transcript. Central to these recommendations is the assumed value of repeated opportunities for listeners to perceive the composite words of an utterance and the use of the written transcript to enable listeners to make explicit links between the phonological and orthographic form of words.

The pedagogical recommendations put forward (Field, 2008a; Hulstijn, 2003; Wilson, 2003) are those which most clearly articulate a pedagogical framework directly aimed at improving the aural word recognition skills of second language learners. A general
synthesis of these recommendations can be summarised in the following pedagogical cycle. A learner repeatedly listens to a section of speech input attempting to reconstruct the text as fully as possible. Presenting the opportunity to repeatedly listen emphasises the assumed importance of frequency of exposure to target language in the aural modality as a key element to improved L2 WRS. This repeated exposure to the speech input enables students to revise hypothesised content of the text and furthermore provides scaffolding for the cognitively demanding task of text reconstruction or transcription. Recommendations to engage in reconstruction of the target speech reflect an assumed importance of the generation of language output in the improvement of L2 WRS. After listening to the aural input an adequate number of times, and attempting to reconstruct the text to the best of their ability, listeners compare their attempt at reconstruction and the written transcript of the speech. This comparison between the learner’s attempt at reconstructing the spoken text into the written form and orthographic form of the spoken text enables students to analyse and correct errors in perception. Importantly, this provides an opportunity for listeners to notice discrepancies between their language production and the target language. This feature of the recommended pedagogy is congruent with Schmidt’s (1990) noticing hypothesis which asserts that drawing attention to input is a vital component to second language acquisition. The final stage of this pedagogical cycle reflects the assumed importance of engaging explicit knowledge, in the form of feedback on performance, in noticing errors in perception and therefore improving L2 WRS.

This general pedagogical approach outlined, which involves: repeated opportunities to listen, opportunities to reconstruct the text and finally opportunities to receive explicit feedback on performance through use of the written transcript of the target text, will be referred to as text reconstruction and comparison (TRC). Although the principles of TRC have been clearly articulated in peer reviewed academic literature (Field, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003), to date the validity of these recommendations have not been the subject of extensive empirical studies in real language learning contexts. In light of the finding that the development of improved word recognition from speech should make up a significant component of L2 listening practice (Graham, et al., 2010; Hulstijn, 2003) an absence of research which investigates the efficacy of recommended approaches to improved L2 WRS represents a significant gap in the
existing literature. Refining and evaluating approaches for the development of L2 WRS is the primary objective of the current research.

2.5 The value of CALL in the development of L2 WRS

As has been outlined the intrinsic attributes of the speech signal make recognising words from speech for L2 learners a challenging prospect. The difficulty experienced by students in regards to L2 WRS presents a significant pedagogical challenge in real language learning contexts. From an emergentist perspective on language acquisition, the development of L2 WRS will in large part depend on providing learners with increased opportunities to be meaningfully engaged with the target spoken language through intensive practice (Hulstijn, 2003). Suggestions that intensive listening plays an important role in the development of word recognition are in line with a SLA consensus that input plays a fundamental role in L2 acquisition (Boyd & Goldberg, 2009; Ellis & Collins, 2009; Ellis & Larsen-Freeeman, 2009). However, ideally there should also be mechanisms put in place which assist to alleviate the cognitive load imposed on the L2 listener while listening to the target speech. The importance of exposure to input and language usage in the development of the L2 WRS must also be reconciled with the key finding that awareness raising mechanisms are consistently shown to play a fundamental role in L2 learning (Norris & Ortega, 2000; Schmidt, 1990). The engagement of explicit knowledge, which is characterised by effortful, conscious and controlled processes, can increase the rate of language acquisition (Ellis, 2002; R. Ellis, 2004). Learning approaches, such as the generation of output and the facilitation of noticing errors in perception, which can be used to focus the learner’s attention to learnable linguistic features are associated with second language acquisition (Izumi, 2002; Schmidt, 1990). This would even appear to be the case for processes which do not lend themselves strongly to conscious monitoring such as word recognition from speech (Al-jasser, 2008; Hulstijn, 2003).

Computer assisted language learning (CALL) provides a pedagogical mechanism which enables learners to increase their exposure to target language in the aural modality and to also be provided with “cognitive scaffolding” which assists in the difficult task of word recognition. The computer is a highly integrated device which can be used to
strategically undertake multiple tasks and present language in a number of modalities (Al-Seghayer, 2001). The key advantage of CALL is the potential it holds to create opportunities for extensive listening opportunities through the delivery of digital sound files (Robin, 2007) with the capability to facilitate awareness raising mechanisms such as opportunities to produce output and the presentation of automated corrective feedback. Such language learning environments are difficult or impossible to create in traditional non-CALL learning contexts.

CALL also has the potential to facilitate the development of L2 WRS by presenting linguistic data in more than one modality. As has been noted, a recurring theme in recommendations for improving word recognition from speech is the association of the phonological form of words with their corresponding orthographic form (Field, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003). CALL has the capability to present extensive opportunities to listen and respond to spoken language as well as to present the orthographic form of the spoken words at the appropriate phase of the learning process. CALL presents a means by which to establish learning contexts within which the ephemeral spoken word form can be meaningfully associated with the corresponding orthographic form of the word. CALL currently has the capabilities to implement the TRC learning protocol for the development of L2 WRS.

A key advantage of web-based CALL is that it has the potential to expand the scope and timing of learning experiences for language learners. For example, web-based CALL provides a means by which purpose designed language learning tools can be made available to widely dispersed cohorts of learners in a targeted manner.

2.5.1 Previous use of CALL to develop L2 WRS

Although, the value of using computers to enhance L2 learner’s listening skills has long been acknowledged (Brett, 1995; Grezel & Sciarone, 1994), the majority of the studies which have addressed the role of computers in L2 listening have focused on the development of global listening comprehension (Absalom & Rizzi, 2008; Grgurović & Hegelheimer, 2007; Guichon & McLornan, 2008; Jones & Plass, 2002; Smidt & Hegelheimer, 2004). Despite digital technology being particularly well suited to the development of perception skills (Vandergrift, 2007), only a few previous studies have
addressed the value of computers for the purposes of developing lower-level listening skills such as L2 WRS.

Hulstijn (2003) has clearly articulated the value of CALL for the development of L2 WRS. Central to these recommendations is the ability for computer mediated environments to present multimodal information according to designer’s specifications. Hulstijn describes LISTEN123, a computer program with a number of functions including the presentation of speech which can be played on multiple occasions before the written transcript of the spoken text is presented to the learner. Notwithstanding an absence of published empirical data verifying the effectiveness of these recommendations in real language learning contexts, Hulstijn’s (2003) recommendations have been cited by preeminent authors in the fields of L2 listening (Vandergrift, 2007) and CALL (Chapelle & Jamieson, 2008; Levy, 2009) as a pedagogical model of best practice for the development of L2 WRS.

Previously, Grezel and Sciarone (1994) investigated a computer test which was implemented in an effort to encourage learners to develop better study skills beyond the classroom. Specifically, the computer test was based on the recorded material which made up a central component of an intensive Dutch as a Second Language learning course. Computers were used to deliver several sentences of spoken text in stages to individual participants. After listening to each sentence the students were required to transcribe what they heard in the sound files. Learners were only able to hear the target sound file section twice before they were asked to reconstruct the text, unlike the recommendations for the development of L2 WRS previously discussed and described as Test Reconstruction and Comparison (TRC) (Field, 2008b; Hulstijn, 2003; Wilson, 2003) which suggested multiple opportunities for learners to listen to the target text. Participants in Grezel and Sciarone (1994) were only given an opportunity to compare their reconstructions and the correct target text sentences after achieving a predetermined transcription accuracy score. Grezel and Sciarone (1994) compared the relative performance of an experimental group (N = 154) which used the computer based listening test and a control group (N = 198) which undertook daily cloze exercises instead. At the end of the course control and experimental groups undertook a summative cloze test and it was determined that the average performance by the experimental group was significantly higher than that of the control group. Grezel and
Sciarone conclude that regular testing of listening through the manner described is an effective means by which to develop L2 acquisition.

2.6 Conclusions and future directions

Models of the cognitive processes which occur during L2 listening comprehension have exerted an influence on the most recent recommendations on how to teach L2 listening (Goh, 2000; Graham, et al., 2010; Hulstijn, 2003). Far from being a passive skill, L2 listening is now recognised as a skill which needs to be actively developed in the language classroom (Field, 2008b; Rost, 2002; Vandergrift, 2007). Investigations into the attributes of less and more skilled L2 listeners have resulted in a recognition that skilled L2 listening is a process which depends on the interaction between lower-level decoding skills and higher order meaning building processes (Buck, 2001, Field, 2008b). The importance of the fluid interaction between these processes has resulted in a current trend in teaching listening which emphasises the importance of both automaticity in decoding skills such as word recognition and development and engagement of listening strategies (Graham, et al., 2010). However, to date the L2 listening research efforts have been primarily focussed towards the use of listening strategies as a way to develop L2 listening comprehension. Historically, lower order decoding skills such as word recognition have been undervalued in L2 listening development (Field, 2008d). However, empirical evidence suggests that lower-level processes such as word recognition from speech are of primary importance in skilled L2 listening (Tsui & Fullilove, 1998). The primacy of lower-level processes in listening has led to suggestions that L2 listeners must possess the ability to recognise a minimum threshold of words in speech before listening strategies can be successfully applied in L2 listening (Graham, et al., 2010). Indeed, it has been shown that a compensatory dependence on top-down L2 listening strategies due to a lack of word recognition can be the source of L2 listening problems (Field, 2008a).

Empirical evidence which points to the importance of L2 WRS in skilled L2 listening should encourage research efforts which develop and evaluate approaches which aim to improve L2 learners’ L2 WRS capabilities. Surprisingly however, very little research has been undertaken which has focussed on the construct of L2 WRS and how this essential component of L2 listening can be effectively developed (Broersma & Cutler,
As described, theoretically and empirically supported recommendations on how to develop L2 word recognition from speech have been articulated in relevant academic literature (Field, 2008a; Hulstijn, 2003; Wilson, 2003). However, extensive reviews of peer reviewed literature indicate that there is a paucity of research which investigates the effectiveness of these recommendations in real language learning contexts. Of central interest here is to investigate the effectiveness of providing learners the opportunity to engage in the general pedagogical approach described here as Text Reconstruction and Comparison (TRC). The TRC approach involves learners having self-determined and multiple opportunities to listen to samples of contextualised target speech, opportunities to reconstruct the text in written form, and finally receiving feedback on performance in reconstruction through presentation of the transcript text.

Furthermore, as the development of L2 WRS requires extensive practice, approaches to develop L2 WRS which are able to be applied in student-centred, out-of-class contexts are of strong interest (Field, 2008b; Hulstijn, 2003). Due to the technological capability of CALL to facilitate learning protocols such as the TRC approach, computer-mediated approaches to the development of L2 WRS potentially have an important role to play in assisting L2 learning to develop better L2 WRS. For this reason investigating and evaluating CALL for the development of L2 WRS is of central interest to this research.

Despite the importance of L2 word recognition and the potential for computers to assist in the development of this demanding skill (Hulstijn, 2003), CALL methods specifically aimed at improving L2 WRS have yet to be adequately described, analysed and evaluated in real language learning contexts. The body of this thesis contains a series of research papers each of which contribute to filling this current gap in knowledge.
Chapter 3

Preamble to research paper 1: “Recognition of high frequency words from speech as a predictor of L2 listening comprehension”

An important theoretical assertion which has been developed in the literature review of this thesis is that word recognition from speech plays a fundamentally important role in L2 listening comprehension. Theoretical accounts of the processes involved in L2 listening (see section 2.3) as well as empirical accounts of the source of listening problems and success (see sections 2.3.3 and 2.3.4) provide broad support for the validity of this assertion. However, as the construct of L2 WRS is the central focus of this thesis, it is appropriate to independently present empirical data which quantifies the relationship between L2 WRS and L2 listening comprehension.

Research paper 1, which is presented in Chapter 3, addresses the primary research question, “Does second language word recognition from speech (L2 WRS) have sufficient influence on second language listening comprehension to make research into the development of L2 WRS a productive enterprise?” Thus, Chapter 3 seeks to present empirical data which validates the construct of L2 WRS as one which is of importance in L2 listening comprehension and thus also to L2 language learning generally.
Research paper 1: Recognition of high frequency words from speech as a predictor of L2 listening comprehension


The content presented in this chapter is a minimally adapted version of the article which appears in the journal *System*. Permission is granted from Elsevier to use the content in the manner it is presented here.

Abstract

This paper investigates the relationship between recognition of high frequency words from speech and second language (L2) listening comprehension among 167 tertiary level Chinese learners of English. It also interrogates the extent to which the ability to recognise words from speech contributes to the prediction of L2 listening comprehension scores. Word recognition from speech (WRS) was assessed with a partial dictation test which targeted high frequency vocabulary. These target words were categorised as belonging to either the first, second or third thousand word frequency levels through comparison with the British National Corpus and the Corpus of Contemporary American English (BNC/COCA) word family lists. L2 listening comprehension was assessed with a version of the International English Language Testing System (IELTS). Multiple regression analysis revealed that recognition of words from the third thousand frequency level alone could predict 52% of the variance observed in the listening comprehension scores. Recognition scores for words below the third thousand frequency range added very little unique predictive power to the regression model. This was the case despite word recognition scores for the first, second and third thousand frequency ranges strongly correlating with listening comprehension scores. Findings suggest the ability to recognise high frequency words from speech is predictive of the aural modality specific word knowledge indicative of successful L2 listening comprehension. Pedagogical implications and applications are discussed.

Keywords: Listening comprehension; Word recognition from speech; High frequency vocabulary; Multiple Regression Analysis
3.1 Introduction

Second language (L2) listening is a fundamental component of L2 learning. Proficient listening comprehension enables learners to understand the spoken discourse of the target language which in turn aids the development of other language skills (Dunkel, 1991; Rost, 2002). The skill of listening is also of strong contemporary significance to L2 learners as it enables engagement with a vast range of online spoken target language samples such as those from video sharing websites and digital audio/video on demand systems (Robin, 2007). Despite the centrality of listening comprehension in L2 learning and the huge range of listening materials available, of the four main language skills, listening comprehension remains arguably the least well understood and researched (Vandergrift, 2007). Listening comprehension is now becoming a more prominent area in L2 teaching and testing (Buck, 2001; Cai, 2013; Field, 2008a, 2008b; Song, 2008). Of significance in relation to this development has been an increased research effort directed towards understanding the processes which underpin successful L2 listening (Field, 2008a, 2008b; Vandergrift, 2007). This emerging picture of the specific knowledge types which support skilled L2 listening provides a useful framework for further advancements in L2 listening research, teaching and testing.

One domain of knowledge which is an important component of skilled L2 listening comprehension is linguistic knowledge at the word level (Graham, et al., 2010). Previous studies have confirmed the value of various constructs of word knowledge in supporting and predicting L2 learners’ listening comprehension proficiency level (Bonk, 2000; Stæhr, 2008, 2009). However, as has been acknowledged within the field of L2 vocabulary development, L2 word knowledge is not a unitary construct (Daller, Milton, & Treffers-Daller, 2007; Henriksen, 1999; Nation, 2001; Qian, 2002). There are multiple aspects of knowing a word, with evidence indicating that each has a differing importance in explaining various aspects of L2 language proficiency (Milton, Wade, & Hopkins, 2010; Stæhr, 2008). Arguably, the construct of word knowledge which is of strongest importance in successful L2 listening comprehension is the ability to recognise words from speech (Field, 2008b; Hulstijn, 2003; Rost, 2002). Word recognition from speech (WRS) is here defined as the ability to map information from the speech signal onto the lexical units that information represents. WRS enables establishment of links between the formal and semantic attributes of known words.
(Hulstijn, 2002) which in turn enables fluid application of both linguistic and non-
linguistic knowledge during the listening process (Graham, et al., 2010). If WRS is as
important to skilled L2 listening as has been suggested by previous research (Field,
2008a, 2008b; Goh, 2000; Hulstijn, 2003; Tsui & Fullilove, 1998), then this construct of
word knowledge should be of direct interest to language teachers. However to date,
research which specifically explores the relationship between WRS and L2 listening
comprehension is scarce (Broersma & Cutler, 2008). This research seeks to contribute
towards filling this current gap in knowledge.

3.2 Literature review

3.2.1 Models of L2 listening comprehension

Many of the challenges associated with teaching and testing listening comprehension
relate to the observation that “the product of listening comprehension is a construction
or representation of meaning in the mind” (Buck, 2001, page 99). This covert attribute
of listening comprehension makes adequate description and subsequent development of
constructs of listening comprehension particularly challenging (Vandergrift, 2007).
However, a number of models of listening comprehension are widely accepted and
provide a valid point of departure for ongoing listening comprehension research.

Bachman and Palmer (1996) use language competence and strategic competence to
broadly describe the competencies required for successful listening comprehension.
Language competence relates to the implicit and explicit forms of language knowledge
that can be applied while listening. Strategic competence relates to the executive
processes, such as metacognitive strategies, which facilitate effective application of
language competence (Buck, 2001). The role of linguistic and non-linguistic
knowledge, and theoretical positions on how these forms of knowledge interact, provide
the basis of contemporary descriptions of L2 listening comprehension. The widespread
application of information processing models which use both top-down and bottom-up
components to conceptualise L2 listening is evidence of this influence (Bonk, 2000;
Cai, 2013; Goh, 2000; Graham, et al., 2010; Vandergrift, 2007; Wilson, 2003). Bottom-
up processes consist of “speech perception and word recognition” and provide the
linguistic input needed for comprehension. Top-down processes are primarily
associated with the use of non-linguistic knowledge such as the application of “semantic expectations and generalisations” (Rost, 2002, page 96). It is broadly accepted that successful listening comprehension depends on the simultaneous application of linguistic and non-linguistic knowledge while processing the incoming speech signal (Buck, 2001).

The exact manner by which bottom-up and top-down information processes interact and the relative contribution these processing types play in successful listening comprehension has been the focus of previous research effort and debate (Field, 2008b; Goh, 2000; Hasan, 2000; Osada, 2001; Tsui & Fullilove, 1998; Wilson, 2003). The value of applying non-linguistic knowledge, in particular metacognitive strategies, has been of central interest in previous L2 listening comprehension research (Macaro, Graham, & Vanderplank, 2007). Another component of the research effort has been directed towards investigating the relative importance of non-linguistic and linguistic knowledge in listening comprehension, with findings suggesting that lower order linguistic competence plays a primary role in L2 listening comprehension.

An early study by Tsui and Fullilove (1998) sought to determine whether bottom-up or top-down processing was more important in determining higher L2 listening comprehension test performance. The study involved approximately 20,000 students of English as a second language (ESL) who were undertaking public examinations in Hong Kong. In order to address the research question, two question types from the exam were investigated: matching schema and non-matching schema types. In matching schema test items, the listeners who formed hypotheses based on the initial input could primarily use top-down processes to select the correct answer. Non-matching schema test items required candidates to process spoken linguistic input which ran counter to the initial input in order to select the correct answer. Analysis indicated that candidates who correctly answered the non-matching question types were the more skilled test takers. Yi’an’s (1998) smaller scale study also suggests lower order linguistic processing is primary in L2 listening comprehension. Ten native Chinese speaking ESL students listened to an English language radio interview that was the stimulus for multiple choice questions and retrospective verbalisations. Yi’an was interested in determining the manner by which linguistic and non-linguistic knowledge was engaged during completion of the multiple choice questions. Analysis of the verbalisation recalls
indicated that partial linguistic decoding of input led to engagement of non-linguistic knowledge to compensate for gaps in understanding and that such use of compensatory strategies did not guarantee selection of the correct answer. Partial linguistic decoding and the resultant dependence on non-linguistic information, often led the portion of the input which had been successfully decoded to be overridden.

The importance of lower order linguistic knowledge has also been highlighted by Field (2008a) who employed a variation of the gating research paradigm to investigate the manner by which native and non-native listeners perceived a series of initially ambiguous spoken utterances. For example, [ˈiːtən] was used as an ambiguous stem for either “eaten up”, “eat an egg” or “eat enough” (Field, 2008a, page 42). Eight utterances were edited such that each started out as ambiguous, but as more gates were reached the ambiguity of the linguistic information encoded in the signal reduced. Native and non-native speakers listened to utterances in four gates, after each gate the participants were required to transcribe what they had heard. A key finding was the high degree of accuracy with which native listeners were able to recast their hypotheses about the segmentation of the words present in the speech signal as the utterances became linguistically unambiguous. Non-native listeners were far less flexible than L1 listeners in their ability to reappraise their segmentation in the instance where incoming linguistic input suggested an error in interpretation was present. These findings complement the studies of Tsui and Fullilove (1998) and Yi’an (1998) in that they highlight adequate levels of speech decoding and word recognition as an attribute of expert listening.

We assert that a useful model of L2 listening comprehension necessarily incorporates the role of both non-linguistic and linguistic knowledge. However, it would appear that “a minimum level of vocabulary recognition is required before non-linguistic knowledge in the form of prior or world knowledge or higher order reasoning can be brought into play effectively” (Graham, et al., 2010, page 14). A theoretical position held here is that word recognition from speech plays a fundamental role in a listener’s ability to apply non-linguistic and linguistic knowledge while processing spoken language. As such, constructs of word recognition from speech are likely to be valuable targets for tests which aim to predict a listener’s ability to deal with L2 listening comprehension tasks.
3.2.2 The application of word knowledge in the prediction of L2 listening comprehension

Vocabulary knowledge is a reliable predictor of various dimensions of L2 proficiency, including the sub-skills of reading, writing, speaking and listening (Meara, 1996; Milton, et al., 2010; Nation, 2001; Stæhr, 2008). Although links have been previously demonstrated to exist between L2 word knowledge and L2 listening comprehension, this relationship is less well researched than the relationship between word knowledge and L2 reading (Qian, 1999, 2002).

Stæhr (2009) investigated the role of vocabulary knowledge, including vocabulary breadth, on listening comprehension among a group of 115 advanced Danish learners of English as a foreign language (EFL). Vocabulary Levels Tests (Nation, 2001; Schmitt, Schmitt, & Clapham, 2001) were used to assess the vocabulary breadth of participants with thirty target words from each of the second, third, fifth and tenth thousand frequency ranges being measured. Results showed a strong correlation between vocabulary size and listening comprehension \( (r = .70, p < .01) \). Multiple regression analysis indicated that measures of vocabulary size were able to predict 49% of the variance observed in the scores for listening comprehension. These results clearly indicate that for the group involved in the study, the construct of vocabulary size was an important predictive variable for listening comprehension. However, a strong limitation of the study was that the vocabulary knowledge tests used were delivered solely in the written modality. Such test formats do not directly tap the ability to recognise and apply word knowledge from the aural modality under the time constraints typically imposed upon listeners processing spoken language in real time. As was noted by Stæhr, research aiming to clearly understand the relationship between vocabulary knowledge and listening comprehension should also measure constructs of vocabulary knowledge which tap the ability to recognise the phonological form of the word. This underrepresentation may be particularly significant in light of the difficulty L2 learners experience in recognising known lexis when presented in the aural modality (Goh, 2000).

Milton, Wade & Hopkins (2010) investigated the contribution of word knowledge, including phonological vocabulary size, on scores obtained by 30 EFL students on a number of language skills, including listening comprehension. Computerised yes/no
tests were administered to measure both orthographic and phonological vocabulary size and these results were correlated with the participants’ sub-skill scores from IELTS. Twenty test items for each of the first, second, third, fourth and fifth word frequency levels were used to measure word knowledge. The relatively modest correlation observed between the measures of orthographic and phonological vocabulary suggested that the equivalent forms of the vocabulary tests delivered in each modality did tap distinct constructs. Phonological vocabulary size and listening scores (Spearman’s $\rho = .67$, $p < .05$) were more strongly correlated than orthographic vocabulary size and listening scores. These findings emphasise that word knowledge is modality specific and also suggest that measures of word knowledge mediated in the aural modality are of strong value in the prediction of L2 listening comprehension. However, it must be noted that the construct of phonological word knowledge used in the study did not measure the tests takers’ ability to produce the target words. The highly receptive nature of this test format may therefore underrepresent the depth of knowledge needed for active recognition and application of word knowledge typical of that needed when processing spoken language in real time. An additional limitation of the study was the relatively small sample size of thirty participants.

Bonk (2000) sought to establish the relationship between lexical familiarity and listening comprehension evident among 59 Japanese EFL learners. Four short L2 spoken texts were used as the stimulus for first language recall protocols which were rated on a four-point scale and used as the measure of listening comprehension. The same four texts were used as the stimulus for dictation tests which were used as the measure for lexical familiarity. Listening comprehension was shown to correlate moderately yet significantly with lexical familiarity ($\tau = .446$, $p < .05$). Results indicated that those participants scoring less than 80% for lexical familiarity were unlikely to achieve good levels of listening comprehension. However, other results from the study demonstrate the complexity of the relationship between word knowledge and listening comprehension. For example, although the dictation scores of 14 of the participants indicated familiarity with up to 100% of the words in the text, these same participants’ were unable to attain a good listening comprehension rating. In contrast, it was also noted that some participants were able to attain good comprehension ratings despite lexical familiarity measures of 75% or less. It must be noted however that the method by which lexical familiarity was measured presents some methodological
issues. Dictation tests, especially of the form used in the study which involved the stimulus material being heard multiple times, tap a range of linguistic and non-linguistic knowledge sources (Buck, 2001). As a result of these issues, questions remain as to the validity of the use of this form of dictation as a direct measure of vocabulary knowledge (van Zeeland & Schmitt, 2013). Despite these methodological concerns, Bonk’s study provides a useful context for the current work. This is the case as the study not only highlights that word knowledge is an important element of listening comprehension, but also that the exact relationship between lexical knowledge and listening comprehension is equivocal, and thus indicates the need for research which further investigates the nature of this relationship.

3.2.3 The present study

This study follows the general methodological approach applied in previous research which has investigated specific constructs of word knowledge in predicting language skills such as reading (Qian, 2002) and listening (Stæhr, 2009). However, the current study differs from the three previous studies reviewed in section 3.2.2 in a number of respects. Firstly, here we focus on a construct of productive WRS which has not been addressed in previous related research. This construct represents the ability to recognise the phonological form of the word, access existing knowledge of that word and produce a representation of it under time constraints. This construct of word knowledge is unique in comparison to those operationalised in the studies reviewed in section 3.2.2 as it reflects aural modality specific word knowledge and the automaticity with which those words can be accessed and produced. These features clearly differentiate it from the form of word knowledge measured by the test instruments used by Stæhr (2009) which involved a form-meaning mapping task undertaken exclusively in the written modality, and that used by Milton et al., (2010) which involved no productive demonstration of word knowledge. As described in additional detail in section 3.4.2.1 below, the measurement of WRS used in this study involves test takers having just one opportunity to listen to the stimulus material and also constrains the target word to a specific lexical item. This approach to measuring word knowledge reduces the methodological concerns associated with multiple opportunities to listen to aural stimulus evident in the approach used by Bonk (2000). Further, constraining the item
target to a single target word ensures more direct measurement of a test taker’s knowledge of a specific range of lexical items.

Secondly, in contrast to previous studies which have investigated word knowledge of a broader range of word frequency levels, here a narrower focus has been applied to include just high frequency words. This decision is based on previous studies which suggest that relatively good L2 listening comprehension is possible with a lexical coverage of spoken discourse at the 95% level which normally demands knowledge of between 2000 and 3000 of the most frequent word families (van Zeeland & Schmidt, 2013). Analysis of large corpuses, such as the spoken component of the British National Corpus (BNC), also suggests that the most frequent 3000 word families are likely to achieve or exceed the 95% coverage level (Adolphs & Schmitt, 2003; Nation, 2006). Additionally, analysis of the spoken language typically used in television and movies also suggests that the 3000 most frequent word families are sufficient to reach a 95% coverage level (Webb & Rodgers, 2009a, 2009b). Such findings suggest that the 3000 most frequent word families are of strongest importance in providing a listener with the basic competence needed to access most spoken texts. In contrast, acceptable levels of reading comprehension are likely to require knowledge of the most frequently occurring 4,000-5,000 word families (Laufer & Ravenhorst-Kalovski, 2010). These findings suggest that knowledge of words beyond the most frequent 3000 word family level is less crucial for listening than for reading. Such findings provide impetus for research which investigates the relationship between word knowledge of the most frequent 3000 word families and L2 listening comprehension. The importance of this frequency range in facilitating basic language competencies, including listening comprehension, has led to recent reassessment of the traditional high frequency word cut-off point. This cut-off point has been extended from the 2000 to the 3000 word frequency level (Schmitt & Schmitt, 2012). It is the most frequent 3000 word families which are considered high frequency vocabulary in the present paper. It is assumed here that if knowledge of high frequency vocabulary underpins the basic competence required to process spoken discourse, then there is theoretical cause to suggest that the measurement of the ability to recognise words from the first 3000 frequency range will be a strong predictor of listening comprehension.
A final point of difference is that previous research has sought to determine the relative contribution of different constructs of word knowledge, such as receptive vocabulary breadth and receptive vocabulary depth, in predicting listening comprehension. In contrast, the present study confines its focus to the relative predictive power of various components of a single construct, namely WRS at the first thousand (1K), second thousand (2K) and third thousand (3K) word frequency levels.

3.3 Research Questions

This study aims to provide a clearer understanding of the relationship between various levels of WRS and listening comprehension. Further, we investigate the value of WRS as a predictor of listening comprehension. Lastly, we seek to align L2 learners’ varying abilities to recognise high frequency words from speech with standardised measures of their listening comprehension proficiency. This final goal is significant as we wish to provide frameworks with which the empirical data relating to L2 WRS attained as part of this investigation can be readily contextualised and applied in real language learning contexts. We do this in an effort to provide tangible links between the relatively specific domain of word knowledge of L2 WRS and the ultimate goal of listening instruction, namely facilitating an improved capability to comprehend L2 spoken discourse.

To this end the following research questions will be investigated:

1. To what extent does the ability to recognise high frequency words from speech correlate with listening comprehension scores?

2. To what extent does the ability to recognise words from the first, second and third thousand frequency level contribute to the prediction of L2 listening comprehension scores?

3. What is the comparative profile of word recognition from speech performance across the first, second and third thousand frequency levels for those participants attaining modest, competent and good listening comprehension scores?
3.4 Materials and Methods

3.4.1 Participants

The participants involved in this study were 167 students studying at a large university in the Peoples’ Republic of China. The group consisted of seven different classes with a range of 24 different majors. The total group consisted of 74 males and 88 females (5 missing values) with an average age of 19.42 years ($SD = 1.37$). All participants possessed Chinese as a first language and had a mean self-reported duration of English language study of 10.18 years ($SD = 2.57$).

3.4.2 Instruments

3.4.2.1 Measurement of word recognition from speech

The construct of WRS was assessed with an 89 item WRS test (see Appendix 3.1). The WRS test is a subtle adaption of the partial dictation test (Coniam, 1998; Cai, 2013; Hughes, 2003). As with partial dictation tests, WRS tests involve test takers listening to contextualised samples of spoken language in an effort to produce a target word present within the spoken input. The primary difference between partial dictation tests and WRS tests is that the latter are developed with a strong focus on ensuring the stimulus sentence within which the target word is embedded does not provide sufficient information for the test taker to systematically identify the target word without first hearing the stimulus sentence. This validation process involves first showing the written stems of the WRS tests to native speakers to ensure the target word cannot be immediately selected through use of written context alone. The second step involves checking to see if native speakers can readily recognise the target word when provided with the listening stimulus. This process ensures that the primary construct being measured is the test taker’s ability to accurately recognise the target words based on their phonological form in connected speech.

WRS tests were also appropriate for our purposes as they enabled selection and assessment of a tightly controlled range of vocabulary items, namely high frequency target words. WRS test items were selected from items previously piloted among similar groups of learners to those involved in the present study. Test items were demonstrated to have strong levels of internal consistency ($\alpha = .91$) and concurrent
validity (Matthews, Cheng & O’Toole, 2015). Each item had an accompanying section of written text with the position for the target word indicated with a blank space. Each item consisted of a single sentence with a single target word. These sentences were written by the first author and were elements of two test instruments which were used as part of a previous investigation which involved the measurement of L2 WRS (Matthews et al., 2015). In total, the WRS test used in the study consisted of 23 target words from the 1K frequency range, 37 words from the 2K frequency range and 29 words from the 3K frequency range. These 89 high frequency words were selected as they were target words present in the validated WRS test instruments previously used in the study mentioned above. These target words were also selected to ensure an adequate number of target words from each of the three word frequency levels of interest (1K, 2K and 3K) were present in the WRS test. The frequency range of the target words was categorised by comparison with the BNC / Corpus of Contemporary American English (COCA) word family lists using analytical tools available on the Compleat Lexical Tutor website (Cobb, 2014). The frequency level of the lexical content of the written contextual sentences for each item was carefully checked. For the items testing the 1K frequency level all contextual lexical items were also from the 1K frequency level and for the 2K and 3K level, contextual lexical items were predominantly from the 1K level with only a few items from the 2K level.

The spoken stimulus for the WRS tests consisted of general factual information relating to countries in the broader region within which the study was undertaken. This was deemed appropriate as the generic nature of the information was assumed to limit the potential influence of variation among participants’ background knowledge. The spoken stimulus for the WRS test was recorded with high quality digital recording equipment by a male, Australian native speaker of English. The total word length of the stimulus was 646 words with each contextual sentence having an average length of 7.26 words. The speech rate of the spoken stimulus was approximately 2.1 words per second and therefore was equivalent to that of the speech rates typical of a lecture delivered at average speech rate (Tauroza & Allison, 1990). There was a four second delay between the end of one test item and the commencement of the next test item. Test items were only heard once. The total duration of the test was just under 17 minutes.
3.4.2.2 Word recognition test scoring procedures

Word recognition from speech fundamentally occurs within the cognitive apparatus of the listener. The ability of a test taker to recognise a word from speech must therefore be inferred from the ability to complete tasks (Vandergrift, 2007). The operationalisation of such tasks necessarily represents a potential threat to the construct validity of any test instrument which aims to measure the ability to successfully process spoken language. In relation to the method of measurement of WRS used as part of this research, the test taker is required to write an orthographic representation of the recognised word. The primary concern was to manage the degree to which the act of writing the target word threatened the construct validity of a test designed to measure word recognition from speech. In order to do this, WRS test scores were assigned using a structured marking rubric designed to ensure responses were not penalised for minor spelling errors. Similar marking rubrics previously employed for the purpose of limiting the influence of spelling errors in WRS test scoring have produced a very high level of inter-rater reliability as determined by Cohen’s Kappa analysis ($\kappa = .91, p < .01$) (Matthews & O’Toole, 2015). Scoring of the partial dictation tests was undertaken by a research assistant who had been trained in the use of the marking rubric.

3.4.2.3 Measurement of listening comprehension

An IELTS listening test (IELTS Examination papers from University of Cambridge ESOL examinations, 2002), which had been sequestered since its publication for research purposes at the research site, was used to assess the participants’ L2 listening comprehension. The listening comprehension test consisted of 4 sections each with 10 questions. The content of the first two sections involved the application of listening comprehension in order to achieve a social function. The last two sections involved language use indicative of educational or training contexts. Section 1 and section 3 involved conversations between two people and section 2 and section 4 involved monologues. The test used included a number of question types, including sentence completion, short answer questions, multiple choice, note completion, table completion and labelling a diagram. Participants answered the questions as they heard the stimulus material. The listening material was heard only once. Test takers were provided 10 minutes at the end of the listening session to check their answers.
It was important to determine if any participants had previously undertaken this particular version of the IELTS listening test. In order to determine this, a single question: “Have you taken this listening test before?” was printed at the end of the IELTS listening test. All 167 participants indicated that they had not previously taken that particular version of the IELTS listening test.

There were several reasons for using this format of listening comprehension test. Firstly, the IELTS listening section is a standardised test with a format well known to test takers and as such had strong face validity. Secondly, IELTS tests have a widely disseminated performance band system associated with published cut-off scores. It was assumed that scores from this test would provide a useful context from which to present and communicate associations between listening comprehension and WRS test scores. Lastly, the IELTS listening test is a task-based language assessment (Stoynoff, 2009). The design of task-based assessment is premised on the assumption that measuring linguistic competence alone is not adequate to explain linguistic performance, and thus includes sociolinguistic competence, strategic competence and discourse competence as part of the test construct (Mislevy, Steinberg, & Almond, 2002). Task-based language assessment therefore assumes to tap both linguistic and non-linguistic knowledge types. In this regard, the theoretical underpinnings of task-based listening tests align strongly with the contemporary theoretical frameworks broadly used to describe listening comprehension (see section 3.2.1).

3.4.2.4 Lexical content of the listening comprehension test

Lexical content of the stimulus material for the listening comprehension test was analysed using Vocabprofile (Cobb, 2014) with the framework of analysis being the BNC-COCA word list. The content of the listening stimulus for the listening comprehension test was transcribed with all proper nouns removed from the text prior to analysis. As can be seen in Table 3.1, words up to and including the 3K frequency level exceeded the 95% coverage level for the stimulus material for the listening comprehension test. The fourth to tenth thousand frequency levels represent only 2.8% of the unique contribution to token coverage of the listening comprehension test stimulus.
3.4.3 Procedure

Test administration occurred in seven testing sessions involving class groups which ranged in size from 17 to 37. The two tests were administered to the participants in a single session which took less than 1 hour. Procedural protocols in regards to timing of tests, instructions to candidates, sequence of tests (WRS test then listening comprehension test) and test room conditions were kept standard between the testing sessions.

Table 3.1: Overview of lexical content of listening comprehension test

<table>
<thead>
<tr>
<th>Frequency level (K)</th>
<th>Word Families (%)</th>
<th>Types (%)</th>
<th>Tokens (%)</th>
<th>Cumulative token percentage (%)</th>
<th>Unique contribution to token coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>352 (65.77)</td>
<td>423 (68.78)</td>
<td>2013</td>
<td>87.37</td>
<td>87.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(87.37)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>95 (18.27)</td>
<td>103 (16.75)</td>
<td>153 (6.64)</td>
<td>94.01</td>
<td>6.64</td>
</tr>
<tr>
<td>3</td>
<td>38 (7.31)</td>
<td>44 (7.15)</td>
<td>58 (2.52)</td>
<td>96.53</td>
<td>2.52</td>
</tr>
<tr>
<td>4</td>
<td>13 (2.50)</td>
<td>13 (2.11)</td>
<td>15 (0.65)</td>
<td>97.18</td>
<td>0.65</td>
</tr>
<tr>
<td>5</td>
<td>9 (1.73)</td>
<td>9 (1.46)</td>
<td>10 (0.43)</td>
<td>97.61</td>
<td>0.43</td>
</tr>
<tr>
<td>6-10</td>
<td>19 (3.66)</td>
<td>18 (3.25)</td>
<td>20 (1.72)</td>
<td>99.33</td>
<td>1.72</td>
</tr>
</tbody>
</table>

3.5 Results

Table 3.2: Descriptive statistics and internal consistencies of tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Maximum possible score</th>
<th>Mean</th>
<th>SD</th>
<th>Mean (%)</th>
<th>SD</th>
<th>Reliability (Cronbach’s α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening comp. test</td>
<td>40</td>
<td>14.68</td>
<td>7.98</td>
<td>36.70</td>
<td>19.96</td>
<td>.87</td>
</tr>
<tr>
<td>WRS test</td>
<td>89</td>
<td>63.82</td>
<td>18.02</td>
<td>71.71</td>
<td>20.25</td>
<td>.91</td>
</tr>
</tbody>
</table>

Note: N = 167
As can be seen from Table 3.3, the general principle inherent to the validity of frequency based testing approaches that words occurring more frequently in the language are more likely to be known than those which occur less frequently, held true for this cohort.

As can be seen from Table 3.2, both test formats had acceptably high coefficients of internal consistency.

Table 3.3: Descriptive statistics for the subsections of the word recognition test

<table>
<thead>
<tr>
<th>WRS test section</th>
<th>Maximum possible score</th>
<th>Mean</th>
<th>SD</th>
<th>Mean (%)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1K</td>
<td>23</td>
<td>18.85</td>
<td>3.76</td>
<td>81.96</td>
<td>16.37</td>
</tr>
<tr>
<td>2K</td>
<td>37</td>
<td>27.37</td>
<td>8.09</td>
<td>73.98</td>
<td>21.86</td>
</tr>
<tr>
<td>3K</td>
<td>29</td>
<td>17.60</td>
<td>6.97</td>
<td>60.71</td>
<td>24.05</td>
</tr>
</tbody>
</table>

Note: N = 167

3.5.1 Research question 1: To what extent does the ability to recognise high frequency words from speech correlate with listening comprehension scores?

Table 3.4: Correlations between listening comprehension tests scores and WRS test scores for the three levels of word frequency and total WRS scores

<table>
<thead>
<tr>
<th></th>
<th>1K</th>
<th>2K</th>
<th>3K</th>
<th>TOTAL (1K, 2K and 3K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening comprehension</td>
<td>.67*</td>
<td>.69*</td>
<td>.72*</td>
<td>.73*</td>
</tr>
<tr>
<td>N</td>
<td>167</td>
<td>167</td>
<td>167</td>
<td>167</td>
</tr>
</tbody>
</table>

* p < .01 (two-tailed)

Pearson product-moment correlation coefficient indicated WRS scores at each frequency level, and also as a total score of all three frequency levels, correlated strongly and positively with listening comprehension test scores.
3.5.2 Research question 2: To what extent does the ability to recognise words from the first, second and third thousand frequency level contribute to the prediction of L2 listening comprehension scores?

Analyses to test that the data satisfied the assumptions of multiple regression were first undertaken. Of particular importance was to ensure that levels of collinearity of predictor variables used in analyses was not an issue. This was confirmed as the minimum tolerance value for predictor variables included in the regression models used (.24) exceeded the minimum acceptable level of .10 (Tabachnick & Fidell, 2001). After ensuring this and other assumptions regarding the data were satisfied, analysis proper was undertaken.

A step-wise multiple regression method was first employed to evaluate the degree to which the predictor variables of WRS test scores for the 1K, 2K and 3K frequency levels could predict variance within the criterion variable of listening comprehension test score. After running the analysis it became apparent that only WRS scores from the 1K and 3K level offered a unique contribution to the prediction of listening comprehension test scores. As WRS scores for the 2K frequency level did not provide a statistically significant and unique contribution to the predictive power of the regression model it was automatically excluded. Thus, two regression models were built in two steps. Model one included only the 3K WRS scores and was able to predict 52% of the variance within listening comprehension tests scores ($F(1, 165) = 180.90, p < .001, R^2 = .52$). Model two included both 3K WRS and 1K WRS test scores and resulted in a statistically significant 2% increase in the model’s power to predict variance within the listening comprehension test scores ($F(2, 164) = 97.56, p < .001, R^2 = .54$). Thus, the scores for 1K and 3K WRS scores both contributed significantly and uniquely to the regression model and in sum were able to account for 54% of the variance observed in the listening comprehension scores (see Table 3.5).

In order to more fully investigate the 2K WRS test scores in predicting variance observed within listening comprehension, an additional stepwise regression model was built, within which only 1K WRS and 2K WRS test scores were inputted as predictor variables (see Table 3.6).
When confining predictor variables to those which fell below the 3K level, that is 1K WRS and 2K WRS scores, it was evident that 2K WRS scores were able to predict 48% of the variance observed in listening comprehension test score. However, as was shown in the results of Table 3.5, when all three WRS word frequency level scores were added to the regression model, the predictive variable of WRS 2K test score was not able to provide a unique contribution beyond that offered by 3K WRS test scores.

Table 3.5: Step-wise multiple regression including 3K and 1K WRS scores (2K WRS scores excluded)

<table>
<thead>
<tr>
<th>Unstandardised coefficients</th>
<th>Standardised coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>Std. Error</td>
</tr>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.11</td>
</tr>
<tr>
<td>3K WRS</td>
<td>.83</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-5.61</td>
</tr>
<tr>
<td>3K WRS</td>
<td>.61</td>
</tr>
<tr>
<td>1K WRS</td>
<td>.51</td>
</tr>
</tbody>
</table>

Note: $R^2 = .52$ for step 1; $\Delta R^2 = .02$ for step 2; * $p < .01$, ** $p < .001$.

Table 3.6: Step-wise multiple regression including 2K and 1K WRS scores (3K WRS scores excluded manually)

<table>
<thead>
<tr>
<th>Unstandardised coefficients</th>
<th>Standardised coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>Std. Error</td>
</tr>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-4.03</td>
</tr>
<tr>
<td>2K WRS</td>
<td>.68</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-8.68</td>
</tr>
<tr>
<td>2K WRS</td>
<td>.45</td>
</tr>
<tr>
<td>1K WRS</td>
<td>.59</td>
</tr>
</tbody>
</table>

Note: $R^2 = .48$ for step 1; $\Delta R^2 = .02$ for step 2; * $p < .05$, ** $p < .001$. 
3.5.3 Research question 3: What is the comparative profile of word recognition from speech performance across the first, second and third thousand frequency levels for those participants attaining modest, competent and good listening comprehension scores?

The next stage of analysis involved investigating the 1K, 2K and 3K WRS test scores obtained by participants attaining modest, competent and good listening comprehension scores. Raw score cut-offs equating to IELTS listening performance bands were used to select 102 of the highest performing listening comprehension test takers from the entire group. Raw scores out of 40 marks were used to categorise these groups into the following performance bands: modest users (raw score of 15-11, \( n = 34 \)), competent users (raw score of 16-22, \( n = 39 \)), and good user (raw scores of 23 or above, \( n = 32 \)) (IELTS, 2014). All participants whose scores fell into any of these three categories were included in the analysis. These listening level categories were used as the grouping variable to analyse differences between WRS scores for the 1K, 2K and 3K word frequency levels.

Table 3.7: 1K, 2K and 3K WRS scores for modest, competent and good listeners

<table>
<thead>
<tr>
<th>WRS test section</th>
<th>( n )</th>
<th>Category of listening comprehension</th>
<th>Mean WRS scores (%)</th>
<th>( SD )</th>
<th>Sig. (2-tailed)</th>
<th>Effect size (( \eta^2 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1K</td>
<td>34</td>
<td>Modest</td>
<td>82.48</td>
<td>9.82</td>
<td>.00*</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>Competent</td>
<td>92.53</td>
<td>5.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>Good</td>
<td>94.16</td>
<td>5.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2K</td>
<td>34</td>
<td>Modest</td>
<td>72.50</td>
<td>16.40</td>
<td>.00*</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>Competent</td>
<td>87.14</td>
<td>9.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>Good</td>
<td>92.65</td>
<td>5.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3K</td>
<td>34</td>
<td>Modest</td>
<td>57.25</td>
<td>19.12</td>
<td>.00*</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>Competent</td>
<td>76.13</td>
<td>12.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>Good</td>
<td>82.60</td>
<td>11.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(*p < .01\)

One-way ANOVA showed that significant differences existed between the mean WRS scores attained by participants of different listening comprehension level: 1K WRS,
\[ F(2,102) = 26.33, p < .05, \eta^2 = .34; 2K \text{ WRS}, F(2,102) = 28.56, p < .05, \eta^2 = .36; 3K \text{ WRS}, F(2,102) = 27.07, p < .05, \eta^2 = .35 \]. Post hoc analyses undertaken with Fisher’s Least Significant Difference (LSD) tests indicated that mean WRS scores for modest category listeners were significantly different from both of the two higher levels of user competence (competent and good) for each of the three word frequency levels involved in the analysis \((p < .01)\). LSD tests indicated that differences between 2K WRS scores for participants from the competent category \((M = 87.14\%, SD = 9.21)\) and good category \((M = 92.65\%, SD = 5.36)\) were also statistically significant \((p < .05)\).

3.6 Discussion

The first point of significance was the overall strong level of association observed between WRS test scores and listening comprehension test scores. Of particular significance was the strength of association observed between the 3K WRS scores, which were obtained from just 29 items, and listening comprehension \((r = .72, p < .01)\). This strength of association is comparable to those observed in previous studies such as that of Staehr (2009) \((r = .70)\) and Milton et al., (2010) (Spearman's \(\rho = .67\)) which employed test instruments containing more than three times the number of target words than those used in the 3K WRS test component of this study. This suggests that vocabulary tests which are mediated through the aural modality and require productive demonstration of word knowledge under time constraints robustly reflect the constructs of word knowledge which are associated with proficient listening comprehension.

Secondly, the ability to recognise high frequency words from speech has good utility in the prediction of listening comprehension. For the participants involved in the present study, the possession of greater WRS speech capabilities was strongly reflective of an increased ability to more successfully negotiate listening comprehension tasks. This finding is suggestive of the importance of a learner’s ability to recognise high frequency vocabulary from speech as a key component of skilled L2 listening comprehension. This finding is strongly suggestive of the practical pedagogical value of measuring WRS of high frequency words. As with other word frequency based vocabulary tests, (Laufer & Nation, 1999; Nation, 2001; Schmitt, et al., 2001), WRS tests provide explicit information about the strengths and weaknesses of a learner’s word knowledge status in relation to different word frequency levels. However, the measurement of WRS also
provides diagnostic information regarding the test taker’s knowledge of spoken word forms. As deficits in word recognition from speech are known to strongly inhibit L2 listening comprehension (Goh, 2000), the potential value of this diagnostic information is significant. Regular low stakes measurement of high frequency words from speech may draw attention to the need for learners to be able to recognise the phonological form of words in the manner these words are likely to be encountered in connected speech. Such testing regimes used in conjunction with learning protocols which develop an increased capacity to recognise L2 words from speech, in our view have the potential to contribute positively to systematic efforts aimed at improving L2 listening comprehension (Matthews et al., 2015; Matthews & O’Toole, 2015).

Thirdly, using a well-known criterion for listening comprehension competence as a contextual frame has enabled insight into the specific levels of WRS associated with different levels of listening comprehension. Due to the correlational nature of this study we are unable to assert that attaining good levels of listening comprehension are a direct result of a strong capacity to recognise high frequency words from speech. However, Table 3.7 does provide concrete information useful for guiding learners towards the levels of WRS associated with their desired category of listening comprehension competence. For example, assuming the results from the present study provide a valid predictive model for other learners with similar linguistic attributes, it would seem that students aiming to develop good or very good scores on standardised listening tests such as IELTS are likely also to be able to attain scores on WRS tests at or beyond 90% at least to the 2K level. In our view, diagnostic information acquired from WRS tests used in conjunction with categories of competence based on standardised test cut-off scores can be usefully applied in practical language learning contexts.

3.6.1 Limitations and recommendations for future research

The current study has sought to provide additional empirical insight into the relationship between high frequency WRS and listening comprehension. Although the current paper in our view has succeeded in providing a clearer understanding of the specificity and practical implications of this relationship, several aspects of this study clearly stand to benefit from future elaborations and more in-depth empirical study. Firstly, although the construct of word recognition from speech was able to account for a large proportion of the variance observed in listening comprehension scores, it must be noted that a
considerable proportion of variance (approximately 46%) was not. Of immediate interest in this regard is to determine whether measuring WRS of words from beyond the high frequency level is able to add predictive power to models seeking to explain variance within listening comprehension test scores. Indeed, investigating the relative contribution of other measures of linguistic and non-linguistic knowledge, as well as WRS, in the prediction of listening comprehension scores holds promise in regards to the development of more empirically grounded theoretical models of listening comprehension in general.

Secondly, of additional interest would be to compare the magnitude of the associations between WRS and listening comprehension observed among the participants of this study with other groups of learners. Determining the degree to which WRS can predict listening comprehension among those with different native languages and proficiency levels to those involved in the current study may provide additional insight into the generalisability of these results in other learning contexts. Further, as has already been alluded to, the correlational nature of this investigation presents clear limitations in the ability to generalise beyond the sample involved in this study. Although a link between WRS and listening comprehension is clearly suggested by these results, a causative link between WRS and listening comprehension cannot be drawn based on the results of this study alone. To explore the potentially causative link between the ability to recognise high frequency words from speech and listening comprehension ability, alternative research paradigms would need to be employed. For example, of high value in this regard would be to gather data from language learners undertaking task-based listening assessments through the application of think-aloud protocols. Gathering such data in conjunction with quantitative collection of WRS test data may enable more robust conclusions to be drawn regarding the nature of the association between WRS and the ability to effectively undertake listening comprehension tasks.

Lastly, in view of the strong link between high frequency WRS and listening comprehension test scores, research exploring the design and evaluation of teaching and testing systems which facilitate measurable improvements in L2 WRS are of strong interest. The design and evaluation of computer assisted language learning and testing systems specifically aimed at addressing this goal are likely to be of significant value in the pursuit of higher levels of L2 listening comprehension.
3.7 Conclusion

The strong link between WRS and listening comprehension clearly suggests that the role of lower order listening skills in listening comprehension instruction should not be undervalued. The ability to recognise high frequency vocabulary from speech is evidently one of the fundamentally important domains of knowledge indicative of skilled listening comprehension. As a consequence, testing the ability of learners to recognise high frequency vocabulary from speech and responding strategically to the diagnostic information derived from such tests, should be an important priority in the listening classroom. Being able to recognise the phonological form of high frequency words provides a broad coverage of the spoken language and establishes a strong platform of linguistic knowledge. Once firmly established, this platform can be a basis from which to build other forms of linguistic and non-linguistic knowledge which can also be usefully applied in an effort to more successfully comprehend the target spoken language.
Summary

The overarching finding of Chapter 3 was that L2 WRS is strongly correlated with, and highly predictive of, L2 listening comprehension success. The findings detailed in Chapter 3 provide an empirical basis from which primary research question one can be answered in the affirmative. The strong link shown to exist between L2 WRS and L2 listening suggests that L2 WRS does impose a sufficient influence on L2 listening comprehension to make research into the development of L2 WRS a productive enterprise. The empirical evidence provided in Chapter 3, provides a strong rationale for efforts to develop effective approaches to develop L2 WRS among learners in real language learning contexts.
Chapter 4

Preamble to research paper 2: “Investigating an innovative computer application to improve L2 word recognition from speech”

The previous chapter confirmed that L2 WRS is a strong correlate with and predictor of L2 listening comprehension. This finding highlights L2 WRS as an important component of successful L2 listening comprehension and thus encourages the investigation of approaches which effectively build L2 learners’ capacity to recognise words from L2 speech. As discussed in the literature review of this thesis (see section 2.5), computer assisted language learning (CALL) does have the capacity to facilitate theoretically sound approaches to the development of L2 WRS. However, to date no research has directly investigated the effectiveness of CALL in the development of L2 WRS in real language learning contexts.

Research paper 2, which is presented in Chapter 4 below, addresses the primary research question, “Does computer-assisted language learning (CALL) provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS a useful way forward?” In order to address this research question, Chapter 4 describes a web-based CALL application which has been developed as part of the research objectives of this thesis. The CALL application has been specifically developed to facilitate the TRC learning protocol outlined in the literature review (see section 2.4.2.4) in order to assist L2 learners to develop their capacity to recognise L2 words from speech. Chapter 4 describes the exploratory implementation of the CALL application among a group of L2 learners. Thus, Chapter 4 seeks to present empirical data which enables a preliminary evidence based assessment on how effective the CALL application is in the development of L2 WRS.
Research paper 2: Investigating an innovative computer application to improve L2 word recognition from speech


The content presented in this chapter is a minimally adapted version of the article which appears in the journal Computer Assisted Language Learning. Presentation of this article in its current form is in line with the copyright policy of the publisher of the article, Taylor and Francis.

Abstract

The ability to recognise words from the aural modality is a critical aspect of successful second language (L2) listening comprehension. However, little research has been reported on computer mediated development of L2 word recognition from speech in L2 learning contexts. This report describes the development of an innovative computer application to improve L2 word recognition from speech, and goes on to describe an exploratory study of its use by 33 ESL university students. The study detected significant differences between word recognition scores measured before and after participants used the application. The manner with which students engaged with the computer application in terms of word recognition, self-determined exposure to speech input and modified output productivity (revisions) are also described and analysed, as well as differences in engagement by learners with different word recognition capabilities. The paper concludes with a discussion of some recommendations for the use of computer assisted development of L2 word recognition from speech and also provides concrete suggestions for much needed further empirical research in this area.

Keywords: CALL; listening; L2 word recognition from speech; engagement

4.1 Introduction

The ability to process and derive meaning from L2 speech is a critical component of L2 language acquisition (Feyton, 1991), and is therefore important to language teachers and students. The foundation of listening comprehension is the ability to recognise words in connected speech (Hulstijn, 2007; Rost, 2002). Automaticity, characteristic of first
language (L1) word recognition from speech, is acquired through extensive exposure to contextualised speech input (Ellis, 2002). The implicit decoding routines that underpin this automaticity enable the L1 listener to readily recognise the words in connected speech, and consequently activate the associated semantic representations in the mental lexicon (Segalowitz & Hulstijn, 2005). Early stage L2 learners, on the other hand, find it harder to recognise the words comprising L2 speech (Goh, 2000). This difficulty stems from the intrinsic blended and ephemeral nature of the speech signal and a relative paucity of exposure to contextualised target language in the aural modality (Hulstijn, 2003; Trofimovich, 2008). Consequently, L2 listeners are often unable to adequately process the incoming L2 speech signal and thus rely strongly on guessing and contextual cues in an effort to derive its meaning (Field, 2008a; Tsui & Fullilove, 1998).

Several pedagogical recommendations on how best to develop L2 word recognition from speech have been advanced (Field, 2003, 2008a; Hulstijn, 2003; Wilson, 2003). Despite current computer assisted language learning (CALL) applications having the technological capability to operationalise many of these recommendations, few published studies have investigated the implementation of CALL approaches specifically aimed at improving L2 word recognition from speech. This apparent lack of research may stem from an entrenched view that deficiencies in L2 word recognition from speech can be countered through employment of contextual information (Field, 2008c). This study aims to address the gap in the current literature.

This paper uses a psycholinguistic model to highlight the value of L2 word recognition from speech in L2 listening pedagogy. It goes on to describe a computer application designed to improve L2 word recognition from speech. The paper then reports the evaluation of an exploratory study of the use of the program among a group of ESL students which determined significant improvements in word recognition capability with its use. Finally, trends in learner engagement over time, and how these were influenced by variation in learners’ word recognition capability, are also reported and discussed.
4.2 Background

4.2.1 A psycholinguistic rationale for complementary approaches to L2 listening pedagogy

Listening comprehension is mediated by rapid, accurate and socially contextualised cognition which draws upon linguistic and non-linguistic knowledge sources (Buck, 2001). The “top-down and bottom-up” information processing model is widely used to conceptualise the cognitive processes and knowledge types drawn upon during listening comprehension, and has exerted a strong influence in the field of L2 listening (Graham & Macaro, 2008; Tsui & Fullilove, 1998; Vandergrift, 2007). The model posits that the cognitive processes involved in listening comprehension exist along a bi-directional spectrum ranging between lower-level (bottom-up) processes and higher-level (top-down) processes. The lower-level end of the processing spectrum involves the serial recognition and coalescence of smaller language units to build larger chunks of meaning (Field, 1999; Vandergrift, 2007). Higher-level or top-down processes are generally described as involving use of context, prior knowledge of content, world knowledge, and metacognitive knowledge in order to facilitate interpretation of speech input (Vandergrift, 2004). Lower and higher level cognitive processes are not mutually exclusive and operate in an interactive manner (Graham & Macaro, 2008; Lynch, 2006). Despite the recognition of the importance of this interactivity of lower and higher order processes in successful L2 listening comprehension, the current emphasis in published research in L2 listening pedagogy has been on higher-level, particularly metacognitive, strategies (Graham, et al., 2008; Graham, et al., 2010; Vandergrift, 2002, 2003a).

Making predictions about content, selectively attending to salient sections of text and monitoring on-going levels of listening comprehension are important aspects of skilled L2 listening (Graham & Macaro, 2008; Macaro, Graham, & Vanderplank, 2007). The employment of listening strategies has been shown to be associated with improvements in listening comprehension (Graham & Macaro, 2008). However, strategies which emphasise engagement of explicit knowledge and conscious control over listening processes must be used in conjunction with lower-level L2 listening skills such as word recognition (Graham, et al., 2010). L2 listeners who are unable to accurately recognise linguistic data may become overly dependent on compensatory strategies such as guessing from context, leading to listening comprehension errors (Field, 2008a).
Therefore it is important that L2 listening pedagogy includes learning activities that emphasise both higher-level and lower-level processes (Graham, et al., 2010).

This realisation should generate interest in pedagogical approaches which improve lower-level processing, such as those involved in mapping sounds onto the corresponding words they represent (Field, 2008c). With the lexical level being the lowest linguistic level at which stable formal-semantic associations occur (Hulstijn, 2002), word recognition represents a conduit through which lower and higher levels of speech processing can effectively interact. Despite this facilitative role of word recognition in listening comprehension, word recognition has been undervalued in L2 listening research and pedagogy (Broersma & Cutler, 2008; Field, 2008b; Wilson, 2003). Developing improved levels of L2 word recognition requires a large time commitment and intensive practice (Hulstijn, 2003). Therefore, devoting time to develop this skill outside of the classroom appears very important (Field, 2008b). Activities that emphasise the development of lower-level listening processes such as word recognition should receive a stronger focus in the L2 learning curriculum (Field, 2003, 2008a, 2008b; Hulstijn, 2003, 2007). Computer applications which deliver learning protocols aimed at improving L2 word recognition from speech have the potential to assist learners to more effectively develop this skill and may provide a useful complement to existing modes of listening pedagogy.

4.2.2 Recommendations for the development of word recognition from L2 speech

Learning opportunities that facilitate cognitive links between the transient speech signal and its corresponding fixed orthographic representation are associated with improved perception of L2 speech (Bird & Williams, 2002; Mitterer & McQueen, 2009). This finding suggests that the systematic use of aural and written input (bimodal input) may improve learners’ perceptual sensitivity to the composite words of target language speech. Studies addressing the benefits of bimodal input have largely done so through investigation of the improved comprehension brought about by synchronous presentation of aural and written modalities (Vanderplank, 1988, 1990). In contrast, other researchers have suggested the pedagogical value of asynchronous presentation of aural and written modalities as a means by which to specifically develop L2 word recognition from speech (Field, 2003, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003). Central to these recommendations is the temporal separation of aural and written input.
presentation. Synchronous bimodal presentation may result in learners attending primarily to the written modality at the expense of aural processing (Bird & Williams, 2002). Allowing an initial focus on the aural presentation alone, with subsequent concurrent presentation of aural and written modalities, circumvents this issue.

The dominant recommendation for improving aural L2 word recognition is to increase the learner’s opportunity to listen to examples of target speech (Field, 2008b; Hulstijn, 2003). This recommendation stems from usage based theories of L2 acquisition which assert that the frequency of language processing events and language performance are closely linked. However, there is also the requirement to ensure that the exposure to input and resultant processing events are mindful and motivating (Ellis, 2002). Therefore, while the quantity and frequency of the learner’s engagement with L2 target speech is important, it is also critical to consider the quality of their engagement.

From a synthesis of existing pedagogical recommendations regarding how best to improve L2 word recognition from speech (Field, 2003, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003) learning activities aimed at improving L2 word recognition should include the following stages. Firstly, a learner listens to a section of speech input, attempting to reconstruct the text as fully as possible. Repeated, self-paced exposure to the speech input enables listeners to have multiple opportunities to revise hypothesised content of the text and furthermore provides scaffolding for the cognitively demanding task of text reconstruction. This phase aims to foster a critical approach to word recognition, where listeners “construct and carry forward provisional interpretations” (Field, 2008a, page 49). This is in contrast to traditional dictation techniques, which typically impose strict teacher centred control over the number and timing of repeated listening opportunities (Jafarpur & Yamini, 1993; Rahimi, 2008). Finally, after listening to the aural input, listeners compare their attempt at reconstruction with the written transcript of the target speech. This last stage provides an opportunity for listeners to notice discrepancies between their language production and the target language, a vital component of L2 acquisition (Schmidt, 1990).

The combination of these stages constitutes intensive listening practice. Here intensive listening is defined as listening which involves the listener closely attending to a text to develop an awareness of the specific language features present in that text. This can be
contrasted with extensive listening which emphasises listening to a text for the purposes of extracting general meaning and global comprehension (Rost, 2002).

4.2.3 Investigating word recognition development and learner engagement with a computer application aimed at developing L2 word recognition from speech

Despite the prolonged acknowledgement of the value of using computers to enhance L2 learners’ listening skills (Brett, 1995; Grezel & Sciarone, 1994), many of the previous studies have focused on extensive listening, aiming for the development of global listening comprehension (Absalom & Rizzi, 2008; Grgurović & Hegelheimer, 2007; Guichon & McLornan, 2008; Smidt & Hegelheimer, 2004). Whilst the value of extensive listening as a component of listening pedagogy is not disputed, few studies have examined the use of computers to develop lower-level decoding processes through intensive listening. The studies that have investigated the use of CALL for lower order listening skills through reconstruction of speech input into writing (Coniam, 1996, 1998; Grezel & Sciarone, 1994) have done so for language testing, with predetermined limits imposed on the number of opportunities to listen to and reconstruct the text. Because of this, these studies conflict with recommendations advocating student control of listening and reconstruction to improve L2 word recognition from speech.

The exploratory study reported in this paper aims to add research momentum to the limited number of published studies which have focussed on computer assisted intensive listening programs. The report addresses three broad areas relating to the use of a prototype web-based computer application designed in line with the previously defined recommendations on how best to develop L2 word recognition from speech. It begins with a description of the application. Then the effect on L2 word recognition development observed in a single group of learners before and after use of the application will be analysed and reported. Finally, it describes learner engagement with the application, defined as the extent to which learners are actively involved in learning activities and the quality of that involvement (Cole & Chan, 1994; Gonida, Voulala, & Kiosseoglou, 2009). The study explores three aspects of learner engagement:

1. Learning protocol word recognition score, defined as the number of words accurately recognised by participants as they undertake the learning protocol delivered by the computer application.
2. Self-determined exposure to speech input, defined as the number of times learners elect to listen to a given section of L2 target language speech.

3. Modified output productivity, defined as the number of times participants amend their efforts at reconstructing the target language into the written form.

Learner engagement is an important determinant of the success of CALL (Hegelheimer & Tower, 2004). Previous researchers have employed CALL as a way to investigate the way different learners use learning applications designed to improve L2 listening skills (Grgurović & Hegelheimer, 2007; Hartwell, 2010; Hegelheimer & Tower, 2004). Learners of different listening capabilities engage in the use of CALL listening applications in different ways (Roussel, 2011). Analysing engagement with CALL by learners of different levels of word recognition ability may contribute to refinements to pedagogical approaches aimed at developing L2 word recognition from speech.

After describing the function of the computer application this paper goes on to investigate the following research questions:

1. Will measurable improvements in word recognition be observed among a cohort of learners after using the application?

2. How will self-determined exposure to input, modified output productivity and learning protocol word recognition scores vary as learners progress through the computer application?

3. How will the self-determined exposure to input, modified output productivity and learning protocol word recognition score vary between learners of relatively low and high word recognition ability?

4.3 Methodology

4.3.1 Participants

The participants were students from a single class of 33 (18 males and 15 females) students enrolled in a course for international students at a Southeast Asian university. This group was drawn from the total class number of 44, 11 of whom declined to be
involved in the use of the computer application. The language of instruction for the course was English. The mean age of participants was 21.7 years (minimum of 19 and maximum 25 years). Participants had a range of self-reported first languages: 10 Chinese, 8 Malay, 7 Malay and Thai, 3 Thai, 1 French, 1 Cham and Khmer and 1 Arabic and Malay (2 missing values). The mean self-reported duration of English language study was 5.4 years (minimum of 1 and maximum 10 years). In line with the human research ethics committee requirements of the author’s institution, written informed consent to participate was obtained from each student. Furthermore participants were only contacted in relation to this study after full human ethics approval had been received.

4.3.2 Materials for the computer application

Monologues containing factual information relating to countries in the broader Asian region were produced to suit the learning needs of participants expressed by facilitating course instructors. Factual information and static image files for each monologue were obtained from online information in the public domain. For each country chosen (Indonesia, Burma, India, Singapore, and Brunei) a monologue was written and recorded into MP3 file format by the author, a male Australian native English speaker.

The lexical content of the monologues was analysed using Vocabprofile (Cobb, 2013; Heatly & Nation, 1994) to determine the proportion of the words from each monologue which were within the first thousand frequency range (1K), second thousand frequency range (2K) (Nation, 2001; West, 1953), academic word list (AWL) (Coxhead, 2000) or low-frequency words (words which fell beyond the 2000 word range and were not from the AWL).

The monologues had a mean duration of 31.6 seconds (minimum of 29 and maximum of 37 seconds) and mean number of words of 65.6 (minimum of 59 and maximum of 75 words). Each of the five monologues was digitally edited to produce 9 shorter monologue sections. The mean number of words of the 45 monologue sections was 7.3 (minimum of 4 and maximum of 12) and the mean duration of these sections was 3.6 seconds. Monologue sections were edited to approximately three seconds to ensure they approximated the average duration of an intonation unit of speech. This decision was made to ensure that the sections of text would impose a burden on phonological
short term memory which approximated that of the short bursts of speech typical of ordinary speech (Rost, 2002). The monologues were edited to correspond with appropriate breaks in natural intonation of the recorded speech.

An analysis of the low-frequency words present in the monologues revealed that 30 of the 50 low-frequency words were either the names of countries in the Asian region or the adjective forms of these names. The presence of these geographic names throughout the listening texts reflects an adherence to requests made by facilitating course instructors to ensure learning content related to countries in the Asian region. Additionally, as the research location was within Southeast Asia, this theme was anticipated to increase the likelihood that students had pre-existing knowledge of the content of the listening texts.

Table 4.1 General description of the five monologues used in the learning protocol

<table>
<thead>
<tr>
<th>Sequence and theme</th>
<th>Total words</th>
<th>Average words / second</th>
<th>Percentage of words in 1K/2K/AWL/low-frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Indonesia</td>
<td>67</td>
<td>2.1</td>
<td>74/7/4/15</td>
</tr>
<tr>
<td>2 Burma</td>
<td>59</td>
<td>2.0</td>
<td>73/7/5/15</td>
</tr>
<tr>
<td>3 India</td>
<td>75</td>
<td>2.0</td>
<td>76/3/8/13</td>
</tr>
<tr>
<td>4 Singapore</td>
<td>66</td>
<td>2.2</td>
<td>74/5/6/15</td>
</tr>
<tr>
<td>5 Brunei</td>
<td>62</td>
<td>2.1</td>
<td>73/2/8/17</td>
</tr>
</tbody>
</table>

All monologues, monologue sections, static images and text files were uploaded to the application’s server, which facilitated the delivery of the web-based learning protocol as described below.

4.3.3 Description of the computer application

The first phase of the protocol was a complete delivery of the full monologue and concurrent on-screen presentation of the monologue’s static contextual image. The static images for each monologue included the country’s name, a geographic location map and an image of the nation’s flag. The images were intended to enable engagement of pre-existing background knowledge relevant to the lexical content of the monologue.
A progress bar was displayed on screen to ensure participants were aware of the progress of the sound file.

Phase two allowed participants to listen to the monologue sections and attempt to reconstruct the content of the speech into the written modality by entering text into an on-screen text box with the keyboard and mouse. On each occasion that text was entered into the text box and the “Submit Attempt” panel was clicked, or if the time limit for submission elapsed, the attempt at reconstruction was captured in the database, but was maintained in the text box allowing amendment after subsequent repetition of the monologue section. Up to 10 attempts to listen to and reconstruct each section were available. The decision to set the maximum number of repetitions at 10 was based on pilot observations that most learners listened to the monologue sections less than 10 times.

Participants could opt to skip some of the available attempts by checking a “Disable Remaining Repetition” check box. Participants were given instruction about the check box to ensure they were made aware of its location and function, and so were able to choose the number of times they listened to each of the monologue sections. The number of listening attempts for each participant was captured in the database, allowing analysis of the construct of self-determined exposure to speech input. Additionally, all attempts at text reconstruction submitted by participants were similarly stored, providing a record of the number of alterations of text reconstruction for each section of the monologue, allowing analysis of the construct of modified output productivity.

After listening to the monologue section the chosen number of times, on-screen feedback was provided (phase three), consisting of an on-screen comparison of the participant’s final attempt at text reconstruction with the target text. A coded algorithm automatically compared the final attempt with the target text. Words incorrectly transcribed or missing were highlighted in red text. Clicking the “Play Transcript” panel allowed the opportunity to listen to the monologue section while text feedback remained on-screen. Participants were then given the choice to progress to the next section or to log out. Returning participants were automatically directed back to the point in the learning protocol from where they previously logged out. Unidirectional movement through the learning protocol was ensured by an application code which disabled the browser “back” button.
4.3.4 Word list

Using spoken texts with few unknown words is pedagogically important in L2 word recognition development as “learners cannot ‘re-cognize’ words which they do not yet ‘cognize’ to some extent” (Hulstijn, 2003, p. 421). Word lists were distributed to participants before commencing use of the web application. All of the words from the five monologues were pooled and arranged into an alphabetically ordered, categorised word list. The word order of the word list differed from that of the word order of the monologues. As the word list contained all of the words present in the five monologues used, the participants had ample opportunity to become familiar with all of the composite words of the monologues before undertaking the word recognition protocol.
The word list was categorised using the following headings: general content words, function words, words relating to religion, words relating to direction and specific locations, words relating to origin and language, words appearing regularly in academic texts, and words relating to trade, agriculture, economy and resources. It was anticipated that the provision of a categorised word list may assist participants to activate appropriate schemata while engaged in the transcription of the monologue sections. The word list was a simple list of words in their isolated form. Participants were free to translate or elaborate on the word forms as they saw fit.

4.3.5 Pre-tests and post-tests

Vocabulary knowledge is modality specific (Song, 2008), with learners often being able to recognise a word in the written modality but being unable to recognise that same word when presented solely in the aural modality (Goh, 2000). As this study is specifically focussed on L2 word recognition from speech, it was important that tests used to measure this construct were channelled through the aural modality. Three test types were used.

4.3.5.1 Partial dictation tests

Partial dictation tests (see Appendix 4.1) involve the presentation of a text in the aural modality accompanied by the text in the written modality with target words missing. Variations of this basic testing format have been validated in both traditional (Cai, 2013) and computer based settings (Coniam, 1998). This test was used to determine word recognition development and the cohorts’ baseline ability to recognise words from the aural modality. The rationale for use of this testing format was that it would provide a highly scaffolded context from which word recognition could be gauged. If learners could not recognise a given word from the aural modality in this test format then it could be assumed that the word was functionally unrecognisable from the aural modality under less structured contexts. The twenty words from the pre and post partial dictation test consisted of sixteen words from the 0-1000 frequency range, 3 words from the 1000-2000 range and 1 low-frequency word.
4.3.5.2 Paused dictation tests

Paused dictation tests (see Appendix 4.2) presented solely in the aural modality, were also used to measure word recognition development. A contextualised sentence was listened to and then paused in order to enable the last word of the sentence to be transcribed by participants. The rationale for use of this testing format was to tap into a construct of word recognition which depended on processing speech for larger-scale meaning (Field, 2008c).

Participants undertook partial and paused dictation tests as a single group in pen and paper format by listening to the stimulus sentences twice each and attempting to transcribe the target words. Partial dictation and paused dictation pre and post-tests varied only in the order with which each item was delivered. The target words for these pre-tests and post-tests were drawn from the word list.

4.3.5.3 Dictation tests

Dictation is an integrative listening comprehension test generally considered to be a valid measure of language ability (Buck, 2001; Cai, 2013; Cohen, 1980; Oller, 1971). Dictation taps into a construct of L2 language ability which depends strongly on aural language processing ability. Dictation tests were delivered via computer in a manner very similar to that of the learning protocol already outlined. A contextual image was presented, a monologue was presented and then learners had the opportunity to transcribe the content of the monologue sections using the keyboard and mouse. Unlike the learning protocol, learners had only three chances to listen to and transcribe the target texts. Different texts were used for the pre-test and post dictation tests (see Appendix 4.3).

Table 4.2 Description of attributes of dictation tests

<table>
<thead>
<tr>
<th>Phase</th>
<th>Total words</th>
<th>Words / sec.</th>
<th>Percentage of words in 1K/2K/AWL/low-frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictation Pre-test</td>
<td>72</td>
<td>2.0</td>
<td>67/4/11/18</td>
</tr>
<tr>
<td>Dictation Post-test</td>
<td>67</td>
<td>2.2</td>
<td>82/2/7/9</td>
</tr>
</tbody>
</table>
As the validity of computerised testing hinges on test takers clearly understanding the procedure of the computerised test (Coniam, 1999), all participants first undertook a trial dictation with the same format as the test but with different content words. The scores of the dictation pre-test were also used to categorise participants into low and high word recognition ability groups.

4.3.6 Attitudinal survey

Surveys are a key evaluative tool in CALL especially when used in conjunction with other data collection methods (Levy & Stockwell, 2006). A brief post-protocol attitudinal survey was administered in an effort to tap into participants’ attitude towards various aspects of the learning protocol.

4.3.7 Procedure

The study was announced by the course coordinator and the opportunity to undertake the learning protocol was offered as a non-compulsory adjunct to course material. The 33 participants were given demographic and language background surveys, pre-tests, a word list and log-on credentials enabling access to the computer application. Students were given 9 weeks to complete the learning protocol in its entirety. Sessions took place on a weekly basis, with the first two and last two sessions being allocated to participant instructions and the administration of tests and surveys. The middle five sessions were allocated to the learning protocol. Participants engaged with the learning protocol for approximately 40 minutes per session, thus, over five sessions approximately 200 minutes were allocated to participants to engage with the five monologues. Students were also able to access the website outside of allocated class time if they so wished. Each student accessed the computer application via the internet with an individual computer and listened through headphones. Post-tests and the post-protocol attitudinal survey were administered at the end of the allotted time period.

4.3.8 Data analysis

4.3.8.1 Self-determined exposure to speech input.

The total number of times participants chose to listen (self-determined exposure to speech input) to each of the 45 monologue sections was quantified by examining the data recorded in the database. This allowed calculation of:
• The mean number of times each participant listened to each of the monologue sections, enabling a comparison of the differences in self-determined exposure between each of the five monologues.

• The mean number of times each participant listened to all 45 monologue sections.

4.3.8.2 Modified output productivity

Similarly, for each of the 45 monologue sections, the total number of revisions made by participants while attempting to reconstruct each section (modified output productivity) was determined. Any change involving alphanumerical keystrokes from one attempt to the next was scored as one revision. This allowed calculation of:

• The mean number of times each participant revised monologue sections for each of the five monologues, allowing comparison of the differences in modified output productivity that occurred between each of the five monologues.

• The mean number of times participants revised their efforts at reconstruction for all 45 monologue sections.

4.3.8.3 Learning protocol word recognition score

This construct of engagement was obtained by scoring participants’ word recognition performance by analysing their efforts at reconstruction. A learning protocol word recognition score was assigned for each participant for each monologue section, using the participant’s best reconstruction attempt (the maximum number of recognised words compared with the total number of words for that monologue section, expressed as a percentage). Each target word written in correct orthographic form, or in a form that contained only minor spelling errors, was awarded a full mark. Half marks were assigned where minor spelling errors in some way hindered the recognition of the target word by the markers. Scoring reliability was achieved by use of a common marking rubric by two independent native speaking markers. A Kappa analysis was undertaken to determine the agreement of word recognition scoring between each scorer. A Kappa
statistic of .91 ($p < .01$) denoted a strong level of agreement between markers. Any discrepancies between markers were reviewed after initial marking and a consensus mark was established by referring to the marking rubric.

4.4 Results

1. Will measurable improvements in word recognition be observed among a cohort of learners after using the application?

Paired sample t-test analysis showed a significant difference between each of the three formats of pre and post-tests indicating significant measurable improvements in word recognition.

- Partial dictation: pre-test ($M = 53.06\%, SD = 19.65$) and post-test scores ($M = 69.84\%, SD = 21.35$); $t(30) = -5.369, p < .01$.
- Paused dictation: pre-test scores ($M = 52.26\%, SD = 21.10$) and post-test scores ($M = 67.63\%, SD = 22.59$); $t(30) = -5.11, p < .01$.
- Dictation: pre-test scores ($M = 46.74\%, SD = 23.57$) and post-test scores ($M = 61.83\%, SD = 27.16$); $t(32) = -3.61, p < .01$.

An indication of the cohorts’ baseline listening vocabulary was established by analysing the results of the partial dictation pre-test, which was the first of the tests administered to the group. A majority of this test’s target words were within the first 1000 frequency range. Results for this pre-test ($M = 53.43\%, SD = 19.45$, Minimum value = 5\%, Maximum value = 90\%) indicated that even under the most highly structured testing format used, collective word recognition capability from speech for the cohort did not significantly exceed the 1000 most frequent word level.

2. How will self-determined exposure to input, modified output productivity and learning protocol word recognition scores vary as learners progress through the computer application?
### 4.4.1 Mean self-determined exposure to input

Results from one-way repeated measures ANOVA showed values for mean self-determined exposure to input differed significantly across the five monologues, $F(4, 128) = 6.70, p < .01$. Paired sample t-test post hoc analyses indicated a significant difference between the mean self-determined exposure to input for each of the following pairs of monologues:

- 1 ($M = 7.07, SD = 2.98$) and 2 ($M = 6.06, SD = 3.32$), $t(32) = 2.85, p < .01$;
- 1 and 3 ($M = 6.10, SD = 3.46$), $t(32) = 3.26, p < .01$;
- 1 and 4 ($M = 5.43, SD = 3.39$), $t(32) = 4.19, p < .01$;
- 1 and 5 ($M = 5.30, SD = 3.56$); $t(32) = 3.75, p < .01$;
- 3 ($M = 6.10, SD = 3.46$) and 4 ($M = 5.43, SD = 3.39$); $t(32) = 2.16, p < .05$.

<table>
<thead>
<tr>
<th>Monologue</th>
<th>Mean self-determined exposure to input</th>
<th>SD</th>
<th>Mean modified output productivity</th>
<th>SD</th>
<th>Mean learning protocol word recognition score</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.07</td>
<td>2.98</td>
<td>2.21</td>
<td>1.57</td>
<td>53.44</td>
<td>27.91</td>
</tr>
<tr>
<td>2</td>
<td>6.06</td>
<td>3.32</td>
<td>1.67</td>
<td>1.35</td>
<td>55.42</td>
<td>29.61</td>
</tr>
<tr>
<td>3</td>
<td>6.10</td>
<td>3.46</td>
<td>1.85</td>
<td>1.50</td>
<td>61.64</td>
<td>32.46</td>
</tr>
<tr>
<td>4</td>
<td>5.43</td>
<td>3.39</td>
<td>1.48</td>
<td>1.21</td>
<td>64.92</td>
<td>31.14</td>
</tr>
<tr>
<td>5</td>
<td>5.30</td>
<td>3.56</td>
<td>1.45</td>
<td>1.15</td>
<td>60.41</td>
<td>28.54</td>
</tr>
</tbody>
</table>

### 4.4.2 Mean modified output productivity

One-way repeated measures ANOVA showed values for mean modified output productivity also differed significantly across the five monologues, $F(4, 128) = 4.58, p < .01$. Paired sample t-test post hoc analyses indicated a significant difference between the mean modified output productivity for each of the following pairs of monologues:
• 1 (M = 2.21, SD = 1.57) and 2 (M = 1.67, SD = 1.35), t (32) = 2.86, p < .01
• 1 and 4 (M = 1.48, SD = 1.21), t (32) = 2.75, p < .05
• 1 and 5 (M = 1.45, SD = 1.15), t (32) = 3.21, p < .01
• 3 (M = 1.85, SD = 1.50) and 4 (M = 1.48, SD = 1.21), t (32) = 2.87, p < .01

4.4.3 Mean learning protocol word recognition score

One-way repeated measures ANOVA showed that mean learning protocol word recognition differed significantly between different monologues, F (4,128) = 2.61, p < .05. T-test post hoc analyses indicated a significant difference between the mean learning protocol word recognition score for each of the following pairs of monologues:

• 1 (M = 53.44, SD = 27.91) and 3 (M = 61.64, SD = 32.46), t (32) = -2.13, p < .05
• 1 (M = 53.44, SD = 27.91) and 4 (M = 64.92, SD = 31.14), t (32) = -2.28, p < .05
• 2 (M = 55.42, SD = 29.61) and 4 (M = 64.92, SD = 31.14), t (32) = -2.05, p < .05

The trends observed over the course of the five monologues were:

• A reduction in both self-determined exposure to speech input and modified output productivity.
• An increase in mean learning protocol word recognition score.

It appears that as participants progressed, they elected to listen fewer times and made fewer changes to their reconstructions however on average they were better able to correctly recognise target words.

3. How will the self-determined exposure to input, modified output productivity and learning protocol word recognition score vary between learners of relatively low and high word recognition ability?

Dictation pre-test results were used to categorise the top 11 and bottom 11 participants into relatively high and relatively low word recognition ability groups respectively.

Independent samples t-tests were conducted to compare means for each of the constructs for low and high word recognition ability groups. Mean values for self-
determined exposure to input and modified output productivity were not significantly different. However, mean learning protocol word recognition score for low ($M = 41.48$, $SD = 31.65$) and high ($M = 78.13$, $SD = 10.27$) word recognition ability groups was significantly different; $t (20) = -3.65$, $p < .01$. In summary, participants with relatively high word recognition ability on average chose to listen more, recast their attempts at transcription more and achieved significantly higher word recognition levels while undertaking the learning protocol.

Table 4.4 Summary of results for dictation pre-test

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean Word recognition (%)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole group</td>
<td>33</td>
<td>46.73</td>
<td>23.57</td>
</tr>
<tr>
<td>Low group</td>
<td>11</td>
<td>20.14</td>
<td>15.44</td>
</tr>
<tr>
<td>High group</td>
<td>11</td>
<td>71.34</td>
<td>7.66</td>
</tr>
</tbody>
</table>

Table 4.5 Summary of engagement by low and high groups across all monologues

<table>
<thead>
<tr>
<th>Construct of engagement</th>
<th>Word recognition ability level</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>self-determined</td>
<td>Low</td>
<td>6.00</td>
<td>3.02</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>7.22</td>
<td>3.08</td>
</tr>
<tr>
<td>modified output</td>
<td>Low</td>
<td>1.43</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>2.10</td>
<td>1.10</td>
</tr>
<tr>
<td>learning protocol word</td>
<td>Low</td>
<td>41.48</td>
<td>31.65</td>
</tr>
<tr>
<td>recognition score (%)</td>
<td>High</td>
<td>78.13</td>
<td>10.27</td>
</tr>
</tbody>
</table>

4.4.4 Attitudinal survey results

In summary, on average, participants perceived the learning protocol as having a positive impact on their interest and motivation towards the target language and their vocabulary knowledge. Perhaps most importantly, considering the overall objective of the learning protocol, participants also reported a perceived positive impact on their English listening skills. In relation to the participants’ attitude towards the difficulty of the task itself and the listening materials used for the task, it seems that these were suitable for the participants’ learning needs. This could not be assumed had responses indicated overwhelmingly that the task and materials were either very easy or very
difficult. Furthermore, it seems that the theme of Asian countries, which was used as a context for the learning protocol, was perceived as being suitably interesting for participants.

Table 4.6 Summary of attitudinal survey results

<table>
<thead>
<tr>
<th>Questions with the format:</th>
<th>Mean score</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very much so = 4 / To some extent = 3 / Not very much = 2 / Not at all = 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the web exercise made you more interested in English than before?</td>
<td>3.30</td>
<td>.78</td>
</tr>
<tr>
<td>Has the web exercise made you more motivated in English than before?</td>
<td>3.19</td>
<td>.75</td>
</tr>
<tr>
<td>Has the web exercise made you more confident about your English listening skill?</td>
<td>3.03</td>
<td>.75</td>
</tr>
<tr>
<td>Has the web exercise improved your knowledge about English vocabulary?</td>
<td>3.33</td>
<td>.84</td>
</tr>
<tr>
<td>How useful was the web exercise for your English listening skill?</td>
<td>3.40</td>
<td>.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions with the format:</th>
<th>Mean score</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very easy = 4 / Quite easy = 3 / Quite difficult = 2 / Very difficult = 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How easy/difficult was the web exercise?</td>
<td>2.60</td>
<td>.62</td>
</tr>
<tr>
<td>How easy/difficult was the speech on the web exercise to understand?</td>
<td>2.67</td>
<td>.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Questions with the format: Very interesting = 4 / Quite interesting = 3 / Not very interesting = 2 / Not interesting at all = 1</th>
<th>Mean score</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>The material on the web exercise was about countries in Asia. How interesting was this material for you?</td>
<td>3.31</td>
<td>.66</td>
</tr>
</tbody>
</table>
4.5 Discussion and Conclusion

Significant improvements in three separate measures of word recognition suggest utility of this computer application in the development of L2 word recognition from speech. The analyses of the learners’ engagement with the computer application also point to the pedagogical value of this prototype. As learners progressed through the learning protocol, they chose to listen and revise their efforts at transcription less frequently, but improved in their ability to recognise words overall. However, it must be emphasised that the exploratory nature of this study does not allow attribution of improvements to the use of the computer application as opposed to the effects of learner maturation. The results of this pilot study warrant further investigation with a larger sample group and quasi-experimental design. Furthermore, tapping qualitative aspects of learner experience with this learning protocol through think-aloud protocols may provide another valuable line of analysis in regards to determining the learning protocol’s pedagogical value and resultant impact on word recognition.

A key area of future research arising from this work is the investigation of the factors which contribute to learners’ levels of engagement. This would appear to be particularly important for listeners of relatively low word recognition ability. Despite recognising fewer words during the learning phase, the lower ability group did not listen or revise their efforts more than higher ability participants. On a practical level it would be helpful to identify factors within a teacher’s sphere of control which may contribute to increased levels of engagement. For example, strategic use of listening materials which reflect the personal interests of the learners using the application may hold promise in this regard. An open question at the end of the survey undertaken with this cohort asked participants about which types of listening materials they would like to use with this application in the future. Opinions were diverse and included areas such as history, music, stories, religion, sport, economics and politics. Furthermore, potentially any word list, grammatical focus, discourse type, or accent type for example could be used as the basis for development of recorded speech input for use with this application. Tailoring materials in response to learner surveys relating to interest or need may assist in the development of learning environments which successfully harness learner motivation to encourage improved levels of engagement.
The application has provided these participants with opportunities to repeatedly recast hypotheses about the lexical content of spoken language in a structured learning environment. Furthermore, it produced accrued data showing levels of engagement and specific word recognition difficulties encountered by language learners over time. Such information may provide teachers and students themselves, with diagnostic information concerning weaknesses in students’ word recognition and could be used to guide explicit teaching in group learning situations. Such approaches, used in conjunction with this computer application, have the potential to enable teachers to help build learners’ metacognitive knowledge in relation to lower order processing, while at the same time providing a means of increasing opportunities for intensive listening practice.
Summary

The overarching finding of Chapter 4 was that engagement with the CALL application described was associated with significant levels of L2 WRS improvement. Chapter 4 provides multiple forms of data which suggest the utility of the application in the development of L2 WRS. As such these results provide initial support that CALL can provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS a useful way forward. However, the one group exploratory design of the research undertaken in Chapter 4 does not enable unequivocal conclusions to be drawn in relation to the effectiveness of the CALL application in the development of L2 WRS. As such the results presented in Chapter 4 are encouraging but demand additional investigation of the same CALL application with a more rigorous quasi-experiment research design.
Chapter 5

Preamble to research paper 3: “Computer-mediated input, output and feedback in the development of L2 word recognition from speech”

The previous chapter provided a preliminary confirmation of the utility of the CALL application in developing L2 WRS in real language learning contexts. Although the results from the previous chapter are not unequivocal, multiple strands of empirical data suggestive of the effectiveness of the CALL application were provided. These data provide an evidence based rationale for a more rigorous investigation of the CALL application in question.

Research paper 3, which is presented in Chapter 5, also addresses the primary research question, “Does computer-assisted language learning (CALL) provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS a useful way forward?” In order to do so, a quasi-experimental pre-test/intervention/post-test research design was undertaken with a control group and two treatment groups. Chapter 5 aims to supplement the depth of the results described in Chapter 4. It aims to do so by applying a more rigorous approach to the quantification of the L2 WRS improvements associated with use of the CALL application described.
Research paper 3: Computer-mediated input, output and feedback in the development of L2 word recognition from speech


http://journals.cambridge.org/abstract_S0958344014000421
Copyright 2014 European Association for Computer Assisted Language Learning

The content presented in this chapter is a minimally adapted version of the article which appears in the journal ReCALL. Permission was granted from Cambridge University Press to present this minimally adapted version.

Abstract
This paper reports on the impact of computer-mediated input, output and feedback on the development of second language (L2) word recognition from speech (WRS). A quasi-experimental pre-test / treatment / post-test research design was used involving three intact tertiary level English as a Second Language (ESL) classes. Classes were either assigned to a control group (n = 31) or to one of two alternative treatment levels which used a web-based computer application enabling self-determined opportunities to repeatedly listen to and reconstruct spoken target text into its written form. Treatment group one (n = 30) received text feedback after each of their efforts at target text reconstruction, whereas treatment group two (n = 35) did not. Results indicated that word recognition gain scores of those who used the application, regardless of treatment level, were significantly higher than those of the control group. The relationship between the quantity of self-determined exposure to input and word recognition improvements was moderate but not linear, with those choosing moderate levels of speech input deriving the greatest measurable improvement. Neither increased levels of modified output nor the provision of text feedback were associated with significant improvements in word recognition gain scores. Implications for computer-mediated approaches for the development of L2 WRS are described and areas for future empirical research are suggested.

Keywords: L2 word recognition from speech; listening; CALL design; feedback; input; output
5.1 Introduction

Word recognition from speech (WRS) is the most important component of spoken language processing and the foundation of listening comprehension (Broersma & Cutler, 2008; Rost, 2002). This is the case as mapping elements of the speech signal onto specific lexical types held in the mental lexicon enables semantic word knowledge to be reliably accessed and put to use in the listening comprehension process (McQueen, 2007). The degree to which this process occurs with automaticity is particularly important for language learners: unlike written words on a page, spoken language is “distributed in time and fades quickly from the perceptual field” (Weber & Scharenborg, 2012, page 387). This attribute makes word recognition from speech a highly time-constrained and implicit process which does not lend itself to the application of explicit language knowledge (Hulstijn, 2003). Consequentially, L2 WRS is an aspect of language learning which L2 learners find particularly difficult (Goh, 2000). A consideration of how to help learners deal with this difficulty is pertinent to language educators as lower-level listening skills differentiate more and less skilled L2 listeners (Tsui & Fullilove, 1989). Despite this importance, research focusing on L2 WRS is scarce (Broersma & Cutler, 2008). The research presented here seeks to tackle the research problem of how to build increased capacity of learners to recognise L2 words from speech in a real language learning context. In an effort to achieve this objective, a computer application specifically developed to improve L2 WRS was designed and evaluated using the tripartite framework of input, output and feedback.

5.1.1 The role of input, output and feedback in developing word recognition from speech

The constructs of input, output and feedback figure prominently in second language acquisition theory (Ellis, Loewen, Elder, Erlam, Philp & Reinders, 2009; Gass & Mackey, 2006; Krashen, 1982, 1985; McDonough, 2005; Swain, 1985), and also align with previously untested pedagogical recommendations on how to improve L2 WRS (Field, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003). Although the themes of input, output and feedback have been applied to previous theoretical and practical research in the field of computer assisted language learning (CALL) (Chapelle, 2009; Heift, 2004; Li & Hegelheimer, 2013; Murphy, 2010), their role in the development of computer-mediated L2 WRS has not been rigorously explored in language learning contexts.
The role of frequency of input has a prominent position in theories which seek to explain the acquisition of fluent language skills. The power law of practice asserts that the speed and accuracy of skills such as word recognition from speech, increase in step with the frequency of skill-specific processing events (Ellis, 2002; Hulstijn, 2002). Providing opportunities for learners to repeatedly listen to spoken input is suggested to assist in developing the ability to recognise the location of boundaries between words and to improve automaticity of word recognition (Field, 2008a; Hulstijn, 2003). Previous studies show that the frequency of occurrence of words in aural input has a positive correlation with vocabulary learning, but that this variable is only one of several which are likely to contribute to the efficacy of L2 word learning (Nation, 2001; Vidal, 2003).

The production of output also plays an important role in language acquisition (Swain, 1985, 1995, 2000). According to Swain (1993, page 159), “producing language forces learners to recognise what they do not know or partially know”. In the case of L2 WRS, Field (2008a) recommends students repeatedly listen to fluent speech and attempt to accurately transcribe lexical content while critically reflecting on perceptual evidence and how it relates to tentative interpretations. Presumably this mode of output production pushes learners to reflect carefully on the exact lexical content of spoken language while also providing an explicit representation of the learner’s perception of the spoken form.

Providing feedback which helps learners to notice deficits in their ability to recognise the lexical content of target language may also improve L2 WRS (Hulstijn, 2003; Wilson, 2003). Evidence suggests that “orthographic information can, under certain circumstances, have a significant impact on the long-term implicit, and explicit, learning of spoken word forms” (Bird & Williams, 2002, page529). Providing written feedback enables learners to engage explicit knowledge, and thus consciously attend to problematic elements of speech, and has practical benefits as the fixed form of written words is more readily accessible to learners in a conscious and explicit manner. Further, the role of explicit instruction has been comprehensively reviewed as an important element of effective L2 instruction (Norris & Ortega, 2000). However, the degree to which explicit knowledge, such as that which is engaged during the production of written target output or when attending to feedback, will facilitate improvements in
language skills which are highly dependent on fluent performance, remains an issue of considerable debate (N. Ellis, 2005; Hulstijn, 2005).

Input, output and feedback have been suggested as a useful frame for the implementation of practical interventions aimed at the development of L2 WRS. This research will address the role of these constructs as design features of a web application specifically designed for this purpose.

5.1.2 CALL for the development of L2 word recognition from speech

Despite assertions that digital technology is well suited for the development of lower-level listening skills (Vandergrift, 2007), research with the core objective of using CALL for the development of L2 WRS is scarce. The work of Hulstijn (2003), an influential exception, describes software specifically designed to improve L2 WRS. The software Hulstijn describes enables learners to listen repeatedly, attempt to reconstruct text and then verify predictions of lexical content using the written text transcript. Notwithstanding an absence of published empirical data verifying the effectiveness of these recommendations in real language learning contexts, Hulstijn’s (2003) recommendations have been cited by leading authors in the fields of CALL (Chapelle & Jamieson, 2008; Levy, 2009) and L2 listening (Vandergrift, 2007) as a model of practice for the development of L2 WRS.

Matthews and O’Toole (2015) developed a web-based computer application designed to improve L2 WRS based on an elaboration of Hulstijn’s recommendations, and investigated its use by a cohort of 33 ESL learners. The application enabled learners to listen multiple times, reconstruct target spoken language in writing in a number of phases and receive on-screen feedback highlighting discrepancies between the learners’ output and the target language. The results were suggestive of the effectiveness of computer-mediated approaches which facilitate increased levels of exposure to input, increased production of output and the delivery of feedback. However, the study’s one-group design did not allow the word recognition improvements observed among participants to be directly attributed to learning effects associated with their use of the application. Furthermore, the design of the study did not enable a refined investigation of the relative impact of input, output and feedback on L2 WRS development.
5.2 Research questions

The current study seeks to expand on the few previous theoretical and empirical investigations which have explored the role of CALL in the development of L2 WRS by addressing the following research questions:

1. *What is the impact of the computer-mediated approach described here in the development of L2 word recognition from speech?*

2. *What is the impact of computer-mediated text feedback on L2 word recognition from speech?*

3. *What is the relationship between computer-mediated exposure to input and L2 word recognition from speech?*

4. *What is the relationship between computer-mediated production of modified output and L2 word recognition from speech?*

5.3 Methodology

5.3.1 Participants

Three first-year undergraduate classes enrolled in the same general English language course, studying at a Chinese university, were involved in the study. Each class was kept intact, but randomly assigned to one of three levels: a control group ($n = 31$), treatment level one ($n = 30$) or treatment level two ($n = 35$). All students spoke Mandarin Chinese as a first language and were aged between 17 and 20. Objectives of the course prescribed by the cooperating institution included the development of listening ability and the use of computers to improve language learning outcomes. All participants were accustomed to regular computer and internet usage.

5.3.2 Materials for the computer application

The listening materials used with the application contained factual information about countries in the Asian region based on piloted material. Eight short monologues, with an average of 67 words were recorded in MP3 format. The lexical content of the monologues was analysed using Vocabprofile (Cobb, 2014; Heatley & Nation, 1994) in order to determine the proportions of high frequency, academic, and low frequency
words (Coxhead, 2000; Nation, 2001; West, 1953). The relative proportions of these word categories in the monologues were comparable to those found in fiction, newspapers, and academic texts (Nation, 2001). The average speech rate of the monologues was 2.1 words per second, which approximates the speech rate of an academic lecture presented at average speed (Tauroza & Allison, 1990). Monologues were edited into shorter monologue sections, with an average length of 7.3 words, such that their average duration approximated that of the intonation units typical of fluent speech (Rost, 2002). For each monologue, a static contextual image was sourced in order to be used as pre-task advanced organiser. The speech input from each monologue section was also transcribed into a text file for use with the application. The combined use of monologues, monologue sections, static images and transcript texts will be discussed later in this paper.

5.3.3 Overview of the application

The desired design elements for the web application were specified and collaboratively developed by the primary author and a small team of professional web-application developers. Open source web server software and database management systems were used. The application was deployed on a remote web server.

Access via the administrator interface allowed selection of the sequence and timing of any number of sound files, text files and image files to be established prior to user access. Key variables which were set in relation to the current study were the number of maximum opportunities learners had to listen to target language, the opportunity to enter text in response to hearing sound files, the presence or absence of text feedback, and the option to allow learners freedom to proceed to the next sound file at their own pace. Key data captured by the database were the number of times users chose to listen to each sound file, the text output produced by users in response to each time a monologue section was heard, and the feedback mode of each participant.

The web application first played the complete monologue while the contextual static image remained on-screen. After listening to the full monologue, participants listened to the nine monologue sections one after the other. Clicking an on-screen panel enabled the participants to control when the monologue section began playing, after which the monologue section played all the way through without pauses. Participants were able to
listen to each monologue section between one and ten times and were instructed to choose to listen as many times as they felt was appropriate for their learning needs. This design feature enabled participants to increase their exposure to target speech input in a manner which students themselves deemed to be most appropriate (Figure 5.1).

Figure 5.1 Opportunities to increase exposure to input and to produce modified output

Each time the monologue section was played, an opportunity to transcribe the aural text by typing reconstructions on-screen was provided. Each time participants chose to listen to the monologue section again, their previous effort at transcription was maintained on-screen such that it could be edited. This design feature enabled the production of output in written form. It also provided an opportunity to recast efforts at transcription in response to the perceptual evidence available from the speech input (Figure 5.2). As mentioned previously, each effort at transcription was recorded in the data base, enabling quantification of the number of times written output was modified by each participant for each monologue section. The construct of modified output was determined as the number of times participants altered their written attempts at text reconstruction for each monologue section. These data were extracted from participants’ text output and coded manually, then checked digitally with a coded algorithm.
After listening to the monologue section the chosen number of times, one of the two feedback formats was made available. Treatment level one participants received a numerical indication of the number of words they correctly transcribed as well as the target text with errors or omissions highlighted in red (Figure 5.3). Treatment level two participants received feedback relating to the number of words correctly transcribed but did not receive target text feedback (Figure 5.4). Both treatment levels were provided the opportunity to listen to the monologue section once again by clicking the “Play transcript” button.
Figure 5.3 Example of feedback mode of treatment level 1 participants (text feedback provided)

Figure 5.4 Example of feedback mode of treatment level 2 participants (no text feedback provided)
5.4 Test instruments

5.4.1 Receptive vocabulary knowledge test

A measure of the receptive vocabulary knowledge of the participants was determined using the second thousand-word level and the academic word level components of a version of the Vocabulary Levels Test (Nation, 1983, 1990; Schmitt, Schmitt & Clapham, 2001). Words from the second thousand-word level are those which are traditionally considered to be high frequency vocabulary items. Academic words are those which occur beyond the second thousand-word level and which occur relatively frequently in a variety of academic texts (Nation, 2001). The portion of the test used consisted of 66 target words with 30 high-frequency and 36 academic target words. As this test format has been validated in a number of contexts (Beglar & Hunt, 1999; Schmitt et al., 2001) it was used as the criterion for concurrent validation of the word recognition pre- and post-tests described below.

5.4.2 Word recognition tests

5.4.2.1 Test structure and specifications

Word recognition was measured with two partial dictation tests specifically designed for the purposes of this study (pre-test, see Appendix 5.1; post-test, see Appendix 5.2). These tests involve test-takers completing a contextual sentence by listening to a spoken stimulus and filling in the blank. Each partial dictation test (hereafter word recognition test) consisted of 60 items: the first 32 target words were high frequency words, while the last 28 were from the Academic Word List (Coxhead, 2000; Nation, 2001). The target words for odd-numbered items for both the pre-test and the post-test were words present within the content of the monologues used in the learning intervention (hereafter intervention target words). Target words for even-numbered items for both the pre-test and the post-test were words which did not appear in the learning intervention monologues (hereafter general target words). Therefore the word recognition pre-test and post-test consisted of 16 high frequency intervention words, 16 high frequency general words, 14 academic intervention words and 14 academic general words. The intervention target words of the pre-test and post-test were the same, but to minimise potential learning effects from one test administration to the next, each word was
embedded in a different contextual sentence. The general target words of the pre-test and the post-test were different.

5.4.2.2 Construct and concurrent test validation

A two-part process was used to ensure the primary construct being measured by the word recognition tests was the ability to recognise the phonological form of the target words. First, a panel of adult native English language speakers completed the word recognition test items without listening to the spoken stimulus. This was to ensure that the target words could not be systematically recognised by using contextual cues present within the written sentences of the items. Second, a new panel of native speakers undertook the tests, but this time with the accompanying spoken language stimulus. All panel members recognised all target words correctly. This process ensured that the contextual information in the item stem was not the primary source of information used to recognise the target word and that the WRS capacity required for high levels of test achievement did not exceed that of expert target language users.

The participants’ receptive vocabulary knowledge scores (criterion) were strongly correlated with both word recognition pre-test scores, \( r = .73, n = 96, p < .01 \) and post-test scores, \( r = .74, n = 96, p < .01 \).

5.4.2.3 Word recognition test scoring procedures

Word recognition scores were assigned by using a structured scoring rubric which limited the influence of spelling errors on scoring (Buck, 2001). This was achieved by assigning a mark to target words written in the correct orthographic form or in a form that was clearly recognisable despite minor spelling errors. Words which were written with a degree of ambiguity, but which were still recognisable as the target word, were ascribed a half mark. In order to standardise this procedure, the scoring rubric was used by two trained, independent native speaking test scorers. Cohen’s kappa analysis indicated a strong level of inter-rater agreement (\( \kappa = .91, p < .01 \)). Any discrepancies between the scorers were resolved by consultation between them with reference to the scoring rubric after initial scoring had been completed.
5.4.2.4 Test reliability

A split halves reliability analysis was used to quantify internal consistency of the word recognition pre-test and the post-test. Scores achieved for odd and even test items were first correlated and then corrected with the Spearman-Brown formula. Both tests were found to have a high level of internal consistency (Table 5.1).

Table 5.1 Summary of internal consistency measures for pre-test and post-tests

<table>
<thead>
<tr>
<th>Test and scoring protocol</th>
<th>Correlation between split halves forms</th>
<th>Spearman-Brown coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test: partial word recognition</td>
<td>.88</td>
<td>.94</td>
</tr>
<tr>
<td>Post-test: partial word recognition</td>
<td>.89</td>
<td>.94</td>
</tr>
</tbody>
</table>

Reliability coefficients between pre- and post-tests were determined through bivariate correlational analysis. Word recognition pre- and post-test scores were strongly and positively correlated; \( r = .89, N = 96, p < .01 \).

5.5 Procedure

The study was undertaken within a seven week period, with the first and last weeks allocated to testing. All students \( (N = 96) \) undertook the receptive vocabulary knowledge test and the word recognition pre-test in the first week of the study. Participants in the treatment groups were provided with log-on credentials and a short in-class demonstration of the use of the computer application. Each of these students accessed the computer application via the internet with an individual computer and listened through headphones. Five 60-minute sessions took place in a computer laboratory once a week for five weeks. All 65 treatment group participants completed the eight monologues within the allocated 300-minute time frame. Word recognition post-tests were administered to all 96 participants in the seventh week of the research.
period. All three groups were taught by the same language instructor and received similar learning activities other than those undertaken by the two treatment groups once a week in the computer laboratories.

5.6 Analysis

In order to gauge improvements in L2 WRS, two gain scores were calculated for each participant:

1. Word recognition gain score for intervention target words.
2. Word recognition gain score for general target words.

Gain scores were calculated by subtracting the appropriate post-test score from the corresponding pre-test score. Research questions were primarily investigated by comparing mean values of these two gain scores grouped in the following ways:

1. Participants grouped based on whether they used the application or not.
2. Participants grouped based on whether they belonged to the control group, treatment group one (text feedback) or treatment group two (no text feedback).
3. Participants grouped on relative exposure to input (treatment group only).
4. Participants grouped on relative production of modified output (treatment group only).

The statistical analyses used to compare the mean gain scores for these groupings were either independent samples t-test or one-way analysis of variance (ANOVA). Shapiro-Wilk tests, coupled with visual appraisal of gain score histograms and normal probability plots, indicated that these data were normally distributed. Levene’s tests confirmed homogeneity of variance among all sets of group means compared in the analyses.

5.7 Results

5.7.1 Research question one

What is the impact of the computer-mediated approach described here in the development of L2 word recognition from speech?
In order to investigate the impact of using the web application on L2 WRS, participants’ gain scores from both treatment groups were combined to form a single group and mean gain scores were calculated. Independent samples t-tests were conducted to compare the mean WRS gain scores attained by those participants who used the application (treatment groups one and two) and those participants in the control group who did not use the application (Table 5.2).

Table 5.2 Mean WRS gain scores for participants who used the application and those who did not

<table>
<thead>
<tr>
<th>Category of target word</th>
<th>Group</th>
<th>Mean word gain score (words)</th>
<th>SD</th>
<th>Sig. (2-tailed)</th>
<th>Effect size (Cohen’s d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention target words</td>
<td>Treatment</td>
<td>6.44</td>
<td>3.26</td>
<td>.04*</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.03</td>
<td>2.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General target words</td>
<td>Treatment</td>
<td>4.93</td>
<td>3.26</td>
<td>.45</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>4.39</td>
<td>3.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

Intervention word recognition gain scores were significantly greater for those participants who used the application ($M = 6.44$, $SD = 3.26$) than for participants in the control group ($M = 5.03$, $SD = 2.71$); $t (94) = 2.08$, $p < .05$. No significant differences were observed between the two groups in relation to general target word gain scores. Although these results suggest that the use of the application, either with or without text feedback, was associated with significant improvements in L2 WRS, these positive effects were confined to intervention target words.

5.7.2 Research question two

What is the impact of computer-mediated text feedback on L2 word recognition from speech?

In order to quantify the impact of text feedback on L2 WRS, mean word recognition scores for each of the three groups (treatment group one, treatment group two and control) were compared using a one-way ANOVA. The main focus of this analysis was
the comparison of mean word recognition gain scores for treatment group one and treatment group two. However, control group mean gain scores were also of interest as they indicate the levels of word recognition development likely to have occurred for various reasons not directly linked to the use of the application. No significant differences in mean gain scores in word recognition for either intervention or general target words were observed (Table 5.3).

Table 5.3 Mean WRS gain scores for treatment one, treatment two and control group

<table>
<thead>
<tr>
<th>Category of target word</th>
<th>Group level</th>
<th>Mean word gain score (words)</th>
<th>SD</th>
<th>Sig. (2-tailed)</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention target words</td>
<td>Treatment 1</td>
<td>6.77</td>
<td>3.89</td>
<td>.09</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Treatment 2</td>
<td>6.16</td>
<td>2.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.03</td>
<td>2.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General target words</td>
<td>Treatment 1</td>
<td>4.35</td>
<td>2.50</td>
<td>.32</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>Treatment 2</td>
<td>5.42</td>
<td>3.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>4.39</td>
<td>3.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.7.3 Research question three

What is the relationship between computer-mediated exposure to input and L2 word recognition from speech?

To determine if differences in the amount of exposure to speech input had a measurable impact on WRS gain scores, treatment group one and treatment group two were pooled into a single group ($n = 65$). Participants from this combined group were then allocated to either low, moderate or high exposure categories based on 33 percentile cut-off points for the mean number of times each participant chose to listen to the 72 monologue sections. The mean number of times each monologue section was listened to by the entire group was 4.44 times ($SD = 1.22$, min. = 2.46, max. = 8.61). The low exposure category had fewer than 3.82 repetitions per monologue section; moderate exposure participants had on average between 3.82 and 4.72 repetitions per section; high exposure category participants had over 4.72 repetitions per monologue section. These categories were used as the grouping variables for analysis of differences between mean WRS gain scores (Table 5.4).
Table 5.4 Mean WRS gain scores for low, moderate and high exposure to input categories

<table>
<thead>
<tr>
<th>Category of target word</th>
<th>Relative exposure to input category</th>
<th>Mean word gain score (words)</th>
<th>SD</th>
<th>Sig. (2-tailed)</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention target words</td>
<td>Low</td>
<td>5.11</td>
<td>2.36</td>
<td>.02*</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>7.86</td>
<td>3.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>6.33</td>
<td>2.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General target words</td>
<td>Low</td>
<td>5.50</td>
<td>2.94</td>
<td>.15</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>5.45</td>
<td>2.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>3.79</td>
<td>3.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*$p < .05$

A one-way ANOVA showed a significant impact of relative exposure to input on WRS gain scores for intervention target words, $F(2,62) = 4.33, p < .05, \eta^2 = .12$. Tukey’s Honest Significant Difference (HSD) tests indicated that only differences between gain scores for low ($M = 5.11, SD = 2.36$) and moderate ($M = 7.86, SD = 3.81$) exposure categories were significant ($p < .05$). A one-way ANOVA showed no significant impact of relative exposure to input on recognition of general target words.

In summary, the amount of self-determined exposure to input had a statistically significant association with the differences in mean WRS gain scores observed between groups. However, this relationship was confined to improvements on intervention words and to differences observed between participants from low and moderate categories of input exposure. A Pearson product-moment correlation coefficient indicated a moderate positive relationship between mean exposure to input and word recognition gain scores among low and moderate exposure category participants only ($r = .35, n = 44, p < .05$).

Correlational analysis of a combined group consisting of participants from all three categories, including the high exposure group, yielded an explanatory model which was clearly non-linear. These results suggest that increased levels of exposure to input were associated with WRS improvements up to a point, beyond which the positive
relationship between exposure to input and word recognition improvements for intervention target words ceased.

5.7.4 Research question four

What is the relationship between computer-mediated production of modified output and L2 word recognition from speech?

Again, treatment group one and treatment group two were pooled into a single group (n = 65) and categorised according to relative levels of modified output productivity. The mean number of times participants modified their efforts at text reconstruction for the 72 monologue sections was 2.23 times per section (SD = .62, min. = 1.00, max. = 4.04). The low output category participants had mean modified output values of less than 1.93 modifications per monologue section, the moderate output category between 1.93 and 2.39 modifications, and the high category more than 2.39 modifications. These categories were used as the grouping variables for the ANOVA in order to analyse differences in mean WRS gain score between the groups (Table 5.5).

A one-way ANOVA showed no significant relationship between mean modified output categories on either intervention or target WRS gain scores. The small effect sizes shown in Table 5.5 confirm the limited practical impact of the relative output modification category on differences in mean WRS gain scores.

Table 5.5 Mean WRS gain scores for low, moderate and high modified output categories

<table>
<thead>
<tr>
<th>Category of target word</th>
<th>Relative production of modified output</th>
<th>Mean word gain score (words)</th>
<th>SD</th>
<th>Sig. (2-tailed)</th>
<th>Effect size (η²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention target words</td>
<td>low</td>
<td>6.13</td>
<td>3.01</td>
<td>.66</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>moderate</td>
<td>6.22</td>
<td>3.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>6.98</td>
<td>3.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General target words</td>
<td>low</td>
<td>5.58</td>
<td>3.19</td>
<td>.57</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>moderate</td>
<td>4.69</td>
<td>2.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>4.60</td>
<td>4.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.8 Discussion and conclusions

The finding that those participants who used the application experienced significantly greater improvements in L2 WRS than those in the control group provides empirical support for the value of CALL in the development of L2 WRS. These results suggest that similar interventions, applied among those with similar linguistic attributes to those involved in this study, are likely to yield positive L2 WRS improvements within relatively short periods of time. However, if the results of this research do indeed provide a useful predictive model for other learning contexts, then any such word recognition improvements are likely to be confined to those words which are present in the speech input used as part of the learning intervention. These findings provide insight into the specificity of such learning approaches and encourage strategic pedagogical decisions about the listening material used for the purposes of developing L2 WRS. Specifically, it is important that the lexical content of listening materials used clearly matches the specific word learning objectives of the language courses or groups of students in question. Ensuring learners have the ability to recognise high frequency vocabulary and the specialised vocabulary particular to their field of anticipated expertise would be a logical and practical starting point.

Another significant finding was that a positive relationship between the quantity of exposure to input and WRS gain scores was only evident among participants in the low and moderate input level groups. Participants choosing to listen to high category levels of input did not significantly outperform those that listened to low and moderate category levels. There are several alternative explanations for the non-linear relationship observed between exposure to input and WRS gain scores. First, as suggested by previous research, increasing exposure to input is likely to only be of benefit to those learners who have the underlying linguistic proficiency necessary to derive its potential benefit (VanPatten, Williams & Rott, 2004; Vidal, 2011). In support of this assertion, a correlational analysis indicated that mean self-determined exposure to input was negatively correlated with word recognition pre-test scores ($r = -.55$, $n = 65$, $p < .01$). This suggests that those participants who chose to listen to greater levels of input had a relatively low baseline L2 WRS proficiency. The implication is that those participants who did derive word recognition improvement in association with increasing levels of input did so with relatively few repetitions (less than 4.72) and had relatively high
baseline levels of word recognition capability. In short, the relative benefit of repeated listening for improving WRS is likely to be influenced by the existing word recognition proficiency status of the individual language learner.

Overall these findings suggest that a careful consideration of input, specifically the learner’s current knowledge status in relation to target language material, is likely to be of great importance in effective and targeted interventions aimed at the improvement of L2 WRS. However, caution must be exercised in drawing conclusions regarding the potentially causative effect of increased levels of input on improvements in WRS among high proficiency level students. The amount of input received by participants was a function of participants’ free choice and uncontrolled by the researchers. It is therefore possible that other factors, such as learner motivation, contributed to the relationship between exposure to input and WRS improvement and thus must also be amply considered in the delivery of future L2 WRS interventions and research.

In contrast to the findings in relation to the role of input, the provision of text feedback was not associated with statistically significant improvements in L2 WRS. Similarly, increased levels of output through transcription of target speech into the written form were not associated with greater levels of WRS gains. The impact of differing levels of these constructs, analysed as isolated independent variables, was not sufficient to elicit statistically significant improvements in WRS over the time frame of the intervention. Contextualised within the framework of this study, these findings cast some doubt over the pedagogical value of explicit feedback in the development of L2 WRS. Although speculative, it may be the case that word knowledge developed through relatively explicit modes of learning is not readily transferrable to highly time constrained language processes such as those which underpin the ability to recognise words from speech. These findings are in line with previous research which shows word knowledge developed through explicit instruction is relatively unstable when applied to spoken word forms (Takimoto, 2009). Additionally, it may be the case that measurable improvements in word recognition linked to the provision of text feedback or the opportunity to reconstruct written output require longer durations of time than those implemented as part of the current research. However, based solely on the results of this study, there is a lack of compelling quantitative evidence to support the role of text feedback and the production of output in the development of L2 WRS.
This study has endeavoured to provide additional empirical research relating to computer-assisted development of L2 WRS. However, there are several limitations of this study, the identification of which may provide a useful framework for future improvement and elaboration on the current work. First, although it was not feasible here, future research of a similar nature would be well served by randomly assigning individual participants to different grouping levels. Secondly, as previously mentioned, the timeframe of this study was relatively brief, involving just five weeks of intensive practice. Increasing the duration of similar future studies may yield learning trajectories discrepant from those summarised here. For example, it is reasonable to postulate that greater amounts of input over longer periods of time may be necessary for lower proficiency participants to derive measurable benefit from similar interventions (Vidal, 2011).

Additionally, for logistical reasons it was only possible to measure word recognition shortly after the completion of the last word recognition learning session. Delayed post-tests would provide additional information on the durability of the WRS improvements associated with use of this or similar applications. Further, this paper only reported on the quantitative results of the investigation; supplementing these with the qualitative information also gathered as part of this study in future reports will add additional insight to the findings presented here. Finally, investigating the comparative success of similar pedagogical approaches with different target languages and through use of mobile technology platforms would also be of interest.
Summary

The finding from Chapter 5 that those learners who used the CALL application attained greater levels of improvement in L2 WRS than those participants who did not use the application provides an empirical basis from which primary research question two can be answered in the affirmative. Such results provide empirical evidence that CALL can provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS an effective option. These results suggest that the CALL application described in Chapter 4 (see section 4.3.3) and Chapter 5 (see section 5.3.3), has good utility in the development of L2 WRS in real language learning contexts. As the CALL application described had a measurable positive effect on the development of L2 WRS, the CALL application described also provides a useful prototype from which future improved iterations of CALL for the development of L2 WRS can be planned and developed.
Chapter 6

Preamble to research paper 4: “The impact of proficiency level on interaction, task success and word learning: design implications for CALL to develop L2 word recognition from speech”

The previous chapter confirmed that the CALL application described is effective in the development of L2 WRS. This finding thus presents the CALL application as a useful prototype from which future improved iterations of CALL for the development of L2 WRS can be based. Although the results described in Chapter 5 provide empirical support for the effectiveness of the CALL application, gathering empirical data which can be used to further refine the CALL application is likely to be of general benefit.

Research paper 4, which is presented in Chapter 6 below, addresses the primary research question, “Which design features might be most effective in improving the utility of CALL for the development of L2 WRS?” The research presented in the following chapter gathers and analyses empirical data drawn from learners’ engaged in the use of the CALL application previously described (see sections 4.3.3 and 5.3.3). This data is used in order to develop recommendations for future improved iterations of CALL for the development of L2 WRS
Research paper 4: The impact of proficiency level on interaction, task success and word learning: design implications for CALL to develop L2 word recognition from speech


The content presented in this chapter is a minimally adapted version of the article which is currently under review for publication in the journal Computer Assisted Language Learning. Presentation of this article in its current form is in line with the copyright policy of the publisher of the article, Taylor and Francis.

Abstract

This paper reports on task interaction, task success and word learning among second language (L2) learners of different levels of word recognition from speech (WRS) proficiency who used a CALL application previously shown to be effective in the development of L2 WRS. Participants (N = 65) were categorised into three levels of L2 WRS proficiency according to their baseline ability to accurately recognise the phonological form of words in connected speech. All participants undertook the same web-based intervention task which involved reconstructing spoken text into the written form in a number of stages. In terms of task interaction, lower WRS proficiency participants chose to listen to stimulus material more often and changed their efforts at reconstruction significantly more frequently than participants of higher WRS proficiency levels. In relation to task success, lower WRS proficiency learners were significantly less accurate overall and less immediate in their arrival at peak reconstruction accuracy than higher WRS proficiency participants. A comparison of gain scores also showed that participants of different L2 WRS proficiency experienced significantly different word learning outcomes after using the application, with moderate proficiency participants significantly outperforming lower proficiency level learners in relation to academic word learning. A synthesis of the results is used to recommend design features of CALL for WRS development which will more adequately cater to learners of different L2 WRS proficiency levels. The paper concludes with some avenues for future research and some implications for the integration of computer-mediated L2 WRS development as a component of L2 listening instruction.
Keywords: Word recognition from speech; L2 listening; CALL design; Proficiency level

6.1 Introduction

Listening comprehension is an essential component of overall second language (L2) proficiency, but one which is particularly challenging for learners (Goh, 2000; Graham, 2006). The challenges of L2 listening can in large part be traced back to the unique features of spoken words. Of these features, arguably the most difficult for L2 learners is the transient nature of the speech signal. The transient nature of speech imparts the need for spoken words to be recognised quickly and at a rate which is not readily controlled by the listener (Hulstijn, 2007). The cumulative processing demands of extracting meaning from fluent speech can result in L2 listeners experiencing cognitive overload (Goh, 2000). Another feature of the speech signal which presents difficulty for the L2 learner is the blended nature of the words encoded in fluent utterances. Unlike written words on a page, demarcation between one word and the next is far less explicit, with fluent speech being characterised by the fusion of the end of one word with the beginning of the next. This feature results in clusters of blended words which have acoustic features significantly different to those of the individual words as they are pronounced in isolation (Field, 2003).

Highly skilled listeners are readily able to deal with the transient and blended nature of spoken language (Field, 2008a, 2008b). For highly skilled listeners, the recognition of words from speech occurs automatically and as such does not impose a significant burden on finite cognitive resources (Hulstijn, 2003). The automaticity of a skilled listener’s recognition of spoken words is linked to a broad foundation of implicit phonological knowledge which has developed through extensive contextualised exposure to the target spoken language (Ellis, 2002). This store of implicit knowledge is important as it frees up cognitive resources for higher-level listening comprehension processes such as ascribing contextualised and nuanced meaning to the spoken message (Rost, 2002). L2 listeners of a relatively low listening skill level are often unable to access and apply existing word knowledge while listening due to insufficient levels of automaticity in word recognition from speech (Goh, 2000). As a result, such L2 listeners must endeavour to apply strategies to extract meaning from a spoken message.
despite an inability to adequately perceive a sufficient proportion of its composite words.

The challenges experienced by L2 listeners provides motivation for L2 researchers and teachers to formulate, validate and refine approaches which alleviate these difficulties. A contemporary view of listening instruction is that it should comprise of two main branches: the development of listening strategies and the development of lower-level listening skills (Graham, Santos, & Vanderplank, 2010). Advocates of the value of strategy instruction contend that conscious monitoring of listening processes assists listeners to strategically accommodate for gaps in their linguistic knowledge and processing capacities (Graham, et al., 2010; Vandergrift, 2007; Vandergrift & Tafaghodtari, 2010). The second, and less broadly researched pedagogical branch, aims to develop lower-level listening skills. The most critical of these lower-level listening skills is word recognition from speech (WRS) (Rost, 2002), which is defined here as the ability to accurately map information encoded in speech onto the corresponding lexical units already stored in the mental lexicon. Advocates of the value of lower-level listening skill development assert that a minimum level of word recognition capability must be attained before higher-order listening strategies can be effectively applied (Graham, et al., 2010). Further, it is unclear to what degree the application of explicit knowledge contributes to the development of skills such as L2 WRS which are primarily dependent on automatic processing at the lexical level (Matthews, Cheng, & O’Toole, 2015). Notwithstanding the various debates which exist regarding the relative value of the two complementary approaches to listening instruction (Field, 2008b; Hasan, 2000; Osada, 2001), lower-level listening skills are of strong importance in successful listening comprehension (Matthews & Cheng, 2015; Tsui & Fullilove, 1998) and as such their systematic development is a very important component of L2 listening pedagogy (Graham, et al., 2010).

The development of L2 WRS is however particularly challenging in learning contexts within which immersion in the target language is unfeasible. This is the case as skills which depend on fluent application of word knowledge need to be developed through extensive engagement with contextualised spoken input (Ellis, 2002; Hulstijn, 2003). It is difficult to provide sufficient levels of contextualised target language input to L2 classroom learners, and a paucity of such input has negative effects on L2 word
knowledge development (Jiang, 2000). CALL offers a means by which to alleviate this language learning difficulty by providing learners opportunities to engage with contextualised input in out-of-class contexts through use of personal devices such as tablets, mobile phones and personal computers. Despite the effectiveness of CALL in the development of improved L2 WRS (Matthews et al., 2015), and its potential to alleviate the difficulties associated with the development of this skill in classroom learning contexts, previous research focussing on this area is limited. This gap in existing knowledge extends to the body of empirical data which can be used to practically inform the improved design of CALL for the development of L2 WRS. As one of the major advantages of CALL for the development of L2 WRS is its potential to facilitate autonomous development of this core component of L2 proficiency outside of regular classroom time, it is important that the design features of such approaches adequately cater to an individual’s learning needs.

The importance of L2 WRS in listening comprehension (Field, 2008b; Hulstijn, 2003; Rost, 2002), the practical difficulties in developing this skill, and the current lack of empirical data which can be used to inform the design of CALL for the development of L2 WRS provide the primary motivations for the current study.

6.2 Literature review

6.2.1 The design of CALL for the development of L2 WRS

Theoretical frameworks for the development of lower-level listening skills in large part contend that improvements in L2 WRS depends on the provision of structured and extensive opportunities to engage with appropriate speech input (Field, 2003, 2008a, 2008b; Hulstijn, 2003; Wilson, 2003). In contrast to traditional dictation techniques which are teacher-centred, and provide an inadequate level of target speech input (Jarapur and Yamini, 1993; Kiany & Shiramiry, 2002; Rahimi, 2008), L2 WRS is most effectively developed through learner-centred text reconstruction tasks which offer multiple opportunities to listen to and reconstruct target text. Field (2008a, page 49) recommends that learners should be “given the opportunity to listen and re-listen as often as they wish until they are satisfied that they have achieved a correct segmentation”. Similarly, Hulstijn (2003) recommends that L2 WRS is most effectively developed through a learner-centred text reconstruction approach by which learners
listen as often as they choose to short sections of speech as they reconstruct what is heard.

Broadly available digital technologies are well equipped to deliver spoken language in structured learning environments, thus making CALL well suited for the development of lower-level listening skills (Vandergrift, 2007). Past studies have provided empirical evidence for the general functionality of CALL in assisting learners of various languages, including Korean, Dutch and English, to make stronger connections between spoken and written word forms (Grezel & Sciarone, 1994; Matthews & O’Toole, 2015; Pyun & Lee-Smith, 2011). Early theoretical frameworks specifically aimed at computer-mediated development of L2 WRS (Hulstijn, 2003) have since been operationalised in real learning contexts with promising results. Matthews & O’Toole (2015) developed and investigated a web-based application with a design based on the second language acquisition (SLA) constructs of input, output and feedback. A single-group pre-test/treatment/post-test research design was applied in order to investigate the effectiveness of the text reconstruction approach for the development of L2 WRS among a group of 33 learners of English as an L2. A feature of the application was the learner-centred control of the timing and number of opportunities to listen to the target text. A battery of word recognition tests indicated that statistically significant levels of improvement in the ability to recognise words from speech occurred among the participants. Of importance also was the finding that participants had a positive attitude towards the usefulness of the student-centred text reconstruction tasks.

Matthews, et al. (2015) provide the clearest evidence for the effectiveness of learner-centred, text reconstruction tasks on the development of L2 WRS. A pre-test/treatment/post-test research design with a treatment and control group was implemented among a group of 96 tertiary level Chinese learners of English. The treatment group of 65 participants undertook the computer-mediated text reconstruction tasks, whereas a control group of 31 participants did not. Carefully validated tests which tapped L2 word recognition from speech indicated that, when compared to the control group, those in the treatment group had significantly greater mean L2 WRS gain scores. Results also indicated that the number of times participants elected to listen to the stimulus listening material had a measurable and positive effect on the magnitude of word learning outcomes experienced by a majority of the treatment group. This study
provides clear empirical support for the pedagogical value of tasks which provide learners repeated opportunities to listen to and attempt to reconstruct spoken language. However, a major criticism of this study is that it did not offer substantive discussion in regard to the need to differentiate design elements to cater to individual difference. The value of CALL for the development of L2 WRS is strongly linked to its utility in the delivery of out-of-class learning opportunities. Accordingly, it is important that such approaches have design features which respond adequately to the individual differences evident between learners. Gathering data drawn from learners involved in the use of a CALL application aimed at developing WRS, and using that data to inform the design of future iterations which cater to individual learner differences, is the central goal of the current paper.

6.2.2 A focus on the learner in the design of CALL for listening development

Despite findings that design decisions have clear impacts on the success of CALL (Cárdenas-Claros, 2014), design research has been underrepresented in CALL literature in recent times (Felix, 2005; Hémard & Cushion, 2006). As the roll-out of digital resources which are of potential value for the development of L2 listening continues (Robin, 2007), so too does the need to more clearly understand how design decisions may impact on the effectiveness of these resources in real language learning contexts (Chapelle, 2009; Garrett, 1991). Using SLA theory to make CALL design decisions is one strand of the CALL design literature (Chapelle, 2004, 2009). However, in order to “prepare a strong and workable design, the designer must be sensitive to individual learner characteristics and the learning context” (Levy & Stockwell, 2006, page 36). Thus although SLA theory is a solid foundation from which the design process can begin, successful elaboration from such theoretical bases requires a careful consideration of how the user interacts with and benefits from CALL and how this information can be effectively applied in CALL design (Hémard & Cushion, 2001, 2006).

Various research paradigms have been used to inform user-centred design approaches in CALL for listening. These include the user and other stakeholders working collaboratively to establish design imperatives for CALL (participatory design) (Cárdenas-Claros, 2014), use of theoretical frameworks to guide the design of help options (Cárdenas-Claros & Gruba, 2013) and qualitative investigation of user’s
attitudes to various CALL design features (Cárdenas-Claros & Gruba, 2012). More closely aligned with the research approach adopted here are those studies which have used quantification of user interactions with established computer-mediated listening applications to draw conclusions about pedagogy and design (Chen, Zhang, & Liu, 2014; Grgurović & Hegelheimer, 2007; Hegelheimer & Tower, 2004; Roussell, 2011; Smidt & Hegelheimer, 2004). This approach to CALL design acknowledges that effective design should be iterative in nature and that revisions to existing designs should be guided by investigations of actual learner use of CALL systems (Heift, 2006).

Among the recurring themes in this research is the strong influence of individual difference on the way learners interact, use and benefit from CALL designed to improve L2 listening. Of the dimensions of individual difference investigated as part of research into CALL for listening, difference in learner proficiency is among the most widely addressed. Previous studies have related varying levels of learner proficiency to aspects of users’ experience with CALL for listening including: the application of different listening strategies, use of help options, and task success. A majority of these studies have concluded that learner proficiency has a significant impact on the way learners use and benefit from CALL for listening. For example, investigations have concluded that learner proficiency level had a significant impact on the types of strategies employed by listeners (Roussell, 2011; Smidt & Hegelheimer, 2004). Proficiency level has also been shown to impact the way learners access CALL help options, with higher proficiency learners using help options more effectively than lower proficiency learners. Additionally, proficiency level has also been found to play a role in the degree of success experienced by learners as they engaged with CALL for listening (Grgurović & Hegelheimer, 2007; Hegelheimer & Tower, 2004). This finding suggests that lower proficiency learners should be given increased levels of scaffolding in order to support higher levels of listening task success. However, specific recommendations on the nature of that scaffolding or recommendations on how that scaffolding may be effectively delivered to CALL users is not well developed in the relevant literature (Smidt & Hegelheimer, 2004).

The developments summarised above highlight the need to carefully consider the relationship between a learner’s proficiency level and the design features of CALL aimed at improving L2 listening skills. Previous research in this area has primarily
focussed on the development of listening comprehension in a global sense, and as with a vast majority of other L2 listening research, has not specifically addressed the recognition of words from speech as a central research objective (Broersma & Cutler, 2008). As such, it is unclear how previous research relates to the way learners of different WRS proficiency levels will interact with CALL specifically designed to develop L2 WRS. Of the few CALL investigations which have focused on the development of L2 WRS as a central research objective, none to our knowledge, have focussed on the impact of learner proficiency on learner interaction, task success and word learning outcomes in an effort to inform CALL design. As such, fundamental questions remain about how dimensions of individual difference, such as learner proficiency, should be catered to by the design features of CALL aiming to improve L2 WRS. It is the goal of this research to begin filling this gap in the existing CALL literature.

At this juncture it is important to address the construct of L2 proficiency. Due to its centrality to the field of SLA, the construct of L2 proficiency has been previously defined through application of a variety of assessment methods and theoretical perspectives (Thomas, 1994). Relatively early models suggest that L2 language proficiency consists of both linguistic and non-linguistic knowledge (Bachman and Palmer, 1996; Canale and Swain, 1980). Empirical investigations have affirmed L2 proficiency as consisting of both linguistic and metacognitive components (Hulstijn, Van Gelderen, Schoonen, 2009; Sasaki, 1993). Based on the balance of evidence from recent research, Hulstijn (2011) categorises the components of L2 proficiency as being either *core* or *peripheral*. The core components are those which relate to “linguistic knowledge and the speed with which this knowledge can be processed” (Hulstijn, 2011, page 238) whereas peripheral components are those which relate to metacognitive competences. According to this framework, the construct of L2 WRS fits into the core component of proficiency. The construct of L2 WRS as operationalised in this study taps both the learner’s breadth of L2 word knowledge and also the speed at which that knowledge can be processed. The term L2 WRS proficiency will be used to refer specifically to a learner’s relative competence in recognising L2 words from spoken language. Thus, L2 WRS proficiency is asserted here as being a construct from the core of L2 proficiency, which along with other constructs of core and peripheral L2 proficiency, constitute overall L2 proficiency. Thus, as has been the case in previous
SLA research, the term proficiency will be used to alternatively refer to either overall language proficiency, a term which encapsulates all components of core and peripheral L2 proficiency, as well as to L2 WRS proficiency, a term which refers to a specific sub-component of core L2 proficiency which involves the ability to recognise and represent words from L2 speech (Leclercq & Edmonds, 2014; Thomas, 1994).

6.3 Research questions

Our initial goal here is to determine the degree to which differences in L2 WRS proficiency impact on the way learners interact with an application designed to improve L2 WRS. Second, we seek to determine the impact of L2 WRS proficiency level on the way learners handle the intervention tasks presented. By doing this we seek to determine if L2 WRS proficiency level is associated with different levels of task success achieved by learners. Lastly, we seek to determine if L2 WRS proficiency level has a significant impact on the word learning outcomes experienced by those who use the application. To achieve these research objectives the following research questions will be addressed:

1. What is the impact of learner L2 WRS proficiency level on the manner by which learners interact with a CALL application for the development of L2 WRS?

2. What is the impact of learner L2 WRS proficiency level on the varying levels of task success experienced by learners during interaction with a CALL application for the development of L2 WRS?

3. What is the impact of learner L2 WRS proficiency level on the word learning outcomes attained by learners after interaction with a CALL application for the development of L2 WRS?

The findings from the following research questions will be synthesised as recommendations for the design of CALL for the development of L2 WRS. The immediate goal of this research is to provide empirical data which will pave the way toward the design of CALL for WRS L2 development which more adequately caters to learners of different L2 WRS proficiency level.
6.4 Methodology

6.4.1 Participants

Participants for this study were 65 members of two first-year undergraduate classes of the same English language course studying at a large university in China. Students were assigned to these classes according to their overall English language proficiency level as determined by tests which were delivered to all first year students at the beginning of the university year. All participants reported Mandarin as their first language and were aged between 17 and 20. The group included 33 males and 32 females. Their mean self-reported duration of English language study was 9.4 years ($SD = 2.4$, min. = 4, max. 18).

6.4.2 Overview of the web-application

The application used in this study has been described in previous research and has been shown to be effective in the development of L2 WRS (Matthews et al. 2014; Matthews & O’Toole, 2015). For the present study the application was set to deliver 8 monologues. The monologues had a mean word length of 67.3 (min. = 59, max. = 75) and contained general information relating to countries of Asia. As the content of the monologues was factual it was possible, with only a few exceptions, to present the content in the present tense. This was used as a mechanism by which to control the degree of syntactic complexity present between monologues. Each monologue was divided into 9 monologue sections. The division of the monologue sections was achieved through use of digital audio editing software (WavePad). The digital audio file of each monologue was imported into the editing suite in order to show a visual representation of the waveform of the monologue. This enabled the identification of the boundaries between the intonation units present within the monologues. Intonation units are utterances of connected speech which mark the natural rhythms for presenting idea while speaking (Rost, 2002). Pauses at the beginning and the end of intonation units were thus considered a valid criterion for dividing the monologues into sections. Each monologue section had a mean word length of 7.3 (min. = 6.3, max. = 8.3) and a mean speech rate of 3.5 syllables per second (min. = 3.3, max. = 3.8). To ensure accuracy, syllable rates were also determined manually for each section. Mean syllable rates reported here are those of the continuous spoken language within each section (intonation unit) and are therefore not affected by the pauses between intonation units.
An overview of monologue word length, speech rate and lexical frequency is provided below (see Table 6.1).

Table 6.1: An overview of monologue word length, speech rate and lexical frequency

<table>
<thead>
<tr>
<th>Monologue</th>
<th>Total words</th>
<th>Mean number of syllables per second</th>
<th>Proportion of words in the 0-2K frequency range (%)</th>
<th>Proportion of words on the Academic Word list (%)</th>
<th>Proportion of words beyond the 2K frequency level and not on the Academic Word list (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70</td>
<td>3.4</td>
<td>68.6</td>
<td>5.7</td>
<td>25.7</td>
</tr>
<tr>
<td>2</td>
<td>72</td>
<td>3.5</td>
<td>70.8</td>
<td>11.1</td>
<td>18.1</td>
</tr>
<tr>
<td>3</td>
<td>67</td>
<td>3.8</td>
<td>80.9</td>
<td>4.4</td>
<td>14.7</td>
</tr>
<tr>
<td>4</td>
<td>59</td>
<td>3.3</td>
<td>76.3</td>
<td>8.5</td>
<td>15.2</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>3.7</td>
<td>78.7</td>
<td>8.0</td>
<td>13.3</td>
</tr>
<tr>
<td>6</td>
<td>66</td>
<td>3.6</td>
<td>78.8</td>
<td>6.1</td>
<td>15.1</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
<td>3.4</td>
<td>74.2</td>
<td>8.1</td>
<td>17.7</td>
</tr>
<tr>
<td>8</td>
<td>67</td>
<td>3.3</td>
<td>83.6</td>
<td>7.5</td>
<td>8.9</td>
</tr>
<tr>
<td>Average</td>
<td>67.3</td>
<td>3.5</td>
<td>76.5</td>
<td>7.4</td>
<td>16.1</td>
</tr>
</tbody>
</table>

A contextual digital image was also sourced for each monologue section. The 8 monologues (72 monologue sections) were delivered via a web-based application accessed via a web-browser and computer. Participants listened to the monologue sections through headphones from an individual computer. The material associated with each monologue was delivered in the following cycle:

- Presentation of the static contextual image on-screen as the entire monologue section is heard once. During this phase learners do not reconstruct text.

- Presentation of each of the monologue’s sections in sequence. During this phase learners had the opportunity to repeatedly listen to the sections and reconstruct
the sections into the written form using the keyboard and mouse. Participants were able to click on an on-screen panel such that they controlled when the monologue section began playing. After clicking the “play” panel the monologue section played through without pauses. Participants were able to modify their efforts at reconstruction after each opportunity to listen. This modification involved participants using the keyboard and mouse to self-correct the text they had written in response to listening to the monologue section multiple times. Participants had the opportunity to listen to each sample of spoken discourse between 1 and 10 times. Participants were free to choose when to move on to the next monologue section and could do so by checking an onscreen checkbox and electing to submit their final attempt by clicking an on-screen panel.

- After submitting their final effort at reconstruction for each monologue section participants received on-screen feedback. This feedback included, a) the number of words correctly reconstructed, b) the concurrent presentation of the participant’s final effort at reconstruction and the correct target text with discrepancies between the two highlighted in red (received by half of the participants only), and c) an opportunity to hear the monologue section while the on-screen feedback remained in view.

6.4.3 Assessment of L2 word recognition from speech and word learning outcomes

6.4.3.1 Test structure

Word recognition from speech was measured before and after the intervention with two equivalent partial dictation tests specifically designed to measure L2 WRS. The tests have been used in previous research and were sourced from published academic literature (Matthews, et al., 2015). The pre-test was used to determine the participants’ baseline L2 WRS proficiency before the intervention. The post-test was used to determine the participants’ ability to recognise words from speech after the participants undertook the intervention.
These tests involve participants listening to a spoken stimulus sentence in order to fill in a single missing word from a contextual written sentence. The stimulus sentence was heard just once with approximately 4 seconds delay between the end of one stimulus sentence and the beginning of the next. The speech rate of the spoken stimulus was approximately 2.1 words per second with each stimulus sentence being on average between seven and eight words in length. As was the case for the speech used for the intervention monologues the spoken stimulus for the tests was recorded to MP3 format with professional quality digital recording equipment by an Australian native speaker of English.

Each test measured knowledge of 60 target words. The first 32 target words were high frequency words, while the last 28 were from the Academic Word List (Coxhead, 2000; Nation, 2001). Word list based structuring of the pre and post-test instruments enabled a measure of WRS gain scores to be established for both high frequency words and academic words. High frequency words were deemed to be appropriate targets, as they make up a majority of the spoken language encountered by listeners (Adolphs & Schmitt, 2003; Nation, 2006). Academic words were considered appropriate target words as they lay beyond the traditional level of high frequency words and are therefore likely to be more challenging for language learners to recognise. Academic words are also of pedagogical value for tertiary level L2 learners as these words appear relatively frequently in academic texts and are thus those words which are likely to meet the students’ actual learning requirements (Folse, 2011). Word learning outcomes were quantified with gain scores (post-test minus pre-test) for total words and for two subsections of the word recognition tests: high frequency words and academic words.

6.4.3.2 Test validation

Following Matthews et al., (2014) a two-stage validation process was undertaken to ensure that the primary construct being tapped by the tests was the ability to recognise the phonological form of words as they are heard in connected speech. The first stage of this validation process involved asking a group of three native speakers to attempt to identify the target words from the blanks in the written contextual sentences without listening to the spoken stimulus. This process confirmed that the target words could not be systematically identified by referring to the contextual sentences within which the target words were embedded. The second stage of validation involved a new group of
three native speakers undertaking the tests, this time while simultaneously listening to the spoken stimulus. All target words were correctly identified by the native speakers. This validation process was important in order to ensure that the tests were tapping test takers’ ability to recognise word from speech through activation of bottom-up listening processes. Further, it ensured that the WRS proficiency levels required for high levels of test achievement were within the range of capabilities typical of those with very high overall target language proficiency levels.

6.4.3.3 Test scoring

As the primary construct of interest is the ability to recognise words from speech, it was important that test takers were not penalised for minor spelling errors (Buck, 2001). In regards to the methodology of scoring word recognition tests, the key concern is to limit the potential threat to construct validity caused by the requirement to represent words in the written form. In order to achieve this goal, a structured scoring rubric was used to ensure categorisation of minor spelling errors and assignment of marks for varying levels of word recognition was systematic (see Appendix 6.1). The rubric assigns full credit to responses which are written in the correct orthographic form. The rubric also assigns full credit to responses which include minor spelling errors which in no way impede the scorer’s ability to recognise the target word. Half marks were assigned for words which could be readily recognised by the scorer, despite a degree of ambiguity introduced due to errors in the representation of the target word. The scoring rubric was piloted by two scorers and was shown to facilitate strong levels of inter-rater reliability. \( \kappa = .91, p < .01 \). Scoring of the tests was undertaken by a research assistant who had been trained in the use of the piloted scoring rubric.

6.4.3.4 Test reliability

Analysis indicated that the pre-test (\( \alpha = .89 \)) and post-test (\( \alpha = .93 \)) both had good levels of internal consistency. Analysis also indicated that scores from pre and post-tests were strongly and positively correlated; \( r = .88, N = 65, p < .01 \).
6.5 Procedures

After written informed consent to participate in the study was obtained from all 65 participants, each undertook the pre-test. Testing took place in two sessions on the same day. After pre-testing each participant received log-on credentials which allowed access to the web-based application. To ensure participants understood how to use the application, an in-class tutorial of its use was provided before intervention sessions began.

Intervention sessions took place in a scheduled weekly one hour listening laboratory class. All participants listened to all 72 monologue sections within the five 60 minute sessions scheduled for the completion of the listening tasks. After completing the intervention sessions each participant undertook the post-test. After completing the post-test, participants were sent an electronic summary of their task performance. These summaries provided a complete record of each written entry participants made in response to hearing the monologue sections. The summaries also showed how many times each participant listened to each of the monologue sections as well as the target text for each of the monologue sections.

6.5.1 Quantifying task interaction, task success and word learning outcomes

6.5.1.1 Task interaction

User task interaction data were automatically captured in the data base associated with the web-application and were therefore unobtrusively collected from the participants as they used the application. Two task interaction variables were determined for each of the participants: repetitions and modifications.

The first interaction variable, repetitions, was the number of times learners elected to listen to each monologue section. This variable was determined by calculating the average number of times each participant listened to the monologue sections, namely by determining the total number of repetitions for all the monologue sections divided by 72.

The second interaction variable, modifications, was defined as the mean number of times participants altered their efforts at text reconstruction for each monologue section. Each effort at reconstruction for a given monologue section, which was in any way
different from the previous attempt at reconstructing the same monologue section, was considered one modification. Thus modification is a measure of the number of times participants self-corrected their effort to reconstruct the text. The total number of modifications was divided by 72 in order to determine the average number of modifications undertaken for each monologue section by each participant.

6.5.1.2 Task success

Two measures of task success were established for each of the participants: reconstruction accuracy and average peak reconstruction attempt.

Reconstruction accuracy is a measure of the total percentage of words correctly transcribed by participants as they attempted to reconstruct the spoken monologue sections into the written form. The total number of words correctly transcribed over the course of the 72 monologue sections was established and this value was then divided by the total number of words in the monologue sections and expressed as a percentage.

The second variable, average peak reconstruction attempt, was a measure of the average repetition number within which the participants attained the maximum degree of reconstruction accuracy for each monologue section. This variable provides information on how immediately each participant was able to attain their greatest number of correctly reconstructed words for each monologue section.

6.5.1.3 Word learning outcomes

Word learning outcomes are defined as the improvements in the ability to recognise words from speech as shown by differences in pre and post-tests scores, namely word recognition gain scores. Gain scores were calculated by subtracting the appropriate component of the pre-test score from the corresponding component of the post-test score. Word learning outcomes are assessed in three categories: gain scores in total WRS, gain scores in the recognition of high frequency words, and gain scores in the recognition of academic words.

6.6 Analysis

The primary method of analysis was to compare the variables of task interaction, task success and word learning outcomes of the 65 participants who were categorised into three groups based on their initial WRS proficiency levels. This was achieved by using
total WRS pre-test scores to establish three groups of equal size with relatively low, moderate and high WRS proficiency level. WRS proficiency level, as defined by these three categories, was the primary grouping variable used for analysis in this study. Analysis of variance (ANOVA) and post hoc testing were the main statistical analyses used to compare the mean values of task interaction, task success and word learning associated with participants within each of the three WRS proficiency levels established.

6.7 Results

The 65 participants were categorised into either relatively low, moderate or high WRS proficiency groups. The mean word recognition scores for the pre-test are shown in Table 6.2. Low WRS proficiency participants \((n = 22)\) were those who achieved 40% or less for total WRS, moderate WRS proficiency participants \((n = 22)\) were those participants who scored between 40% and 55.83%, and the high group \((n = 21)\) were those who scored above 55.83%.

Table 6.2: Pre-test mean L2 WRS scores for the total, high frequency and academic words.

<table>
<thead>
<tr>
<th>Test component</th>
<th>N</th>
<th>Maximum possible score</th>
<th>Mean raw score</th>
<th>SD</th>
<th>Mean WRS score (%)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total WRS</td>
<td>65</td>
<td>60</td>
<td>28.68</td>
<td>9.91</td>
<td>47.79</td>
<td>16.52</td>
</tr>
<tr>
<td>High frequency WRS</td>
<td>65</td>
<td>32</td>
<td>18.14</td>
<td>5.74</td>
<td>56.68</td>
<td>17.95</td>
</tr>
<tr>
<td>Academic WRS</td>
<td>65</td>
<td>28</td>
<td>10.53</td>
<td>4.63</td>
<td>37.64</td>
<td>16.54</td>
</tr>
</tbody>
</table>

To establish a picture of the low, moderate and high participants’ initial level of L2 WRS proficiency, scores for total and test sub-sections, are shown in Table 6.3. ANOVA and post hoc tests were undertaken to determine if the differences between the three groups were statistically significant. Mean scores for each proficiency level, within each test component, were significantly different \((p < .01)\). The expectation that academic words would be more challenging than high frequency words was confirmed.
with each WRS proficiency level achieving higher WRS scores for the high frequency words than for their corresponding scores for academic words.

Table 6.3: Mean scores for total, high frequency and academic words for participants from low, moderate and high L2 WRS proficiency levels

<table>
<thead>
<tr>
<th>Test component</th>
<th>WRS proficiency group</th>
<th>n</th>
<th>Maximum possible raw score</th>
<th>Mean raw score</th>
<th>SD</th>
<th>Mean WRS score (%)</th>
<th>SD</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total WRS</td>
<td>Low</td>
<td>22</td>
<td>60</td>
<td>17.52</td>
<td>5.35</td>
<td>29.20</td>
<td>8.91</td>
<td>.00**</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td></td>
<td>29.43</td>
<td>2.87</td>
<td>49.05</td>
<td>4.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td></td>
<td>39.57</td>
<td>3.70</td>
<td>65.95</td>
<td>6.16</td>
<td></td>
</tr>
<tr>
<td>High frequency WRS</td>
<td>Low</td>
<td>22</td>
<td>32</td>
<td>11.84</td>
<td>3.86</td>
<td>37.00</td>
<td>12.06</td>
<td>.00**</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td></td>
<td>18.89</td>
<td>2.06</td>
<td>59.02</td>
<td>6.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td></td>
<td>24.00</td>
<td>2.39</td>
<td>74.85</td>
<td>7.48</td>
<td></td>
</tr>
<tr>
<td>Academic WRS</td>
<td>Low</td>
<td>22</td>
<td>28</td>
<td>5.68</td>
<td>2.12</td>
<td>20.29</td>
<td>7.57</td>
<td>.00**</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td></td>
<td>10.55</td>
<td>2.01</td>
<td>37.66</td>
<td>7.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td></td>
<td>15.62</td>
<td>2.58</td>
<td>55.78</td>
<td>9.22</td>
<td></td>
</tr>
</tbody>
</table>

**p < .01

6.7.1 Research question one

What is the impact of L2 WRS learner proficiency level on the manner by which learners interact with a CALL application for the development of L2 WRS?

Mean values of the interaction variables of repetitions and modifications for each WRS proficiency category were compared with one-way ANOVA (Table 6.4). One-way ANOVA showed a significant difference between the mean number of times participants of different WRS proficiency (low, moderate and high) elected to listen to each of the monologue sections (repetitions), $F(2,62) = 11.54, p < .01, \eta^2 = .27$. Post hoc comparison with Tukey’s Honest Significant Different (HSD) tests indicated that differences between mean repetitions for the low WRS proficiency group ($M = 5.30, SD$
and the moderate WRS proficiency group ($M = 4.20, SD = .77$) and the low WRS group and the high WRS proficiency group ($M = 3.80, SD = .91$) were significant ($p < .05$). Mean differences in repetitions between the moderate and high groups were not significant to the .05 level. The general trend observed in the data is an inverse relationship between WRS proficiency level and the number of times participants elected to repeatedly listen to the aural stimulus. A Pearson product-moment correlation coefficient affirmed the negative relationship between WRS proficiency level and the number of times participants repeatedly listened, $r = -.56, N = 65, p < .01$. Further, the large effect size ($\eta^2 = .27$) indicates the high practical significance of WRS proficiency level on this interaction variable. This appraisal of the magnitude of effect size is based on the guidelines of 0.01 = small, 0.06 = medium, 0.14= large (Cohen, 1988).

Table 6.4: Comparison of two task interaction variables for three L2 WRS proficiency levels

<table>
<thead>
<tr>
<th>Interaction variable</th>
<th>WRS proficiency group</th>
<th>$n$</th>
<th>Mean value of variable</th>
<th>$SD$</th>
<th>Sig. (2-tailed)</th>
<th>Effect size ($\eta^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Repetitions</td>
<td>Low</td>
<td>22</td>
<td>5.30</td>
<td>1.39</td>
<td>.00**</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td>4.20</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td>3.80</td>
<td>.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Modifications</td>
<td>Low</td>
<td>22</td>
<td>2.70</td>
<td>.63</td>
<td>.00**</td>
<td>.31</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td>2.14</td>
<td>.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td>1.85</td>
<td>.34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**$p < .01$**

These results appear to contradict those of previous research which has investigated the task interaction of L2 learners with CALL designed to improve L2 WRS. This previous research found that learners of relatively low proficiency listened fewer times, with a mean value of 6.00 repetitions, to stimulus material than those of relatively high proficiency, with a mean value of 7.22 repetitions (Matthews & O’Toole, 2015). However, comparison between the studies is difficult as the entire cohort from the aforementioned study was only able to successfully reconstruct on average
approximately 46% of the target text while engaged in the L2 WRS development intervention, and thus had an average baseline WRS proficiency far lower than those in the low WRS proficiency category of the current study. Overall, preliminary evidence from this research suggests an inverse relationship between L2 WRS proficiency and the number of times learners select to listen to speech samples while engaged in L2 WRS development tasks.

ANOVA also showed a significant difference between the mean number of times participants of different WRS proficiency groups revised their efforts at text reconstruction (modifications), $F(2,62) = 14.09, p < .01, \eta^2 = .31$. Again Tukey’s HSD tests indicated that differences between mean modifications for the low WRS proficiency group ($M = 2.70, SD = .63$) and the moderate WRS proficiency group ($M = 2.14, SD = .56$) and the low WRS proficiency and the high WRS proficiency group ($M = 1.85, SD = .34$) were significant ($p < .05$). As with repetitions, mean differences in modifications between the moderate and high groups were not statistically significant. The general trend observed in the data was that the lower WRS proficiency participants revised their efforts at reconstruction more than those of higher WRS proficiency level. A large effect size was evident ($\eta^2 = .31$), as too was a negative correlation between WRS proficiency level and the number of times participants modified their efforts at text reconstruction ($r = -.58, N = 65, p < .01$). Again this affirmed the inverse relationship between proficiency and the number of modifications undertaken and also highlighted the practical significance WRS proficiency had on this interaction variable.

In summary, these results suggest that listeners with moderate or high WRS proficiency listen to significantly fewer repetitions before moving on than do listeners of low proficiency. Further, listeners with moderate or high WRS proficiency make significantly fewer modifications to their effort at text reconstruction than do those of low proficiency. The number of repetitions and number of modification of those with moderate and high WRS proficiency were not significantly different.

6.7.2 Research question two

What is the impact of learner L2 WRS proficiency level on the varying levels of task success experienced by learners during interaction with a CALL application for the development of L2 WRS?
Mean values of the task success variables of reconstruction accuracy and peak reconstruction attempt were compared with one-way ANOVA (Table 6.5).

Table 6.5: Comparison of two variables of task success for three L2 WRS proficiency levels

<table>
<thead>
<tr>
<th>Task success variable</th>
<th>WRS proficiency group</th>
<th>n</th>
<th>Mean value of variable</th>
<th>SD</th>
<th>Sig. (2-tailed)</th>
<th>Effect size (η²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reconstruction accuracy</td>
<td>Low</td>
<td>22</td>
<td>68.46 %</td>
<td>11.63</td>
<td>.00**</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td>81.80 %</td>
<td>8.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td>86.72 %</td>
<td>5.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peak reconstruction attempt</td>
<td>Low</td>
<td>22</td>
<td>3.86</td>
<td>.80</td>
<td>.00**</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td>3.29</td>
<td>.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td>2.94</td>
<td>.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p < .01

A one-way ANOVA showed a significant difference between the mean reconstruction accuracy of participants of different WRS proficiency groups, $F(2,62) = 23.35, p < .01, \eta^2 = .43$. Post hoc comparison with Tukey’s HSD again indicated that differences in reconstruction accuracy for the low WRS proficiency group ($M = 68.46\%, SD = 11.63$) and the moderate group ($M = 81.80\%, SD = 8.97$) and the low WRS group and the high WRS group ($M = 86.72\%, SD = 5.52$) were significant ($p < .01$). Mean differences in reconstruction accuracy between the moderate and high groups were not significant to the .05 level. The general trend observed in the data was that WRS proficiency level was positively correlated with reconstruction accuracy, $r = .73, N = 65, p < .01$. As was expected, the variable of WRS proficiency had a very strong practical impact on reconstruction accuracy ($\eta^2 = .43$).

A one-way ANOVA also showed a significant impact of relative WRS proficiency grouping on peak reconstruction attempt, $F(2,62) = 12.58, p < .01, \eta^2 = .29$. As with the three previous analyses, Tukey’s HSD indicated that differences in peak reconstruction attempt for the low WRS proficiency group ($M = 3.86, SD = .80$) and the moderate WRS group ($M = 3.29, SD = .51$) and the low WRS group and the high WRS group ($M = 2.94, SD = .43$) were significant ($p < .01$). Again mean differences in peak
reconstruction attempt between the moderate and high groups were not significant to the .05 level. A negative correlation between WRS proficiency level and peak reconstruction attempt was evident, \( r = -.56, N = 65, p < .01 \). WRS proficiency also had a strong practical impact on the mean peak reconstruction attempt (\( \eta^2 = .29 \)).

In summary, these results suggest that listeners with high and moderate L2 WRS proficiency levels reconstructed text more accurately and more immediately than those of low WRS proficiency level. However, the difference in reconstruction accuracy and peak reconstruction attempt between those of moderate and high WRS proficiency groups was not statistically significant.

### 6.7.3 Research question three

**What is the impact of learner WRS proficiency level on the word learning outcomes attained by learners after interaction with a CALL application for the development of L2 WRS?**

Gain scores for total improvements in L2 WRS indicated that each WRS proficiency level achieved the desired overall word learning outcome of improved L2 WRS. This word learning outcome was evident for each proficiency level ranging between a mean gain score of 10.10 and 13.14 words. As the ability to recognise high frequency words from speech has a strong association with L2 listening comprehension (Matthews & Cheng, 2015) another targeted learning outcome was an improved ability to recognise high frequency words from speech. Each WRS proficiency level was shown to have achieved word learning outcomes suggestive of an improved ability to recognise high frequency words from speech, with gain scores ranging between 4.41 and 5.82 words. Although improvements in regards to both total WRS and WRS of high frequency words were evident, one-way ANOVA showed no significant difference between the word learning outcomes for total words or high frequency words attained by participants from the three WRS proficiency levels.

Another targeted learning outcome was an improved ability to recognise academic words. Improvements in the ability to recognise academic words for each proficiency level were noted, with mean gain scores ranging between 4.66 and 7.45 words. One-way ANOVA indicated a statistically significant difference between the mean academic word gain scores achieved by the three proficiency level groups, \( F(2,62) = 3.91, p < .05, \).
\(\eta^2 = .11\) (see Table 6.6). Tukey’s HSD indicated that mean academic word gain scores between the low (\(M = 4.66, SD = 4.12\)) and moderate WRS proficiency level groups (\(M = 7.45, SD = 3.10\)) were significant (\(p < .05\)). Differences between the mean academic word gain scores of the low and high and moderate and high groups were not statistically significant to the .05 level. The moderate to high effect size (\(\eta^2 = 0.11\)) suggests that WRS proficiency grouping had a practical impact on mean academic word gain scores. Learners of moderate L2 WRS proficiency level evidently attained significantly greater improvements in the ability to recognise academic words from speech when compared to those in the lower WRS proficiency level category.

Table 6.6: Comparison of total, high frequency and academic L2 WRS gain scores for three WRS proficiency levels

<table>
<thead>
<tr>
<th>Test component</th>
<th>WRS proficiency group</th>
<th>n</th>
<th>Raw gain score (words)</th>
<th>SD</th>
<th>Sig. (2-tailed)</th>
<th>Effect size ((\eta^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total WRS</td>
<td>Low</td>
<td>22</td>
<td>10.48</td>
<td>6.35</td>
<td>.12</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td>13.14</td>
<td>5.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td>10.10</td>
<td>3.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High frequency WRS</td>
<td>Low</td>
<td>22</td>
<td>5.82</td>
<td>3.67</td>
<td>.22</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td>5.68</td>
<td>3.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td>4.41</td>
<td>2.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic WRS</td>
<td>Low</td>
<td>22</td>
<td>4.66</td>
<td>4.12</td>
<td>.03*</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>22</td>
<td>7.45</td>
<td>3.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>21</td>
<td>5.95</td>
<td>2.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* \(p < .05\)
6.8 Discussions

6.8.1 Implications for the design of CALL for the development of L2 word recognition from speech

Although the participants of this study had been assigned to classes according to their overall English language proficiency level, results from the partial dictation pre-test clearly showed that among this group, a significant variation in WRS proficiency existed. This finding is significant as it indicates that overall categorisations of proficiency level based on assessment instruments which tap both core and peripheral components of proficiency, such as task based reading, listening and writing tests, may not be sensitive to differences in core components of proficiency such as L2 WRS. As shown in the results of this study, these differences in word recognition proficiency have significant impacts on the manner by which learners interact with processing tasks, as well as the degree to which learners are able to handle those tasks successfully. Further, the differences in WRS proficiency had a measurable effect on the specific word learning outcomes experienced by learners. This finding suggests the merits of a differentiated approach to L2 WRS development according to an individual’s L2 WRS proficiency. However, from a practical perspective it is difficult for an individual teacher to facilitate such differentiated learning experiences without the benefit of well-designed CALL.

It is important to note that a single partial dictation test robustly predicted differences in task interaction, task success and word learning outcomes. Although pre and post-testing was undertaken in pen and paper format for the purposes of this study, partial dictation tests are able to be effectively administered with computers (Coniam, 1996, 1998). Computerised versions of the partial dictation tests used as part of the present study, represent a useful design feature of computer-mediated development of L2 WRS. The administration of such tests at the beginning of the intervention process could provide a quantification of WRS proficiency and be used to automatically align learners of differing proficiencies with appropriate listening texts. Additionally, administration of computerised post-testing after intervention phases may be a convenient way to obtain data relating to the WRS improvements experienced by learners.
A number of the results from the present study suggest the importance of aligning learners of different WRS proficiency levels with appropriate listening materials. For example, despite the opportunity to listen multiple times, low proficiency learners were on average only able to accurately reconstruct approximately 68 percent of the target text. This level of word recognition is well below the level necessary for the comprehension of spoken discourse (van Zeeland & Schmitt, 2013). In light of recommendations that spoken texts containing only a few unknown words are appropriate for the training of word recognition skills (Hulstijn, 2003), it would seem that for the low proficiency participants involved in this study, the listening texts used were too challenging. How might this information usefully inform the design of CALL which aims to cater to the individual learning requirements of those with a relatively low capability to recognise words from speech? Automated quantification of task success, such as reconstruction accuracy, can be used as a method of formative assessment. Such measures of task success can then be used as a point of comparison with pre-selected achievement benchmarks. Such comparisons can be used to trigger the automated alignment of the learner with an appropriate listening text for the training of L2 WRS. For the purposes of this study, in order to standardise the listening texts encountered by all participants, a single set of texts were used. However, it is suggested that a range of listening materials could be made available to learners at different levels of WRS proficiency. Just as individual difference in reading proficiency is catered to by graded readers of differing degrees of complexity, so too could the needs of learners of different WRS proficiency be catered to with graded listening materials (Wilson, 2003). Although a labour intensive and technically challenging task, a number of levels of listening materials which are suitable for learners of different WRS proficiency could be developed. A collection of listening materials with a known range of word frequency profiles, speech rates, and levels of semantic and syntactic complexity would be of strong value in this regard (Révész & Brunfaut, 2013). CALL users that are consistently unable to accurately recognise spoken texts with a known level of complexity, could be automatically directed towards less challenging texts. Empirical data relating to task interaction and task success presented here offers a starting point from which to loosely base performance standards. Although speculative due to the preliminary nature of these findings, perhaps for those with linguistic attributes similar to those of the participants involved in the present study, a cut-off of an 80 percent accuracy level is a useful rule of
thumb below which learners are provided listening texts with higher proportions of high frequency vocabulary, slower speech rates and reduced semantic and syntactic complexity.

The principles outlined to ensure spoken texts are not overly challenging could also be used to ensure that listening materials are not pitched at an unsuitably low level for those with higher levels of WRS proficiency. For learners involved in the study presented here, accuracy tending towards the 85 to 90 percent level, achieved within four repetitions, may be a suitable trigger for access to more challenging spoken texts. Again it is suggested that the difficulty levels of texts be increased by reducing the relative proportion of high frequency words, increasing syntactic and semantic complexity, as well as by increasing speech rates. The relatively straightforward mechanism suggested here of using on-task measures, such as reconstruction accuracy, as a trigger for alignment of learner proficiency with suitable listening texts is an example of what Heift (2007) refers to as a microadaptive approach to the individualisation of CALL. The advantage of this approach is that it has the potential to cater to the temporal nature of learner’s abilities. Providing the performance benchmarks are set at appropriate levels, this form of automated delivery of listening materials provides a feasible mechanism by which to ensure listening materials are neither too challenging nor too simple for the learning needs of CALL users.

Other aspects of the results reported here have the potential to inform the design of CALL for the development of L2 WRS. For example, assuming the number of times participants elect to listen to a section of spoken text reflects a listener’s proficiency and likelihood of task success, then such measures of interaction could also be used to trigger the provision of individualised scaffolding. Low levels of reconstruction accuracy coupled with high levels of repetition, perhaps beyond five or six repetitions, could trigger the automated provision of scaffolding such as glosses, cloze style structures for the target utterances or by providing the onsets of low frequency words (Conaim, 1998; Hartwell, 2010; Jones, 2006). In the instance where very high levels of scaffolding were required, an array of composite words and distractors could be presented in “drag and drop” panels to assist in the text reconstruction process. Such scaffolding would provide learners a highly structured form of help which would greatly reduce task difficulty, provide opportunities to effectively apply compensatory
strategies and thus increase the likelihood of task success. Individualised provision of help options which are triggered by learners’ relative success and task interaction, in our opinion hold great promise. If learners are explicitly aware, for the sake of argument, that needing more than 5 repetitions to reconstruct a text to an 80 percent accuracy level is an indication that they require “help”, then it may be the case that learners will be more likely to take advantage of the help options made available. Such microadaptive provision of help may be particularly beneficial for lower proficiency learners who typically use help options ineffectively or not at all (Grgurović & Hegelheimer, 2007; Hegelheimer & Tower, 2004; Pujolà, 2002).

The finding that the academic word learning improvements of moderate WRS proficiency participants were significantly greater than those of low WRS proficiency participants also has implications for CALL design. The finding suggests that design mechanisms which effectively guide learners toward appropriate listening materials and adequately scaffolded learning tasks, may be important in enhancing word learning outcomes. Although speculative, the listening material and learning tasks used here appeared to be most suitable for those participants of moderate WRS proficiency. It seems likely that moderate proficiency participants had adequate WRS capabilities to meaningfully engage with the spoken input and at the same time also had sufficient scope for significantly greater word learning improvements.

Generally the vocabulary learning outcomes observed among the participants relate in a number of ways to the existing body of research on vocabulary learning. Firstly, there was a general improvement in the ability of participants from each proficiency level to recognise the phonological form of target words. This learning outcome indicates that the computer application described provides a practical way by which L2 WRS can be improved among L2 learners. This finding is significant in light of the difficulty language learners have in recognising known words from spoken language (Goh, 2000) and also in light of the finding that a minimum level of word recognition from speech is necessary in order to enable L2 learners to effectively engage non-linguistic knowledge to enhance listening comprehension (Graham, et al., 2010). Indeed, the participants in this study had higher levels of word knowledge of high frequency words in the written modality than of they did of high frequency words in the aural modality. A comparison of participants’ knowledge of high frequency words, as measured by the high frequency
words component of a version of the Vocabulary Levels Tests (Schmitt, Schmitt and Clapham, 2001) \((N = 65, \bar{M} = 87.95\%, \bar{SD} = 11.24)\) and as measured by the high frequency words component of a WRS test (Matthews, et al., 2015) \((N = 65, \bar{M} = 56.68\%, \bar{SD} = 17.95)\) presents important insight in this regard. The differences between these two mean scores for words from the same frequency range clearly show that participants had a more robust knowledge of words in the written modality as operationalised by the Vocabulary Levels Tests than they did for the same category of words when presented in the spoken form as operationalised by a WRS test. Such findings are suggestive that the depth of knowledge of words, namely the addition of knowledge of the phonological form of known words, may be developed relatively rapidly through strategic use of computer applications such as the one described in this study. This finding is significant in light of empirical research which shows aspects of depth of vocabulary knowledge can contribute significantly in the prediction of L2 listening comprehension success (Stæhr, 2009). Further, the specific improvements observed among participants from each level of WRS proficiency in the ability to recognise high frequency words from speech is significant as a large proportion of the words which make up spoken discourse are high frequency words (Adolph & Schmitt, 2003; Nation, 2006; Webb & Rodgers, 2009a, 2009b). Indeed the ability to recognise high frequency words from speech is strongly correlated with and predictive of successful performance on task based global listening comprehension tests (Matthews & Cheng, 2015).

6.8.2 Conclusions and future research

The measures of task interaction and task success investigated here are suggested as useful variables which can be used to trigger automated alignment of listening materials with learners of different WRS proficiency. Establishing benchmarks of performance for these measures of task interaction and success also holds promise in the automated delivery of performance based scaffolding. Ensuring learners of differing L2 WRS proficiency are aligned with the most appropriate listening materials and modes of support is an important component of future efforts to more effectively use CALL for L2 WRS development.

It must be emphasised that the guidelines put forward here are based on empirical data drawn from speakers of a common L1 background, Mandarin Chinese, and therefore
may not be generalisable to speakers of other language backgrounds. A potential avenue for future research is to investigate the effect L1 background may have on L2 learners’ interaction with listening materials of various known levels of complexity. Of further interest would be to ascertain the influence listening text complexity may have on the manner by which L2 learners interact with listening material used in the development of L2 WRS. Data drawn from such investigations would provide a more valid basis from which to recommend general guidelines for learners of a broader ranges of L1 backgrounds.

Achieving success in developing more refined approaches to L2 WRS development, especially those with effective triggers for microadaptive individualisation, has significance for listening pedagogy generally. Out-of-class computer-mediated learning which brings about effective and adequately individualised development of L2 WRS provides an attractive option for language educators. Adequate L2 WRS proficiency is essential for good listening comprehension, but its development in the classroom is challenging. As many aspects of overall language proficiency need to be formally addressed in the language classroom, there is typically insufficient time to facilitate adequate levels of WRS. Improved iterations of the application investigated here may not only have a significant role in enhancing learners’ L2 WRS proficiency, but also play a role in reinventing the structure of listening pedagogy generally. Computer-mediated (or mobile device-mediated) WRS development could be the mainstay of independent out-of-class listening skill development. While in the classroom, time could be more fully dedicated towards other essential components of listening pedagogy, such as the development of effective listening strategies and two-way interactive listening tasks.

It is suggested that CALL has the potential to deliver effective WRS development to learners in relatively autonomously, out-of-class contexts. However, crucial to the feasibility of this suggestion is the degree to which design features can cater to CALL users of different WRS proficiency levels. This research has sought to provide empirical data which can be used to begin working towards this design objective. However, the design recommendations based on these data should be viewed as just one phase of an ongoing cyclical design process. Applied research investigating the design features recommended here is necessary to assess their validity in real language learning.
contexts. For example, investigating the way learners of different WRS proficiency levels interact with and derive benefit from CALL for WRS development which operationalise the recommendations put forward here would be of strong interest. Further, tapping deeper into the individual user experience through walkthroughs or think-aloud protocols obtained from learners as they use such CALL approaches would also provide data which could inform future design revisions.

Developing methods to improve the efficiency of L2 WRS assessment is also an important avenue of future research. Although manual modes of scoring of L2 WRS do provide strong levels of inter-rater reliability (Matthews & O’Toole, 2015; Matthews, et al., 2015), a major limitation of the approach is that it requires ongoing reference to a detailed scoring rubric and as such is labour intensive and time consuming. Although it was not always a straightforward task to differentiate between an error in L2 WRS and an error in spelling, scoring rubrics such as those used as part of this research do enable such decisions to be made in a systematic and reliable manner. The development of algorithms which apply principles such as those reflected in the scoring rubric used in this research, may provide a useful starting point from which computers can be used to quantify L2 WRS in a valid, reliable and time effective manner.

The results presented in this paper suggest that variables such as repetition, modification and reconstruction accuracy provide useful indicators for different levels of WRS listening proficiency. CALL applications allow the capacity to respond differentially to the levels so indicated. Such differential responses could include automatic guidance down pathways involving listening texts of differing difficulty and provision of varying degrees of task scaffolding beyond the aural modality. It may seem intuitive to suggest that learner interactions might scaffold CALL design, but these results provide empirical support for that intuition and such support could make appropriate developments more likely.
Summary

Chapter 6 presents a number of recommendations for future iterations of CALL applications for the development of L2 WRS. Key to these recommendations are design features which enable the CALL application to effectively cater to the individual learner’s proficiency level. It is suggested that future iterations of the CALL application which implement the design features described in Chapter 6 may improve the effectiveness of CALL applications aimed at improving L2 WRS.

Research paper 4 is the last research paper associated with this thesis. The following chapter brings together the findings of the four research papers by addressing each of the research objectives and their corresponding primary research questions in detail (see section 1.2). Chapter 7 presents the key findings, implications for practice, and limitations and implications for future research.
Chapter 7: Discussions and recommendations for future research

7.1 Overview

This chapter summarises the key findings in relation to each of the three research objectives and their corresponding research questions which frame the research undertaken as part of this thesis. The three research objectives will be discussed in the order that they appear below.

- Section 7.2.1. Research objective: to quantify the importance of the construct of WRS in L2 listening comprehension. This section addresses the research question: Does second language word recognition from speech (L2 WRS) have sufficient influence on second language listening comprehension to make research into the development of L2 WRS a productive enterprise?

- Section 7.2.2. Research objective: to investigate the effectiveness of a CALL approach for the development of L2 WRS. This section addresses the research question: Does computer-assisted language learning (CALL) provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS a useful way forward?

- Section 7.2.3. Research objective: to provide design recommendations to guide future improved iterations of the CALL approach described. This section addresses the research question: Which design features might be most effective in improving the utility of CALL for the development of L2 WRS?

For each section key findings, implications for practice, limitations and implications for future research are provided.

7.2 Summary of findings and implications for practice and future research

7.2.1 The importance of the construct of WRS in L2 listening comprehension

7.2.1.1 Key findings

Research paper 1 of this study addressed the research objective which was to demonstrate that L2 WRS was a construct of importance in L2 listening. Previous studies had addressed the relationship between L2 word knowledge and L2 listening
comprehension (Bonk, 2000; Milton, Wade and Hopkins, 2010; Stæhr, 2008, 2009). However these studies either did not measure word knowledge in the aural modality, did not measure productive and time constrained knowledge of those words or did not ensure that lower-level word recognition capability was being targeted. The L2 WRS test applied as part of this research successfully tapped a lower-level construct of word knowledge which was mediated through the aural modality, time constrained and productive. The L2 WRS test was shown to have excellent levels of internal consistency (Cronbach’s $\alpha = 0.91$).

L2 WRS scores obtained from participants ($N = 167$) of this phase of the study were found to correlate strongly with scores obtained on a task-based L2 listening comprehension test (IELTS) ($r = .73$). All subcomponents of the L2 WRS tests scores also had strong levels of correlation with L2 listening comprehension scores. The level of correlation observed between L2 WRS scores for words from the third thousand frequency range and L2 listening comprehension ($r = .71$) was particularly noteworthy. Previous studies which have sought to establish the relationship between word knowledge and L2 listening comprehension, such as that of Stæhr (2009) ($r = .70$) and Milton et al., (2010) (Spearman’s $\rho = .67$) have established similar levels of correlation between word knowledge and L2 listening comprehension. This was the case despite the fact that the 3K subsection of the test used in research paper 1 contained far fewer test items than those tests used by Stæhr (2009) and Milton et al., (2010). Therefore a key finding for research paper 1 was that the construct of L2 WRS, as measured by the L2 WRS test, was sensitive to the underlying skills needed to successfully negotiate an authoritative task-based L2 listening test.

The participants’ L2 WRS scores obtained during this phase of the study were shown to have good utility in predicting the variance observed among standardised IELTS L2 listening comprehension test scores. The third thousand and the first thousand L2 WRS scores were able to account for a total of 54% of the variance observed among L2 listening comprehension scores. The importance of L2 WRS in skilled L2 listening was also made clear by the L2 WRS profiles of those participants categorised as possessing modest, competent and good L2 listening comprehension. These three categorises of L2 listening comprehension were established by using the published benchmarks of competence used by the standardised IELTS test (see section 3.5.3). Participants
categorised as good L2 listeners were able to successfully recognise no less than 90% of the words from the first and second thousand L2 WRS test subsections. This finding suggests that a robust capability to recognise high frequency words from speech, as measured by the L2 WRS test, is strongly associated with good levels of L2 listening comprehension, as measured by standardised task-based L2 listening tests.

In direct response to the research question: “Does second language word recognition from speech (L2 WRS) have sufficient influence on second language listening comprehension to make research into the development of L2 WRS a productive enterprise?”, findings from research paper 1 strongly support an answer in the affirmative. The findings from this research paper are strongly suggestive that L2 WRS is an important element of L2 listening teaching and learning. This finding provides robust empirical support for previous work which asserts that lower order or bottom-up listening skills, such as word recognition, are an essential component of skilled L2 listening (Field, 2008a, 2008b; Graham, et al., 2010; Hulstijn, 2003; Tsui and Fullilove, 1998; Wilson, 2003).

7.2.1.2 Implications for practice

The importance of L2 WRS in relation to L2 listening comprehension has a number of implications for practice. The first of these is that diagnostic measurement of L2 WRS should be emphasised in L2 language learning programs. As knowledge of high frequency vocabulary provides high levels of coverage of the majority of words likely to be encountered in L2 spoken discourse, establishing a clear picture of the ability of L2 learners to recognise these words from the aural modality is of strong pedagogical value. Administering L2 WRS tests at the beginning of learning programs will provide diagnostic information which can be used to make reasoned decisions about the relative need to focus on developing L2 WRS among a given cohort of learners. Such information may enable learners to be allocated to different streams of listening classes within which the development of L2 WRS is emphasised proportionately to the specific learning needs of the learners in the group.

The strong association observed between L2 WRS and L2 listening comprehension also suggests the value of focussing on the development of L2 WRS as a part of teaching and learning approaches broadly aimed at the development of L2 listening
comprehension. As a consequence, implementing learning protocols which effectively develop the capacity of L2 learners to recognise L2 words from speech, such as those developed and investigated as part of the research presented as part of this thesis, are likely to be a useful adjunct to existing approaches to L2 listening development.

7.2.1.3 Limitations and implications for future research

A limitation of this study was that the test data used in the analysis of the relationship between L2 WRS and L2 listening comprehension were taken from a single administration of two test instruments at one point in time. This phase of the study thus employed a correlational research paradigm, which has an inherent limitation: correlation does not imply causation. Although a strong association between these constructs can be asserted, due to the correlational nature of the analysis, a conclusion that improved L2 WRS relates directly to improved L2 listening comprehension should not be made based on this data alone. This limitation to the research presents a number of avenues for future research. Firstly, interviews with and direct observations of L2 learners undertaking task-based L2 listening assessments would be of interest. Such data may provide additional information about the relative importance of L2 WRS in the ability to successfully negotiate task-based L2 listening tests. Coupling such data with analysis of results from a battery of L2 WRS tests would provide a more direct route to establish the relative importance of L2 WRS in the achievement of L2 listening comprehension tasks. Additionally, measurement of the L2 WRS and L2 listening comprehension of a cohort of learners over an extended period of time would be of interest. Tracking the relationship between L2 WRS and L2 listening comprehension over longer periods of time would provide a deeper understanding of the strength of association of these two constructs.

Another limitation of this phase of the study was that it was undertaken within a single group of learners with a homogenous linguistic background. On one hand, drawing on results from a cohort of learners who share the same linguistic background is a strength in that it enables more valid generalisations to be made in relation to the likely relationship between L2 WRS and L2 listening comprehension among others with the same linguistic background. However, on the other hand, it limits the ability to generalise more broadly about the relationship between L2 WRS and L2 listening comprehension among cohorts from different linguistic backgrounds to those involved
in the research. This limitation is magnified by findings that show that knowledge of L2 words in the aural modality does vary in association with the linguistic background of the language learner (Milton & Hopkins 2006). A line of future research is thus to investigate the relative strength of association between L2 WRS and L2 listening comprehension of learners from different linguistic backgrounds. The results of such research would provide a more robust empirical basis from which to assert the strength of association between L2 WRS and L2 listening comprehension more generally, as well as provide insight into the degree to which demographic variables such as linguistic background influence the strength of this association.

A further limitation of this phase of the study is that the L2 WRS tests measured the ability to recognise words from only the first three thousand word frequency levels. As such, no information was provided which casts light on the strength of association between the ability to recognise words of lower level frequency, such as those from the fourth and fifth thousand frequency level, and L2 listening comprehension. It is acknowledged that the proportion of words in spoken discourse, including those which make up the stimulus for task based listening assessments, are likely to contain only a relatively small proportion of words which extend beyond the third thousand frequency level. However, the relative importance of being able to recognise low frequency words has on the ability to comprehend L2 speech is not known. Further, the value of the ability to recognise words of lower levels of frequency in predicting L2 listening comprehension levels is also unknown. It is therefore of interest for future research efforts to establish the strength of association between the ability to recognise words beyond the three thousand frequency level as well as to investigate to what degree including such measurement of word knowledge contributes to the power of models which seek to predict L2 listening comprehension.

Lastly, a significant limitation of the procedures of this study related to the feasibility of the L2 WRS test scoring procedures. As it was deemed appropriate to ensure that minor spelling errors did not unduly influence measures of word recognition, a structured scoring rubric was implemented (see appendix 6.1). The manual scoring protocol for the L2 WRS test was shown to be reliable, but these procedures were costly in regards to time. Using the scoring protocols developed as part of this phase of the research and translating them into coded algorithms which enable computer based assessment of L2
WRS would be of strong value. The computerised testing of L2 WRS would make the large scale testing of L2 WRS more feasible and would enable the L2 WRS capability of large cohorts of learners to be measured and the resultant diagnostic information stored. The development, implementation and evaluation of computerised testing of L2 WRS represents an area of future research of strong importance in the development of L2 listening skills.

7.2.2 The utility of computer assisted language learning (CALL) for the development of L2 WRS

7.2.2.1 Key findings

Research paper 2 and research paper 3 (presented in Chapters 4 and 5 respectively) each addressed the research objective which was to examine the utility of computer assisted language learning (CALL) for the development of L2 WRS. The results from both research paper 2 and research paper 3 each provided different forms of empirical evidence for the effectiveness for the CALL approach described in the development of L2 WRS. Research paper 2 was exploratory in nature and provided empirical data which showed learners who used the CALL application experienced significant improvements in L2 WRS. Further, the data obtained from this phase of the study suggested that learners had a positive attitude toward the application and felt that the use of the application had a positive effect on their vocabulary knowledge and English language listening. Research paper 3 employed a pre-test/treatment/post-test research design with two levels of treatment and a control group (see section 5.5). The L2 WRS improvements observed among those who used the CALL application were significantly greater than those who did not use the application. These findings provide an empirical basis from which to support the use of CALL for the development of L2 WRS in L2 language learning contexts. The findings from research paper 3 are significant in that they provide empirical support for the general recommendations put forward by Hulstijn (2003). His model of practice for the development of L2 WRS has hitherto lacked empirical data supporting the validity of his recommendations in real language learning contexts.

The findings from research paper 2 and research paper 3 provide an empirical basis from which to strongly argue for the effectiveness of the CALL approach described in
the development of L2 WRS. As such, in direct response to the research question: “Does computer-assisted language learning (CALL) provide sufficient learning affordances to make web-based applications aimed at improving L2 WRS a useful way forward?”, the answer is in the affirmative. The results of research paper 2 and research paper 3 also provide a clear overview of the design features of CALL which may be assumed to be beneficial to the development of learners L2 WRS and therefore provides a solid foundation from which future iterations of the CALL application described can be implemented, evaluated and improved.

7.2.2.2 Implications for practice

The positive results obtained from research paper 2 and research paper 3 were obtained from groups of learners from two different L2 learning contexts. This result is suggestive that the CALL approach described has a generalisable utility in other similar learning contexts (South East Asian and Chinese tertiary language learning contexts). As this approach is noted to have a positive impact on L2 WRS it is suggested that longer term implementation of similar approaches may yield similar or greater L2 WRS improvements in other equivalent L2 learner contexts. The provision of such computer-mediated approaches provides a feasible and time effective way to improve L2 WRS. The web-based CALL application is able to be accessed and used by learners in an out-of-class manner such that the development of L2 WRS can be undertaken in the students’ own time. The compatibility of this learning approach to web-based delivery suggests that this learning approach has the capacity to be delivered to multiple cohorts of learners simultaneously. Such web-based capability enables L2 WRS development to be implemented across multiple locations, such as to multiple campuses of a given institution. Further, the web-based capability of this application opens the possibility of delivering L2 WRS learning protocols to hand-held devices such as tablets and mobile phones.

The CALL application described as part of this research (see section 4.3.3 and section 5.3.3) can be strategically used in conjunction with L2 WRS test results. A battery of L2 WRS tests can be used to establish areas of weakness in L2 WRS of a cohort of learners. Learning material can then be developed in response to the test results and a targeted intervention can by delivered by providing learners access to the CALL approach described.
The results from research paper 3 also emphasised the importance of the frequency of exposure to target language in the development of L2 WRS. For a significant proportion of the learners in the treatment groups, a positive correlation was observed between exposure to input and L2 WRS gains scores for words which were present in the learning materials. This finding highlights the specificity of this learning approach and the need for language teachers to be strategic in the selection of listening materials for the purposes of developing the capacity to recognise L2 words in speech.

7.2.2.3 Limitations and implications for future research

Neither the production of output nor the provision of text feedback was associated with improved L2 WRS but research paper 3 did not resolve this issue. A failure to obtain a clearer insight into the underlying causes of this lack of association impedes the ability to present highly specific recommendations regarding future iterations of similar CALL applications. As a consequence of this limitation future research which seeks to more fully investigate the relative importance of the provision of text feedback and opportunities to generate output is warranted. In this regard, research which more directly taps the learning behaviour and thought processes employed when students use this application may be of value. For example, more directly measuring the effect of the provision of text feedback on incorrectly recognised words on subsequent capability to recognise those same words would be of interest. Further, it would be valuable to seek learner opinions about what they believe they have learnt after receiving corrective text feedback on incorrectly recognised words. Similarly, using think aloud protocols to elicit the thought processes in play when learners attempt to reconstruct spoken language into the written form would provide valuable data which would enable a clearer understanding of the relative benefit or otherwise of this learning technique.

Another limitation of research paper 2 and research paper 3 relate to the relatively short intervention duration of the studies described. Research paper 2 involved a 7 week intervention duration and research paper 3 involved a 5 week intervention duration. Learning effects that occurred as a result of the production of output and/or the provision of feedback may not have been sufficient for the test instruments used as part of these studies to detect improvements.
Interventions which apply the use of similar applications over longer periods of time may produce results which could provide information as to whether the apparent lack of impact of text feedback and output production was a result of limited time for learning effects to become evident in the group or if these design features are actually not beneficial to the improvement of L2 WRS.

The CALL application was trialled among a group of learners with relatively homogenous linguistic and demographic background variables. Although, an advantage in many regards, the limitation of such research is that there is a limit to the degree to which conclusions from the group can be generalised among learners with other linguistic and demographic profiles. To counter this research limitation, investigating the implementation of the CALL application described and the resultant L2 WRS improvements associated with this implementation, among a variety of learners of various demographic and linguistic backgrounds would be of strong interest.

Another limitation of the research presented in research paper 2 and research paper 3 is that the application described was only investigated in traditional computer based settings. With the increasing availability and use of powerful handheld devices, investigating mobile-assisted language learning (MALL), and how MALL can be used to develop fundamental language skills is a growing research area. As the application described was found to be effective in the development of L2 WRS through use of CALL, empirical data which provides insight into whether the application delivered through MALL is also effective in the development of L2 WRS would be of considerable interest. For example, how the differences between MALL and CALL, such as screen size and control mechanisms, effect the learner’s experience while undertaking interventions designed to improve L2 WRS is a fruitful area for further research. With the ubiquitous nature of mobile technology, it is also of interest to determine if delivery of the application through MALL results in improved learning outcomes when compared to those that use the CALL application. Further, it would be of interest to determine if undertaking L2 WRS interventions through MALL and CALL delivery modes respectively result in differences in interaction patterns and L2 WRS learning outcomes. The need to initiate further research investigating the effectiveness of MALL in the delivery of language learning outcomes is particularly
pressing considering the current paucity of rigorous research in this growing area (Burston, 2015).

7.2.3 The contribution to the body of knowledge which can be used to inform future computer-mediated approaches to develop L2 WRS

7.2.3.1 Key findings

Research paper 4 of this thesis addressed the research objective which was to contribute to the body of knowledge which can be used to inform future computer-mediated approaches to develop L2 WRS. Specifically, this research paper investigated the effect of learners’ L2 WRS proficiency level on their interaction patterns, levels of success and word learning outcomes experienced after using the CALL application. Differences in L2 WRS proficiency had a significant impact on both the way learners interacted with the application and the success they experienced while undertaking the learning tasks delivered by the application. Further, differences in L2 WRS proficiency had a significant effect on the word learning outcomes achieved by learners. A key finding is that the varying levels of linguistic proficiency evident within the target cohort should be adequately considered in the CALL design process.

In response to the research question: “Which design features might be most effective in improving the utility of CALL for the development of L2 WRS?”, this research suggests the following broad recommendations. Firstly, computerised L2 WRS tests should be used at the beginning of a learning cycle so that learners of varying L2 WRS proficiency can be aligned with appropriate listening texts. Findings from research paper 4 indicated that the listening texts used were excessively difficult for some learners and not challenging enough for others. Therefore, mechanisms which automatically guide learners towards appropriate listening texts are warranted. It is also recommended that learner interaction patterns are used as triggers to dynamically align learners with appropriate texts and scaffolding modes.

7.2.3.2 Implications for practice

The primary implication for practice of this phase of study is the need to closely consider differences in proficiency levels of L2 word recognition skills when designing CALL for the development of L2 WRS. Research paper 4 indicates that it would be
beneficial to develop and use graded listening materials as part of interventions aimed at developing L2 WRS. The development of such materials would enable the L2 WRS proficiency levels of language learners to be aligned with listening materials of an appropriate difficulty level. Further, it is important that learners of different L2 WRS proficiency levels are afforded appropriate levels of task scaffolding to enhance task success and potential word learning outcomes.

7.2.3.3 Limitations and implications for future research

This phase of the study only used a single variable of individual difference in an effort to determine its impact on task interaction, task success and word learning. Additionally, the study was undertaken among a relatively homogenous group in relation to linguistic background. Investigating other variables of individual difference such as motivation level, linguistic background, gender, and general language proficiency (rather than specifically L2 WRS proficiency) would cast more light on the influence of individual difference on the manner by which learners used the CALL application described.

An additional limitation to this research is that qualitative data was not gathered in regards to the reasons why learners interacted with the CALL application differently. Undertaking walkthroughs and interviews of learners of different proficiency levels while using the computer application may provide deeper insight into the factors other than L2 WRS proficiency which contributed to the differences in interaction, task success and word learning outcomes observed between learners of different proficiency level.

A final and major limitation is that the recommendations which have been put forward in regards to future improved iterations of the CALL approach described, are yet to be empirically verified. In order to counter this limitation these recommendations would need to be developed as part of a CALL program and investigated in real language learning contexts. The implementation of a CALL application which operationalises these recommendations would be of interest. Additionally, the resultant gathering of both qualitative and quantitative data from learners engaged with the use of improved iterations of the CALL approach described would also be illuminating.
7.3 Concluding remarks

The ability to recognise L2 words from speech is an essential component of skilled L2 listening. Accordingly, the systematic development of L2 WRS should be an important component of L2 language learning curricula. There has been little recent research effort directed towards investigating approaches which could be used to effectively develop L2 WRS among groups of language learners. The research undertaken as part of this thesis confirms the value of CALL in the development of L2 WRS in real language learning contexts and therefore has successfully contributed to efforts aimed at filling this gap in current knowledge. However, due to the limited amount of research which has been undertaken in relation to the development of computer-mediated development of L2 WRS, there is a need to expand on the empirical base of information established here. A challenge for future research is to test refined iterations of the CALL approach described here, in a variety of research contexts using a variety of data collection methods.
References


Rost, M. (2002). *Teaching and researching listening*. Harlow, UK: Longman.


Vandergrift, L. (2002). It was nice to see that our predictions were right: Developing metacognition in L2 listening comprehension. *The Canadian Modern Language Review, 58*(2), 555-575.


Appendices

Appendix 1.1

This thesis, “Investigating a computer-assisted language learning approach for the development of second language word recognition from speech”, incorporates four chapters which comprise the substance of articles published in and submitted to scholarly journals.

The articles are:


We, the authoring team for these articles, attest that Research Higher Degree candidate Joshua Matthews made the major contribution to the preparation of these articles in the areas of conceptual development; literature review; instrument design, production and implementation; data collection and analysis; text production; and publication facilitation.

(Signature of Co-Author)
Dr John Mitchell O’Toole
Date: 5 May 2015

(Signature of Co-Author)
Associate Professor Junyu Cheng
Date: 5 May 2015

(Signature of Co-Author)
Associate Professor Shen Chen
Date: 5 May 2015

(Signature of Candidate)
Joshua Aron Matthews
Date: 5 May 2015

(Signature of Assistant Dean Research Training)
Associate Professor Ros Smith
Date: 19 May 2015
Appendix 3.1 (see section 3.4.2.1)

You will see 89 sentences. Each sentence has only one word missing. Listen to the recording and fill in the blank according to what you hear. Each sentence will be heard only once.

1. The most ............................................ language is South Korean
2. This country depends on ............................................
3. ............................................ 50 million people live in this country
4. Parts of this country are covered in ............................................
5. The city of Busan is in the ............................................
6. Great ............................................ has been put into improving this country
7. This country ............................................ on making cars
8. There are many ............................................ in this country
9. The ............................................ of this country is quite large
10. Many ............................................ Korea as a very beautiful country.
11. South Korea is a ............................................ small country.
12. Korea doesn’t have many ............................................
13. Only one other country shares a ............................................ with South Korea.
14. ............................................ is not a large industry in this country.
15. The ............................................ language of this country is Korean.
16. Daegu is the ............................................ biggest city in South Korea.
17. ............................................ is important to this country.
18. A product of this country is ............................................
19. Korea ............................................ business people from around the world.
20. Many people are ............................................ by Korea’s natural beauty.
21. The ............................................ port in this country is in Busan.
22. Many people ............................................ Korea every year
23. ............................................ about Korea can be found on the internet.
24. Korea has a very good ............................................ system.
25. There are several ............................................ schools in Korea.
26. Korean food is ............................................ popular
27. Korea has a large number of ............................................
28. Most of this country is surrounded by ............................................
29. Korea produces ............................................
30. The ............................................ in this country changes very quickly.
31. The number of people in this country is ............................................ growing.
32. The cost of living in this country is ............................................
33. This country has many ............................................
34. Great ............................................ has been made to develop this country.
35. There are many ............................................ in this country.
36. Many ............................................ Vietnam as a very beautiful country.
37. An important food in Vietnam is ............................................
38. ............................................ is an important industry in this country.
39. Ba Ria is the ............................................ largest city in Vietnam.
40. One product of this country is ............................................
41. Various types of ............................................ are popular in Vietnam.
42. ............................................ are very important to this country.
43. This country has a good ............................................ system
44. Vietnamese food is ............................................ popular.
45. On the coast of this country there is a large ............................................
46. The ............................................ in this country changes very quickly.
47. The number of people in this country is ............................................ growing.
48. Korea is a country with a strong ............................................
49. Many ............................................ work in South Korea
50. It ............................................ cars and computers.
51. Busan and Seoul are very ............................................ cities.
52. The ............................................ language is standard Korean.
53. This country has a good ............................................ system.
54. Recently this country has experienced ............................................ growth.
55. Korea has many ............................................ to improve production.
56. Korea has received ............................................ from many different countries.
57. Korea needs to ............................................ coal from other countries.
58. Milk is an important ............................................ product of this country.
59. Korea has the ............................................ to continue to modernise.
60. This country has many ............................................ varieties of food.
61. Korea is a ............................................ country.
62. China and Korea have strong ............................................
63. Korea is a ............................................ country in East Asia.
64. Korea’s second largest city is ............................................ in the south.
65. There are few ............................................ in South Korea.
66. ............................................ languages are uncommon in this country.
67. Korea has ............................................with many countries.
68. Korea produces many types of ............................................
69. Korea has developed a new trading ............................................
70. In the past Korea has been ............................................
71. ............................................ sports are popular in Korea.
72. Recently the industrial ............................................ of this country has increased.
73. Recently Korea has ............................................ more trade with other countries.
74. The ............................................of Korean food makes it popular around the world.
75. Korea will ............................................ several new laws next year.
76. Vietnam produces a ............................................ of things.
77. Industry is the largest ............................................ in this country.
78. Within Vietnam there are many different ............................................
79. The ............................................ in this country is very important.
80. Vietnam uses many ............................................ to improve production.
81. This country has experienced a ............................................ of market reforms.
82. Vietnam has many important ............................................ sites.
83. Vietnam has experienced ............................................ development.
84. This country is making a ............................................ to Southeast Asian growth.
85. Vietnam has several oil ............................................
86. The protection of Vietnam’s forests is an important ............................................
87. ............................................ travel is cheap in this country.
88. Recently Vietnam has ............................................ trade with other countries.
89. There are various forms of ............................................ within this country.
Target words

These target words show the correct entries for the test above. They are provided as a reference for the reader and were not supplied as a component of the test.

**Appendix 4.1** (see section 4.3.5.1)

1. Many different ________________ are practised in India.
2. China is a world leader in ________________.
3. Languages often have many different ________________.
4. Malaysia depends on the export of natural ________________.
5. Brunei is on the northern ________________ of the island of Borneo.
6. Korea and Japan are Asian ________________.
7. Thailand shares ________________ with Malaysia and Burma.
8. Indonesia produces ________________.
9. Thailand has never been occupied by European ________________.
10. Indonesia is made up of a huge number of ________________.
11. ________________ is an important product of Asia.
12. Chinese is one of the most commonly spoken ________________ in Asia.
13. Two important farming ________________ include wheat and rice.
14. A large ________________ of people speak English.
15. Burma depends on the ________________ of wood.
16. America has many ________________.
17. Japan is an Asian ________________.
18. There are a large number of ________________ living in India.
19. There are hundreds of countries in the ________________.
20. The global ________________ is growing rapidly.

**Target words** These target words show the correct entries for the test above. They are provided as a reference for the reader and were not supplied as a component of the test.

Religions / Industry / Dialects / Gas / Part / Countries / Borders / Oil / Forces / Islands / Rubber / Languages / Products / Number / Production / States / Nation / People / World / Population
Appendix 4.2 (see section 4.3.5.2)

| 1 | The farmer sold the **rubber**          |
| 2 | The woman talks to all of the **people**|
| 3 | Thailand and Burma are quite **different** |
| 4 | The ship will wait in the **port**      |
| 5 | Thailand is a **country**               |
| 6 | A million is a large **number**         |
| 7 | Italy is a country that depends on **tourism** |
| 8 | The man is a good **leader**            |
| 9 | The bus drove along the road towards the **southeast** |
|10 | Many types of animals are found in the **world** |
|11 | The school in the city is **huge**      |
|12 | It was a good **investment**            |
|13 | The people at the party ate all of the **rice** |
|14 | The man is **Christian**                |
|15 | The people voted for **democracy**      |
|16 | All the students in the school are **European** |
|17 | Islam is a **religion**                 |
|18 | Thailand is in the Asian **group**      |
|19 | The stove uses **gas**                  |
|20 | Singapore is located at the end of a **peninsula** |
|21 | The food is **great**                   |
|22 | Farming is **agriculture**              |
|23 | The customer was very happy with quality of the **service** |
|24 | All the people from the village are very **successful** |
|25 | China has a very strong **economy**     |
|26 | Oil is a very important **export**      |
|27 | Sweden has a good **government**        |
|28 | The girl came **second**                |
|29 | The boy really wants to sell the **product** |
|30 | The students always ask for more **information** |
**Appendix 4.3** (see section 4.3.5.3)

**Pre-test**

Malaysia is located on the peninsula bordering Thailand and the northern part of the island of Borneo. A majority of the population are Muslim, but there are also Buddhist, Christian, Hindu, and other minority religions. The languages spoken are Malay, English, and Chinese. The economy depends on the production of rubber and palm oil, and also on the export of oil and natural gas. Its primary export partners are Singapore and China.

**Post-test**

China is a large east Asian country which shares borders with many different nations. It is currently the most populated country in the world. Standard Chinese is the official language, but many dialects are also spoken. China is the largest exporter in the world and has a strong and growing economy. China is the world leader in industrial output. Its main export partner is the United States.
Appendix 5.1 (see section 5.4.2.1)

Word recognition pre-test

1. The most ............................................ language is South Korean.
2. This country depends on ............................................
3. ............................................ 50 million people live in this country.
4. Parts of this country are covered in ............................................
5. The city of Busan is in the ............................................
6. Great ............................................ has been put into improving this country.
7. This country ............................................ on making cars.
8. There are many ............................................ in this country.
9. The ............................................ of this country is quite large.
10. Many ............................................ Korea as a very beautiful country.
11. South Korea is a ............................................ small country.
12. Korea doesn’t have many ............................................
13. Only one other country shares a ............................................ with South Korea.
14. ............................................ is not a large industry in this country.
15. The ............................................ language of this country is Korean.
16. Daegu is the ............................................ biggest city in South Korea.
17. ............................................ is important to this country.
18. A product of this country is ............................................
19. Korea ............................................ business people from around the world.
20. Many people are ............................................ by Korea’s natural beauty.
21. The ............................................ port in this country is in Busan.
22. Many people ............................................ Korea every year.
23. ............................................ about Korea can be found on the internet.
24. Korea has a very good ............................................ system.
25. There are several ............................................ schools in Korea.
26. Korean food is ............................................ popular.
27. Korea has a large number of ............................................
28. Most of this country is surrounded by ............................................
29. Korea produces ............................................
30. The ............................................ in this country changes very quickly.
31. This country doesn’t produce natural ............................................
32. The number of people in this country is ............................................ growing.
33. Korea is a country with a strong ............................................
34. Many ............................................ work in South Korea.
35. It ............................................ cars and computers.
36. Busan and Seoul are very ............................................ cities.
37. The ............................................ language is standard Korean.
38. This country has a good ............................................ system.
39. Recently this country has experienced ............................................ growth.
40. Korea has many ............................................ to improve production.
41. Korea has received ............................................ from many different countries.
42. Korea needs to ............................................ coal from other countries.
43. Milk is an important ............................................ product of this country.
44. Korea has the ............................................ to continue to modernise.
45. This country has many ............................................ varieties of food.
46. Korea is a ............................................ country.
47. China and Korea have strong ............................................
48. Korea is a ............................................ country in East Asia.
49. Korea’s second largest city is ............................................ in the south.
50. There are few ............................................ in South Korea.
51. ............................................ languages are uncommon in this country.
52. Korea has ............................................ with many countries.
53. Korea produces many types of ............................................
54. Korea has developed a new trading ............................................
55. In the past Korea has been ............................................
56. ............................................ sports are popular in Korea.
57. Recently the industrial ............................................ of this country has increased.
58. Recently Korea has ............................................ more trade with other countries.
59. The ............................................ of Korean food makes it popular around the world.
60. Korea will ............................................ several new laws next year.
The target words show the correct entries for the test above. They are provided as a reference for the reader and were not supplied as a component of the test.

common, shipping, currently, mountains, east, effort, depends, towns, population, regard, relatively, lakes, border, mining, standard, fourth, agriculture, fruit, attracts, charmed, busiest, tour, information, health, international, especially, islands, ocean, rice, weather, rubber, steadily, economy, contractors, exports, similar, major, financial, significant, projects, investment, source, primary, potential, regional, secure, links, dominant, located, immigrants, minority, partnerships, technology, scheme, occupied, professional, output, predicted, diversity, initiate

Appendix 5.2 (see section 5.4.2.1)

Word recognition post-test

1. Vietnamese is the most ............................................ language.
2. The cost of living in this country is ............................................
3. ............................................ 90 million people live in Vietnam.
4. This country has many ............................................
5. The city of Binh Dinh is in the ............................................
6. Great ............................................ has been made to develop this country.
7. This country ............................................ on manufacturing.
8. There are many ............................................ in this country.
9. The ............................................ of this country is large.
10. Many ............................................ Vietnam as a very beautiful country.
11. Vietnam is a ............................................ long country.
12. An important food in Vietnam is ............................................
13. Vietnam ............................................ borders with three other countries.
14. ............................................ is an important industry in this country.
15. Vietnamese is the ............................................ language of this country.
16. Ba Ria is the ............................................ largest city in Vietnam.
17. ............................................ is very important to Vietnam.
18. One product of this country is ............................................
19. Vietnam ............................................ visitors from around the world.
20. Various types of ............................................ are popular in Vietnam.
21. One of the ............................................ cities in Vietnam is Hanoi.
22. ............................................ are very important to this country.
23. ............................................ about this country can be found on the internet.
24. This country has a good ............................................ system.
25. This country has several ............................................ schools.
26. Vietnamese food is ............................................ popular.
27. This country has several famous ............................................
28. On the coast of this country there is a large ............................................
29. This country produces ............................................
30. The ............................................ in this country changes very quickly.
31. Vietnam produces a large amount of natural ............................................
32. The number of people in this country is ............................................ growing.
33. The ............................................ of Vietnam is growing.
34. Vietnam produces a ............................................ of things.
35. ............................................ are very important to Vietnam.
36. Industry is the largest ............................................ in this country.
37. Vietnamese is the ............................................ language.
38. Within Vietnam there are many different ............................................
39. This country has a ............................................ number of different languages.
40. The ............................................ in this country is very important.
41. Many foreign countries have made ............................................ in this country.
42. Vietnam uses many ............................................ to improve production.
43. Fish is an important ............................................ product of this country.
44. This country has experienced a ............................................ of market reforms.
45. Different parts of Vietnam have different ............................................ languages.
46. Vietnam has many important ............................................ sites.
47. Vietnam has ............................................ with Cambodia.
48. Vietnam has experienced ............................................ development.
49. Vietnam’s second largest city is ............................................ in the north.
50. This country is making a ............................................ to Southeast Asian growth.
51. Many ............................................ languages exist in the country.
52. Vietnam has several oil ............................................
53. ............................................ is important to this country’s development.
54. The protection of Vietnam’s forests is an important ............................................
55. This country has been ............................................ in the past.
56. ............................................ travel is cheap in this country.
57. In recent times Vietnam’s coal ............................................ has reduced.
58. Recently Vietnam has ............................................ trade with other countries.
59. There is a large ............................................ of languages and cultures.
60. There are various forms of ............................................ within this country.

**Target words**

These target words show the correct entries for the test above. They are provided as a reference for the reader and were not supplied as a component of the test.

- common, rising, currently, forests, east progress, depends, villages, population, recognise, relatively, fish, shares, farming, standard, third, agriculture, coffee, attracts, dance, busiest, tourists, information, education, international, extremely, islands, bay, rice, temperature, rubber, gradually, economy, range, exports, sector, major, cultures, significant, environment, investments, processes, primary, series, regional, cultural, links, constant, located, contribution, minority, corporations, technology, task, occupied, domestic, output, promoted, diversity, transport
When scoring word recognition tests it should be remembered that:

1. The test is not a spelling test. However, it is important to ensure that there is compelling evidence that the phonological form of the word has been recognised by the test taker in order for a mark or half mark to be ascribed.

2. It will be necessary to make some decisions with some degree of subjectivity as part of the scoring procedure. It is therefore of primary importance to make such decisions in a systematic manner. Making systematic decisions based on the principles inherent to the rubric will assist in the reliability of the scoring process.

### Appendix 6.1 (see section 6.4.3.3)

#### WRS Scoring rubric with rationale and examples

**General Instructions**

A rubric which outlines the principles of the scoring system is provided below. Examples which illustrate the guiding principles to be used for scoring word recognition from speech (WRS) are given. The example answers provided are not a comprehensive list. A list will need to be compiled by the marker and each unique answer will need to be considered on a case by case basis. As marking decisions are made, the list of previously marked words will need to be considered in order to work through problematic answers in a systematic manner. Knowledge of the spelling and sound system conventions of English will enable the marker to make a decision based on implicit knowledge of the target language words. At times reasoned decisions must be made based on the scoring rubric or previous marking decisions.

When scoring word recognition tests it should be remembered that:

1. The test is not a spelling test. However, it is important to ensure that there is compelling evidence that the phonological form of the word has been recognised by the test taker in order for a mark or half mark to be ascribed.

2. It will be necessary to make some decisions with some degree of subjectivity as part of the scoring procedure. It is therefore of primary importance to make such decisions in a systematic manner. Making systematic decisions based on the principles inherent to the rubric will assist in the reliability of the scoring process.

**General Instructions**

A rubric which outlines the principles of the scoring system is provided below. Examples which illustrate the guiding principles to be used for scoring word recognition from speech (WRS) are given. The example answers provided are not a comprehensive list. A list will need to be compiled by the marker and each unique answer will need to be considered on a case by case basis. As marking decisions are made, the list of previously marked words will need to be considered in order to work through problematic answers in a systematic manner. Knowledge of the spelling and sound system conventions of English will enable the marker to make a decision based on implicit knowledge of the target language words. At times reasoned decisions must be made based on the scoring rubric or previous marking decisions.

When scoring word recognition tests it should be remembered that:

1. The test is not a spelling test. However, it is important to ensure that there is compelling evidence that the phonological form of the word has been recognised by the test taker in order for a mark or half mark to be ascribed.

2. It will be necessary to make some decisions with some degree of subjectivity as part of the scoring procedure. It is therefore of primary importance to make such decisions in a systematic manner. Making systematic decisions based on the principles inherent to the rubric will assist in the reliability of the scoring process.

### Score Rubric

<table>
<thead>
<tr>
<th>Score</th>
<th>Principle</th>
<th>Comments / Advice</th>
<th>Examples Target word → example answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>The word has been written in the correct orthographic form (both American and English spelling conventions accepted)</td>
<td>This is the easiest type of answer to score as there is no subjectivity involved.</td>
<td>agriculture → agriculture technology → technology regional → regional attracts → attracts predicted → predicted relatively → relatively charmed → charmed recognise → recognize / recognise</td>
</tr>
<tr>
<td>1.0</td>
<td>Additional/incorrect consonant or vowel which doesn’t affect phonological form of the word (marker must apply knowledge of the various ways sounds in English can be represented in various orthographic forms)</td>
<td>Subjectivity involved. First identify that there has been an orthographic error (spelling error). Then consider the orthographic error and decide whether, despite the error, the manner by which the word has been represented in writing enables the target word to be pronounced with the correct phonological form. As the recognition of the phonological form of the word is the target construct in this test, this type of orthographical error should not be penalised.</td>
<td>agriculture → agirculture technology → technology regional → regional attracts → attracts predicted → predicted relatively → relatively</td>
</tr>
<tr>
<td>0.5</td>
<td>Minor alteration to phonemic form of word through addition/deletion/substitution of vowels or consonants with some ambiguity of phonological form of the resultant word</td>
<td>Subjectivity involved. First identify that the manner by which the word has been represented in writing results in the target word being pronounced in an erroneous phonological form. However despite these minor errors, the scorer is still satisfied that the test taker has recognised the target word. These words typically contain one incorrect vowel or consonant sound.</td>
<td>agriculture → agrilutner technology → tactknilage diversity → daversity immigrants → immgrance significant → significan temperature → tapemature mountains → mtains</td>
</tr>
<tr>
<td>0.5</td>
<td>Incorrect conjugation / incorrect form of verb but clear evidence that the root word has been recognised</td>
<td>Some subjectivity involved. The core element of the target word has been correctly recognised however an error in inflection (for example occupy instead of occupied) or word form (for example economic instead of economy) has been made.</td>
<td>located → locate / locating occupied → occupy predicted → predict initiate → initiated economy → economic relatively → relative</td>
</tr>
<tr>
<td>0.5</td>
<td>Incorrect pluralisation / use of “s” / failure to identify a plural</td>
<td>Simple to identify. If a word is correct except that it has been made a plural or a singular in error.</td>
<td>contactors → contractor processes → process regard → regards</td>
</tr>
<tr>
<td>0.0</td>
<td>Significant alteration of vowels/consonants leading to significant phonemic alteration to word or significant ambiguity in the recognition of the target word – this may involve one of more significant vowel/consonant alterations</td>
<td>Subjectivity involved. The scorer is not sufficiently satisfied that the test taker has clearly recognised the target word. This may involve a single or multiple incorrect vowel or consonant sounds which radically alters the phonological form of the word when pronounced.</td>
<td>attracts → attracts occupied → occupied ocean → ocean corporations → coptions environment → inverment professional → perfessional tourist → taris villages → vellege common → comond</td>
</tr>
<tr>
<td>0.0</td>
<td>No answer provided</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Comments / Advice*

**Comments / Advice**

The scorer is not sufficiently satisfied that the test taker has clearly recognised the target word. This may involve a single or multiple incorrect vowel or consonant sounds which radically alters the phonological form of the word when pronounced.